

Rainer Alt and Bogdan Franczyk (Editors)

**Proceedings of the 11th International Conference on
Wirtschaftsinformatik (WI2013)**

Feb 27 – Mar 01 2013

University Leipzig, Germany

Volume 2

The logo for WI2013 is a stylized, blocky representation of the text. The 'W' is red, the 'I' is red, the '2' is magenta, the '0' is magenta, the '1' is magenta, and the '3' is purple. The characters are thick and have a slightly irregular, hand-drawn appearance.

© R. Alt and B. Franczyk 2013

The papers in this book comprise the proceedings of the 11th International Conference on Wirtschaftsinformatik. The copyrights of papers belong to the authors. The papers reflect the authors' opinions, and in the interests of timely dissemination, are published without change. Their inclusion in this book does not necessarily constitute endorsement by the editors. The editors have no responsibility for the persistence and accuracy of URLs referred to in this publication and do not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Volume 1 of the conference proceedings contains papers submitted to the following tracks:

- Track 1: Individualization and Consumerization
- Track 2: Integrated Systems in Manufacturing Industries
- Track 3: Integrated Systems in Service Industries
- Track 4: Innovations and Business Models
- Track 5: Information and Knowledge Management

ISBN 978-3-00-041359-9

Volume 2 of the conference proceedings contains papers submitted to the following tracks:

- Track 6: Business Process and Service Management
- Track 7: Sustainability and Security
- Track 8: Modeling and Development Methods
- Track 9: Integration and Architectures
- Track 10: Research Methods and Philosophy
- Track 11: General Track - Visionary and Cross-Disciplinary Issues
- Student Consortium

ISBN 978-3-00-041360-5

Cover layout by: André Apel and Marius Czernetzki
EWERK IT GmbH Leipzig, Markt 16, D-04109 Leipzig
www.ewerk.com

Printed by: Merkur Druck- & Kopierzentrum GmbH & Co. KG
Ritterstraße 10, D-04109 Leipzig
www.merkurdruck.de

Table of Contents

Business Process and Service Management

Track Chairs: *Jörg Becker, Jan Mendling*

Vorhersagemodell für die Verfügbarkeit von IT-Services aus Anwendungssystem-landschaften <i>Sascha Bosse, Matthias Splith, Klaus Turowski</i>	935
Measuring IT System Value with Entity-Specific Factors Influencing Process Runtime Efficiency <i>Jan Muntermann, Moritz C. Weber, Carola Wondrak</i>	951
Agilität von Geschäftsprozessen trotz Compliance <i>Kai Kittel</i>	967
Risikomodellierung in strategischen Liefernetzwerken für hybride Wertschöpfung <i>Holger Schrödl, Laura Geier, Matthias Geier, Paulina Simkin</i>	983
A Holistic Approach for Integrating Methods in Quality Management <i>Florian Johannsen</i>	999
Leveraging Service Incident Analytics to Determine Cost-Optimal Service Offers <i>Axel Kieninger, Detlef Straeten, Steven Kimbrough, Björn Schmitz, Gerhard Satzger</i>	1015

Sustainability and Security

Track Chairs: *Hans-Knud Arndt, Kai Rannenberg*

Unrealistischer Optimismus der Cloud Computing Anbieter bezüglich IT Sicherheitsrisiken – Eine Bedrohung für die Nutzer? <i>André Loske, Thomas Widjaja, Peter Buxmann</i>	1035
Managing the Access Grid - A Process View to Minimize Insider Misuse Risks <i>Stefan Meier, Ludwig Fuchs, Günther Pernul</i>	1051
Systematic Review and Meta-Analysis of IS Security Policy Compliance Research. First Steps towards Evidence-Based Structuring of the IS Security Domain <i>Danijel Milicevic, Matthias Goeken</i>	1067

Experimental Evaluation of a Process Benchmarking Tool in a Green Business Process Management Context <i>Matthias Gräuler, Frank Teuteberg</i>	1083
Nachhaltigkeit in IT-Organisationen - Ein Forschungsrahmen für das Nachhaltige Informationsmanagement <i>Koray Ereğ, Fabian Löser, Rüdiger Zarnekow</i>	1099
Kennzahlenbasierte Erfolgsmessung von Green IT-Maßnahmen – Eine empirische Analyse zum aktuellen Stand in Forschung und Praxis <i>Nicky Opitz, Florian Thies, Koray Ereğ, Lutz M. Kolbe, Rüdiger Zarnekow</i>	1115
Modellgestützte Analyse und Optimierung der Energieeffizienz betrieblicher Informations- und Kommunikationstechnik <i>Steffen Vaupel, Jörg Leukel</i>	1131
Energy-aware Service Allocation for Cloud Computing <i>Tobias Widmer, Marc Premm, Paul Karaenke</i>	1147

Modeling and Development Methods

Track Chairs: *Stefan Strecker, Klaus Turowski*

A Formal Specification of the Horus Modeling Language Using FDMM <i>Hans-Georg Fill, Susan Hickl, Dimitris Karagiannis, Andreas Oberweis, Andreas Schoknecht</i>	1165
Applicability of Business Process Model Analysis Approaches – A Case Study in Financial Services Consulting <i>Jörg Becker, Mathias Eggert, Dennis Bruning, Jan Saat</i>	1181
An Empirical Investigation on the Design of Process Architectures <i>Monika Malinova, Henrik Leopold, Jan Mendling</i>	1197
Analyse des Qualitätsdiskurses zur Modellverständlichkeit in experimentellen Studien <i>Constantin Houy, Armella-Lucia Vella, Tom Thaler, Peter Fettke, Peter Loos</i>	1213
Vom SOM-Geschäftsprozessmodell zur vollständig dokumentenorientierten RESTful SOA – Ein modellbasierter Ansatz <i>Matthias Wolf, Thomas Benker</i>	1229

Using a Generic Model Query Approach to Allow for Process Model Compliance Checking – An Algorithmic Perspective <i>Sebastian Bräuer, Patrick Delfmann, Hanns-Alexander Dietrich, Matthias Steinhorst</i>	1245
Nutzung des Viewpoint-Konzepts zur Unterstützung kollaborativer Modellierung – Konzeption und prototypische Implementierung <i>Julian Krumeich, Dirk Werth, Peter Loos</i>	1261
Eine Untersuchung der Potentiale automatisierter Abstraktionsansätze für Geschäftsprozessmodelle im Hinblick auf die induktive Entwicklung von Referenzprozessmodellen <i>Jana-Rebecca Rehse, Peter Fettke, Peter Loos</i>	1277
Spezifikation von funktionalen und nichtfunktionalen Systemanforderungen auf Basis von Geschäftsprozessmodellen <i>Beate Hartmann, Andree Teusch, Matthias Wolf</i>	1293
SIMchronization: Eine Modellierungsmethode zur Synchronisation von Material- und Informationsflüssen in der Instandhaltung <i>Christoph Prackwieser</i>	1309
Modeling Clinical Pathways - Design and Application of a Domain-Specific Modeling Language <i>Martin Burwitz, Hannes Schlieter, and Werner Esswein</i>	1325

Integration and Architectures

Track Chairs: *Witold Abramowicz, Robert Tolksdorf*

Ontology-Based QoS Aggregation for Composite Web Services <i>Paul Karaenke, Joerg Leukel, Vijayan Sugumaran</i>	1343
Eine Bestandsaufnahme von Standardisierungs-potentialen und -lücken im Cloud Computing <i>Robin Fischer, Christian Janiesch, Joachim Strach, Nicolai Bieber, Wolfgang Zink, Stefan Tai</i>	1359
Six Principles for Governing Mobile Platforms <i>Julia Manner, David Nienaber, Michael Schermann, Helmut Krcmar</i>	1375
Integration von Planungssystemen in der Logistik – Ansatz und Anwendung <i>Christoph Augenstein, Stefan Mutke, André Ludwig</i>	1391

Evaluating the App-Store Model for Enterprise Application Software and Related Services <i>Stefan Wenzel, Francesco Novelli, Christoph Burkard</i>	1407
---	------

Research Methods and Philosophy

Track Chairs: *Ulrich Frank, Frédéric Thiesse*

Theorieentwicklung in der Akzeptanzforschung: Entwicklung eines Modells auf Basis einer qualitativen Studie <i>Kristin Vogelsang, Melanie Steinhüser, Uwe Hoppe</i>	1425
--	------

Building Taxonomies in IS and Management – A Systematic Approach Based on Content Analysis <i>Dennis M. Steininger, Manuel Trenz, Daniel J. Veit</i>	1441
---	------

Qualitative Comparative Analysis in Information Systems and Wirtschaftsinformatik <i>Roy Wandler, Helena Bukvova, Sandra Leupold</i>	1457
---	------

General Track - Visionary and Cross-Disciplinary Issues

Track Chairs: *Jörn Altmann, Stefan Kirn*

Envy on Facebook: A Hidden Threat to Users' Life Satisfaction? <i>Hanna Krasnova, Helena Wenninger, Thomas Widjaja, Peter Buxmann</i>	1477
--	------

Students' Intentions to Use Wikis in Higher Education <i>Christian Kummer</i>	1493
--	------

Effects of the Network Structure on the Dynamics of Viral Marketing <i>Marek Opuszko, Johannes Ruhland</i>	1509
---	------

A Comprehensive Analysis of E-Government Adoption in the German Household <i>Cigdem Akkaya, Petra Wolf, Helmut Krcmar</i>	1525
--	------

The Coevolution of Network Structure and Perceived Ease of Use <i>Johannes Putzke, Kai Fischbach, Detlef Schoder, Peter Gloor</i>	1541
--	------

A Structural Perspective on Belief Formation <i>Daniel Grgecic</i>	1557
---	------

Teaching the Chief Information Officers: An Assessment of the Interrelations within their Skill Set <i>Matthias Boehm, Carl Stolze, Oliver Thomas</i>	1573
Towards Understanding Social Software and Its Impact on Corporate E-Learning Motivation <i>Thomas Wirtky, Sven Laumer, Andreas Eckhardt, Tim Weitzel</i>	1589
Resistant Use of Project Management Methodologies – Using Psychology to Rethink the Influence of Methodology Attributes <i>Kunal Mohan, Frederik Ahlemann</i>	1603
Predicting The Disclosure of Personal Information on Social Networks: An Empirical Investigation <i>Thomas Buckel, Frédéric Thiesse</i>	1619
Personal Information Markets AND Privacy: A New Model to Solve the Controversy <i>Alexander Novotny, Sarah Spiekermann</i>	1635
Legal Literacy and Users’ Awareness of Privacy, Data Protection and Copyright Legislation in the Web 2.0 Era <i>Katharina Steininger, David Rückel</i>	1651
A Household-Oriented Approach to The Benefits of Vehicle-To-Grid-Capable Electric Vehicles <i>Tobias Brandt, Stefan Feuerriegel, Dirk Neumann</i>	1667
A Decision Support Tool for the Risk Management of Offshore Wind Energy Projects <i>André Koukal, Michael H. Breitner</i>	1683
Simplified Bid Languages – A Remedy to Efficiency Losses in Large Spectrum Auctions <i>Stefan Mayer, Pasha Shabalin</i>	1699

Student Consortium

Track Chair: *Jan vom Brocke*

Erfolgsfaktoren von On-Demand-Enterprise-Systemen aus der Sicht des Anbieters – eine explorative Studie <i>Alexander Wieneke, Sebastian Walther, Rüdiger Eichin, Torsten Eymann</i>	1717
--	------

Cloud Computing lohnt sich (noch) nicht <i>Carsten Ingo Berendes, Markus Ertel, Thomas Röder, Thomas Sachs, Thomas Süptitz, Torsten Eymann</i>	1733
Foundations for the Integration of Enterprise Wikis and Specialized Tools for Enterprise Architecture Management <i>Max Fiedler, Matheus Hauder, Alexander W. Schneider</i>	1749
Green IT-Service-Management: Eine empirische Voruntersuchung der konzeptionellen Grundlagen <i>Jörg Schneider, Markus Reiter</i>	1765
Auswirkungen von In-Memory-Datenmanagement auf Geschäftsprozesse im Business Intelligence <i>Patrick Wessel, Sebastian Köffer, Jörg Becker</i>	1781
Coupling Metrics for EPC Models <i>Daniel Braunnagel, Florian Johannsen</i>	1797
Information Systems Success - A Quantitative Literature Review and Comparison <i>Sebastian Dörr, Sebastian Walther, Torsten Eymann</i>	1813
Knowledge Creation in Requirements Engineering – A Systematic Literature Review <i>Linda Schneider, Kerim Hajji, Andreas Schirbaum, Dirk Basten</i>	1829
Developing an Organizational Culture Supportive of Business Process Management <i>Sanja Tumbas, Theresa Schmiedel</i>	1845
IT-basierte Unterstützung virtueller und realer Selbsthilfegemeinschaften in Katastrophenlagen <i>Oliver Heger, Christian Reuter</i>	1861
Unterstützung mobiler Geo-Kollaboration zur Lagebeurteilung von Feuerwehr und Polizei <i>Christian Reuter, Michael Ritzkatis</i>	1877
Entwicklung neuer Services in Lehre und Praxis – ein Abgleich von personellem Angebot und Nachfrage <i>Peer Benholz, Helena Preiß</i>	1893

Business Process and Service Management

Jörg Becker¹ und Jan Mendling²

¹ Westfälische-Wilhelms-Universität, Münster, Germany

² Wirtschaftsuniversität, Wien, Austria

Aufgrund des Prozesscharakters von Dienstleistungen kommt der Integration von Geschäftsprozess- und Servicemanagement eine zentrale Bedeutung zu. Insbesondere setzt die erfolgreiche Etablierung serviceorientierter Geschäftsmodelle die Konzeption und Implementierung geeigneter Geschäftsprozesse zur Entwicklung, Vermarktung und Erbringung von Dienstleistungen voraus. Zur Erfüllung ggf. individueller Kundenanforderungen sind Geschäftsprozesse dabei unternehmensintern und unternehmensübergreifend zu gestalten, zu optimieren und zu integrieren, damit Ressourcen und Kernkompetenzen bedarfsgerecht zusammengeführt werden können.

Zielsetzung des Tracks Geschäftsprozess- und Servicemanagement ist es, ein Forum bereitzustellen, in dem Methoden des Geschäftsprozessmanagements im Hinblick auf ihre Eignung zur Unterstützung serviceorientierter Geschäftsmodelle hin untersucht und weiterentwickelt werden. Aus Sicht des Servicemanagements werden Anforderungen serviceorientierter Geschäftsmodelle an die Analyse und Gestaltung von Geschäftsprozessen identifiziert. Neben der funktionsübergreifenden Integration im serviceorientierten Unternehmen werden insbesondere auch Aspekte der unternehmensübergreifenden Vernetzung betrachtet. Hierbei stellen sich weiter führende Fragen der Vernetzungsstrategien, Arbeitsteilung und der prozessualen Gestaltung. Aus Sicht des Geschäftsprozessmanagements werden Methoden und Modelle entwickelt, die diese Anforderungen im Hinblick auf strategische, analytische, gestaltungsorientierte und implementierungsspezifische Aspekte erfüllen und somit zur weiteren Etablierung serviceorientierter Geschäftsmodelle beitragen. Hierbei stellen sich Fragen der ganzheitlichen integrierten Betrachtung sowie der Umsetzung und Etablierung des Geschäftsprozessmanagementlebenszyklus.

Die 29 eingereichten Beiträge spiegeln die insgesamt breite Ausrichtung der betrachteten Themenfelder wider. Eine aus unserer Sicht weitere interessante Beobachtung ist, dass mit 18 deutschsprachigen Beiträgen und elf englischsprachigen Beiträgen – alle Beiträge stammen aus dem deutschsprachigen Raum – eine gewisse Tendenz zur Einreichung englischsprachiger Artikel erkennbar ist.

Mit den insgesamt 6 angenommenen Beiträgen konnten beide im Track fokussierten Themengebiete in etwa gleichgewichtig berücksichtigt werden. Im Themenfeld Geschäftsprozessmanagement stellen Weber, Muntermann und Wondrak ein Konzept zur Messung des Wertes von Informationssystemen vor, während Kittel die Eigenschaften von Geschäftsprozessen im Spannungsfeld zwischen Agilität und Compliance untersucht. Im Themenfeld Service Management präsentiert Johannsen einen Ansatz zur Integration von Methoden für das Qualitätsmanagement. Kieninger, Straeten, Kimbrough, Schmitz und Satzger stellen eine Methode zur Analyse von Service Incidents vor, auf deren Basis Dienstleistungsangebote kostenoptimal ausgewählt

werden können. Bosse, Splieth und Turowski entwickeln ein Vorhersagemodell für die Verfügbarkeit von IT-Services. Schrödl, Geier, Geier und Simkin erweitern Modellierungstechniken zur Abbildung von Risiken in Liefernetzwerken der hybriden Wertschöpfung.

Vor dem Hintergrund der Zielsetzung des Tracks als Zusammenführung von Geschäftsprozess- und Servicemanagement erscheint die Ausrichtung der Beiträge insgesamt fokussiert auf spezifische Themen in einer der beiden Betrachtungsebenen. Hier kann für die weitere Forschung ein darüber hinausgehender Bedarf an einer Untersuchung von Themen vermutet werden, die beide Perspektiven gezielt weiter zusammenführen. Dies gilt für eine Erweiterung von Prozessmodellierungssprachen im Dienstleistungsbereich genauso wie für die Implementierung, Analyse und Optimierung von Geschäftsprozessen in Service Systemen.

Unser besonderer Dank gilt den 21 Mitgliedern im Programmkomitee sowie den Gutachtern, die ihre Expertise in insgesamt 87 Gutachten zur Bewertung und weiteren Verbesserung der Artikel zur Verfügung gestellt haben.

Programmkomitee

Daniel Beverungen, Westfälische Wilhelms-Universität Münster

Tilo Böhmman, Universität Hamburg

Peter Buxmann, Technische Universität Darmstadt

Helmut Krcmar, Technische Universität München

Christine Legner, Université de Lausanne

Susanne Leist, Universität Regensburg

Peter Loos, Universität Saarbrücken

Oliver Müller, Universität Liechtenstein

Markus Nüttgens, Universität Hamburg

Andreas Oberweis, Karlsruher Institut für Technologie

Jens Pöppelbuß, IFIB, Universität Bremen

Michael Räckers, Westfälische Wilhelms-Universität Münster

Jan Recker, Queensland University of Technology

Manfred Reichert, Universität Ulm

Stefanie Rinderle-Ma, Universität Wien

Michael Rosemann, Queensland University of Technology

Gerhard Satzger, Karlsruher Institut für Technologie

Oliver Thomas, Universität Osnabrück

Jan vom Brocke, Universität Liechtenstein

Mathias Weske, Hasso Plattner Institut, Universität Potsdam

Michael zur Muehlen, Stevens Institute of Technology

Vorhersagemodell für die Verfügbarkeit von IT-Services aus Anwendungssystemlandschaften

Sascha Bosse, Matthias Splieth, und Klaus Turowski

Otto-von-Guericke-Universität, Magdeburger Research and Competence Cluster for Very Large Business Applications, Magdeburg, Germany
{sascha.bosse,matthias.splieth,klaus.turowski}@ovgu.de

Abstract. Serviceorientierung und Cloud Computing haben dazu geführt, dass sich inzwischen immer mehr Unternehmen darauf spezialisieren, IT-Dienstleistungen für Organisationen anzubieten. Erhöhter Wettbewerb in dieser Branche führt zu Kosten- und Qualitätsdruck, was eine bessere Beherrschung des damit einhergehenden Leistungserstellungsprozesses durch quantifizierbare Vorhersagen für Anpassungen an den Anwendungssystemlandschaften der Service-Anbieter wünschenswert macht. In diesem Beitrag wird ein neuer Ansatz entwickelt, um die Verfügbarkeit von IT-Services aus Anwendungssystemlandschaften vorherzusagen. Entgegen bisher bekannter Lösungen geht dieser Ansatz nicht davon aus, dass die einzelnen Systemkomponenten unabhängig voneinander ausfallen. Es wird gezeigt, dass der Lösungsansatz ohne diese Abhängigkeiten die Ergebnisse analytischer Methoden erreicht. Somit können Entscheidungsprozesse unterstützt werden.

Keywords: Anwendungssystemlandschaft, Verfügbarkeit, IT-Service

1 Einleitung

Die effektive und effiziente Nutzung von Informationstechnologie (IT) spielt eine immer entscheidendere Rolle für den Erfolg von Unternehmen. Neben der klassischen Strategie, Geschäftsprozesse durch eine betriebseigene IT-Abteilung zu unterstützen, ist auch das Outsourcing der IT zu einem externen Dienstleister in den Fokus gerückt [1]. Unternehmen, die sich auf das Anbieten solcher IT-Dienstleistungen bzw. IT-Services spezialisiert haben, können unter dem Begriff IT-Service-Industrie zusammengefasst werden.

Das IT-Outsourcing hat Vor- und Nachteile [1]: Wichtigste Motivation ist in der Regel die Vermeidung von Fixkosten und das Weiterreichen von Risiken. Größter Nachteil des IT-Outsourcing ist der Kontrollverlust, sei es bspw. bezüglich sensibler Daten, Ausfallsicherheit oder Performanz. Um diese Nachteile zu egalisieren, werden so genannte Service Level Agreements (SLAs) als Verträge zwischen Service-Anbieter und Kunde geschlossen [2]. Sie enthalten unter anderem Garantien für die Servicequalität (engl. Quality of Service – QoS).

Um konkurrenzfähig zu sein und zu bleiben, muss ein Dienstanbieter zwischen den Kosten der Service-Bereitstellung und deren Qualität abwägen. Nicht zuletzt aufgrund der steigenden Energiepreise und der Nachhaltigkeitsdiskussion werden verschiedene Ansätze erforscht, die Dienstleistung durch Anwendungssystemlandschaften effizienter zu gestalten. Unter anderem sind dies Virtualisierungstechniken, intelligente Lastverteilung oder verbesserte Netzwerkstrukturen [3-5]. Für einen Service-Anbieter stellt sich dabei die Frage, ob diese Ansätze zu geringeren Kosten führen und die QoS gehalten oder sogar verbessert werden kann. Vorhersagemodelle könnten diese Entscheidungen unterstützen.

Während die Qualität von IT-Systemen traditionell eng an die Performanz der angebotenen Funktionen geknüpft ist, geraten nun auch andere nicht-funktionale Eigenschaften in den Fokus: Neben Wartbarkeit und Skalierbarkeit betrifft dies vor allem die Verfügbarkeit von Systemen [6]. Nach einer Untersuchung der *Aberdeen Group* von 134 Unternehmen mit Rechenzentren im Jahre 2012 erleiden diese durchschnittlich über eine Million US-Dollar an Umsatzeinbußen aufgrund von IT-Ausfällen [7]. Meldungen über Rechenzentrumsausfälle sind selbst bei spezialisierten Dienstleistern wie *Amazon.com* immer wieder aktuell [8-9].

1.1 Ziele und Methodologie

Ziel dieser Arbeit ist, ein Vorhersagemodell für die Verfügbarkeit von Services aus Anwendungssystemlandschaften auf Basis einer Diskreten-Ereignis-Simulation zu entwickeln. Dieses Modell würde es Betreibern ermöglichen, die Servicequalität bezüglich der Verfügbarkeit bei Änderungen oder dem Neuaufbau von Systemlandschaften zu bestimmen und damit die Entscheidungsfindung zu unterstützen. Diese Anwendungssystemlandschaften sind äußerst komplexe Gebilde, die sich durch eine hohe Dynamik auszeichnen [10]. Generelle Aussagen über Änderungseffekte sind in der Regel schwierig bis unmöglich. Dies führt unter anderem dazu, dass geplante Kosteneinsparungen geringer als erwartet ausfallen oder sogar nicht eintreten [11].

Die Entwicklung des Vorhersagemodells folgt einem gestaltungsorientierten Forschungsdesign, vgl. z. B. [12]. Das entstehende Artefakt besteht dabei aus einem Modell und Algorithmen zur Vorhersage. Diese erfolgt quantitativ durch die Simulation. Durch die hohe Komplexität der untersuchten Anwendungssystemlandschaften und die spätere Erweiterung auf andere Eigenschaften ist diese Methode geeignet, bei formal beschreibbaren, aber schwierig analytisch zu lösenden Problemen die nötige Flexibilität zu bieten [13]. Evaluert wird das Artefakt durch einen Vergleich mit einer analytischen Methode an einem ausgewählten Anwendungsbeispiel.

1.2 Struktur des Beitrags

Nachdem im ersten Abschnitt auf Motivation und Zielsetzung des Forschungsbeitrags eingegangen wurde, wird in Abschnitt 2 das Forschungsgebiet abgegrenzt. Dazu werden Definitionen des Begriffs Anwendungssystemlandschaft und der Größe Verfügbarkeit erörtert.

Anschließend wird in Abschnitt 3 der Stand der Technik bei der Vorhersage von Verfügbarkeit von Systemen beleuchtet, bevor in Abschnitt 4 ein eigener Lösungsansatz präsentiert wird.

Dieser Lösungsansatz wird in Abschnitt 5 auf das Beispielszenario eines international agierenden *SAP Application Service Provider (ASP)* angewendet. Auf Basis dieses Szenarios erfolgt dann eine Verifikation des vorgestellten Ansatzes. In Abschnitt 6 werden die Ergebnisse des Beitrags zusammengefasst und ein Ausblick auf weitere Forschungsarbeiten gegeben.

2 Definitionen

2.1 Anwendungssystemlandschaft

Eine Anwendungssystemlandschaft (auch IT-Systemlandschaft oder IT-Landschaft) ist ein System von Systemen, die miteinander integriert sind [14-15]. Sie ist in der Regel historisch gewachsen, komplex und besteht aus heterogenen Einzelteilen [10], [16]. Die Systeme der Landschaft werden von Menschen mit unterschiedlichen Zielen benutzt [15]. Neben Anwendungssystemen enthält sie alle benötigten Unterstützungssysteme [10].

Der Begriff *Enterprise Architecture* wird als Obermenge der Anwendungssystemlandschaft gesehen [14]. Diese bietet eine ganzheitliche Sicht auf ein Unternehmen, meistens Ebenen-orientiert. Diese reichen von der Ebene der IT-Infrastruktur über die betrieblichen Anwendungen, deren Integration untereinander bis zur Ebene der Geschäfts- bzw. Organisationssicht und der strategischen Ebene [15]. Auf jeder Ebene interagieren Menschen [17].

Eine Anwendungssystemlandschaft ist dann eine Teilmenge bzw. die Umsetzung einiger Teile dieser Enterprise Architecture. Neben den Anwendungen und deren Integration beschreibt die Anwendungssystemlandschaft alle nötigen Infrastruktursysteme. Auch enthalten ist die Nutzung der Systeme durch Menschen, um bestimmte Geschäftsprozesse zu unterstützen.

2.2 Nicht-funktionale Eigenschaften – Verfügbarkeit

Funktionale Eigenschaften eines Systems adressieren die Funktionen oder das Verhalten, die das System anbieten soll. Nicht-funktionale Eigenschaften beschreiben dabei alle qualitativen Aspekte eines Systems mit Ausnahme der funktionalen [18]. Leitfaden zur Aufstellung von qualitativen Eigenschaften wie z. B. die ISO 9126 (in der ISO 25000 2005 aufgegangen) basieren dabei auf dem so genannten *Factor Criteria Metrics (FCM) Model* [19]. Dabei werden zu Qualitätsfaktoren wie Effizienz oder Übertragbarkeit Qualitätskriterien definiert. Im Fall Übertragbarkeit können diese bspw. Anpassbarkeit oder Ersetzbarkeit bedeuten. Im letzten Schritt muss für jedes Kriterium eine Metrik definiert werden, um die Erfüllung quantitativ messen zu können.

Verfügbarkeit ist nach diesem Modell ein Qualitätskriterium, welches dem Faktor Zuverlässigkeit zugeordnet ist. Gemessen werden kann sie z. B. nach [20] mit der normierten Betriebsdauer, die sich aus dem Verhältnis der tatsächlichen Betriebsdauer zu der geplanten Betriebsdauer ergibt.

Eine probabilistische Definition aus [21] definiert Verfügbarkeit hingegen als die „Wahrscheinlichkeit, dass ein System zu einem bestimmten Zeitpunkt in Betrieb und in der Lage ist, die angeforderten Dienste zu liefern“. Hierbei ist der Fokus statt auf die absolute Zeit eher auf die Nutzungsanfrage gesetzt. In [22] werden diese beiden Definitionen verbunden, um auf der einen Seite die Verfügbarkeit aus historischen Daten zu berechnen und sie auf der anderen Seite in Zukunft schätzen zu können. Alle Definitionen gehen dabei davon aus, dass die geplante Betriebszeit entscheidend ist, also geplante Wartungen aus der Berechnung heraus genommen werden. Auch in dieser Arbeit werden beide Definitionen berücksichtigt.

3 Stand der Technik

Vorhersagemodelle für Verfügbarkeit wurden bisher u. a. in der Softwareentwicklung eingesetzt. Bei der komponentenorientierten Softwareentwicklung hängt die Verfügbarkeit des Gesamtsystems von den einzelnen Komponenten ab. In [23] wird eine Literaturanalyse präsentiert, die den Stand der Technik bei der Vorhersage der Verfügbarkeit eines Softwareprodukts in einer frühen Entwurfsphase untersucht. Die gefundenen Ansätze werden in drei Kategorien klassifiziert: Zustandsbasiert- oder additiv-analytisch sowie pfadbasiert-quantitativ Modelle. Letztere erfordern zumindest eine experimentelle Datenerfassung, während die analytischen Ansätze ausnahmslos vom unabhängigen Ausfall der Komponenten ausgehen. Dies wird als signifikanter Schwachpunkt aller bekannten Ansätze identifiziert. Die vorgeschlagene Lösung dieses Problems ist die Erweiterung der Markov-Eigenschaft auf höhere Grade, was jedoch Modellierung und Parametrisierung schnell komplex werden lässt [24].

Neben Software sind auch verteilte Systeme Gegenstand von Vorhersagemodellen für die Verfügbarkeit. In [25] wird z. B. eine Vorhersage für ein OSCAR¹-Cluster vorgenommen. Hierbei werden drei Ebenen modelliert, die für die Verfügbarkeit entscheidend sind. Neben der Serverebene, auf der Anfragen angenommen werden, gibt es die Netzwerkebene für die Kommunikation und die Clientebene für die angefragte Berechnung. Jede Ebene enthält redundante fehlerbehaftete Komponenten. Bei einem Ausfall der primären Komponenten übernehmen nach kurzer Zeit Standby-Komponenten den Betrieb. Zur Modellierung des Systemverhaltens werden für jede Ebene so genannte *Stochastic Reward Nets* (SRN) verwendet.

Dieses Modell wird in [26] eingeführt. Zwar erlauben kombinatorische Modelle wie Fehlerbäume eine schnelle Auswertung, aber die Abbildung von Abhängigkeiten gestaltet sich schwierig. Zustandsbasierte Modelle wie Markov-Ketten auf der anderen Seite leiden häufig unter der so genannten Zustandsexplosion, welche die Hand-

¹ Open Source Cluster Application Resources

habbarkeit dieser Modelle stark verringert. Eine Möglichkeit, diesen Nachteil zu verringern, bieten Stochastische Petri-Netze, die eine automatische Erzeugung des Zustandsgraphen ermöglichen. Der Zustand des Systems wird dabei über die Kombination der markierten Stellen repräsentiert, während zeitbehaftete Transitionen die Markierung des Netzes verändern.

Eine Erweiterung dieser Netze stellen die so genannten *Generalisierten Stochastischen Petri-Netze* (GSPN) dar. Hierbei wird das SPN um Konstrukte erweitert, die einzelne Transitionen bei bestimmten Systemzuständen deaktivieren können. Weiterhin sind auch zeitlose Transitionen möglich. Verschiedene Farben von Marken ermöglichen die Zusammenfassung strukturell gleicher Netze [27].

Ein SRN ist dann ein GSPN mit *Reward Functions*. Diese werden unterteilt in *Rate Rewards* und *Impulse Rewards*. Erstere verändern Variablen, wenn eine bestimmte Stelle markiert ist, während Letztere Variablen bei Ereignissen verändern. Das Systemverhalten kann dann in einem GSPN modelliert werden, während die Veränderungen der Verfügbarkeit durch Rewards berechnet wird [26].

Die Vorhersage der Verfügbarkeit von IT-Diensten wird unter anderem in [28] adressiert. Dort werden existierende Verfahren in analytische, qualitative und quantitative Methoden eingeteilt. Petri-Netze und ihre Erweiterungen werden als analytische Verfahren klassifiziert, auf die Lösung dieser Netze durch Simulation wird nicht eingegangen. Quantitative Ansätze setzen hier ebenfalls eine Datenerfassung aus laufenden Systemen voraus. Sie sind daher auf hypothetische Systeme nicht übertragbar.

Für den eigenen, quantitativen Lösungsansatz wird ebenfalls angenommen, dass die Verfügbarkeit eines zusammengesetzten Systems aus den einzelnen Komponenten berechnet werden kann. Die zunächst vorgestellten Ansätze aus der Softwareentwicklung gehen dabei ohne Ausnahme davon aus, dass diese Komponenten unabhängig ausfallen [23]. In den Ansätzen für die Verfügbarkeit verteilter Systeme werden unter anderem Stochastic Reward Nets verwendet. Die entwickelten Modelle sind nicht auf beliebige Systeme übertragbar, die Modellierungsmethode jedoch geeignet, auch Abhängigkeiten zwischen den einzelnen Komponenten abzubilden [27]. Die analytische Lösung dieser Netze wird jedoch durch das Problem der Zustandsexplosion erschwert [28]. Aus diesem Grund werden die Netze in diesem Beitrag nicht analytisch, sondern durch eine Diskrete-Ereignis-Simulation mit anschließender Datenanalyse gelöst.

4 Lösungsvorschlag

In einem Experteninterview bei einem international agierenden SAP ASP wurde festgestellt, dass vier Ereignisse zu einer Nichtverfügbarkeit eines Services führen können: Fehler in einem Rechenzentrum bei der Bearbeitung des Dienstes, Fehler in der Infrastruktur um das Rechenzentrum, menschliche Fehler bei der Administration sowie Wartungsarbeiten. Da letztere zu geplanten Downtimes zählen, die bei der Berechnung der Verfügbarkeit ausgeblendet werden [20], müssen nur die ersten drei Ereignisse betrachtet werden. In einer Datensammlung über Serverausfälle von 1996 bis 2005, bereitgestellt vom Los Alamos National Lab, werden analog zu diesem

Interview Hardware-, Software- und Infrastrukturfehler sowie menschliche Fehler unterschieden [29]. Bei der Beschreibung von einzelnen Ressourcen können weiterhin transiente, intermittierende und permanente Fehler unterschieden werden [28]. Jede dieser Fehlerarten zeichnet sich dabei durch unterschiedliche Ausfall- oder Reparaturzeiten aus.

In Abbildung 1 sind die relevanten Schichten der Enterprise Architecture für das Problem, die Verfügbarkeit von Services aus Anwendungssystemlandschaften zu bestimmen, auf Basis des Schichtenmodells aus [17] dargestellt. Die Bereitstellung eines IT-Dienstes erfordert dabei, dass bestimmte Komponenten einer Anwendungssystemlandschaft intakt sind. Die Verfügbarkeit des angebotenen Service hängt dann von diesen Komponenten ab, denen jeweils für jede Fehlerart ein Wert *Mean Time to Failure* (Mittlere Zeit bis zum Fehler - MTTF) und *Mean Time to Recover* (Mittlere Zeit zur Reparatur - MTTR) zugeordnet ist. Die Annahme ist dabei, dass Ausfall- und Reparaturzeiten exponentialverteilt sind. Formal ist eine Komponente c dann ein 6-Tupel.

$$c = (MTTF_i, MTTR_i, MTTF_p, MTTR_p, MTTF_s, MTTR_s) \quad (1)$$

Menschen administrieren diese Komponenten und können dabei Fehler erzeugen. Dies wird hier der Vollständigkeit halber erwähnt, aber im Modell aufgrund schwieriger Abstraktion zunächst nicht aufgenommen.

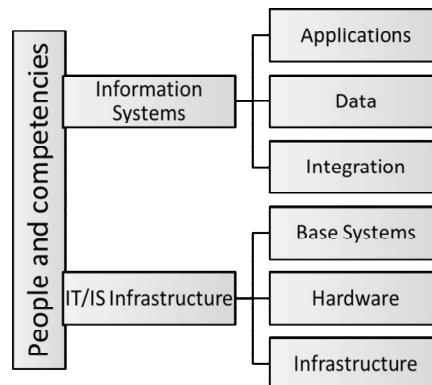


Abb. 1. Relevante Schichten für die Vorhersage der Verfügbarkeit von Services aus Anwendungssystemlandschaften basierend auf [17]

Eine Komponente kann dabei den zwei Schichten Informationssysteme und IT/IS Infrastruktur zugeordnet werden, wobei diese noch weiter unterteilt werden können: Anwendungen (z. B. ein ERP-System) mit Daten- und Integrationskomponenten, Unterstützungs- (z. B. Betriebssysteme), Hardware- (z. B. eine Festplatte) und Infrastruktursysteme (z. B. Stromversorgung).

Dabei wird davon ausgegangen, dass nur Komponentenklassen betrachtet werden, die für die Bereitstellung des Dienstes notwendig sind. Eine Komponentenklasse kann jedoch mehrfach realisiert werden, so dass sie dann redundant vorhanden ist. Im

nächsten Schritt können nun die Komponenten und deren Zusammenspiel modelliert werden.

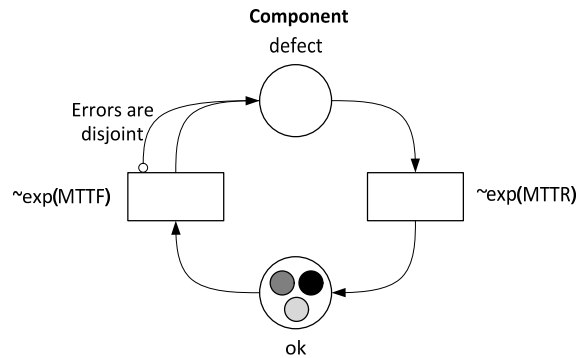


Abb. 2. Farbiges GSPN einer einzelnen Komponente

Einzelne Komponenten werden als farbiges GSPN modelliert (Abbildung 2). Dieses besteht aus zwei Stellen, die den jeweiligen Zustand „defekt“ oder „arbeitet“ widerspiegeln. Dieser Kreislauf enthält drei farbige Marken, die den jeweiligen Fehlerarten zugeordnet sind. Die beiden Transitionen zwischen diesen Stellen schalten nach exponentialverteilter Dauer mit Erwartungswert MTTF beziehungsweise MTTR. Dabei wird angenommen, dass das Auftreten von Fehlerarten disjunkt ist, daher wird die Transition nach „defekt“ durch eine Inhibitor-Kante für die anderen Fehlerarten deaktiviert. Der Vereinfachung halber wird in den folgenden Abbildungen auf die Berücksichtigung von Farben verzichtet.

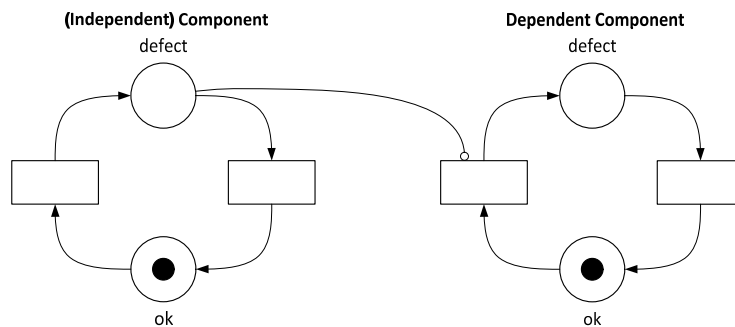


Abb. 3. Zwei Komponenten mit Abhängigkeitsbeziehung

Ist der Ausfall einer Komponente nur möglich, wenn eine andere Komponente arbeitet (z. B. Softwarefehler bei Hardwareausfall), wird dies ebenfalls über eine Inhibitor-Kante modelliert. Zwei Komponenten und eine solche Ausfallabhängigkeitsbeziehung sind in Abbildung 3 dargestellt. Die Transition zu „defect“ in der abhängigen Kom-

ponente ist genau dann deaktiviert, wenn die Stelle „defect“ in der unabhängigen Komponente markiert ist.

Eine weitere mögliche Abhängigkeit zwischen zwei Komponenten ist die so genannte Standby-Redundanz, bei der die redundante Komponente erst aktiv wird, wenn die Primärkomponente ausfällt. Dieser Vorgang nimmt eine bestimmte Übernahmzeit in Anspruch [25]. Erst nach dieser Zeit, wenn sich eine Marke in „Tookover“ befindet, ist die Sekundärkomponente in Betrieb und kann ausfallen. Ein solches Paar von Komponenten wird dann wie in Abbildung 4 modelliert.

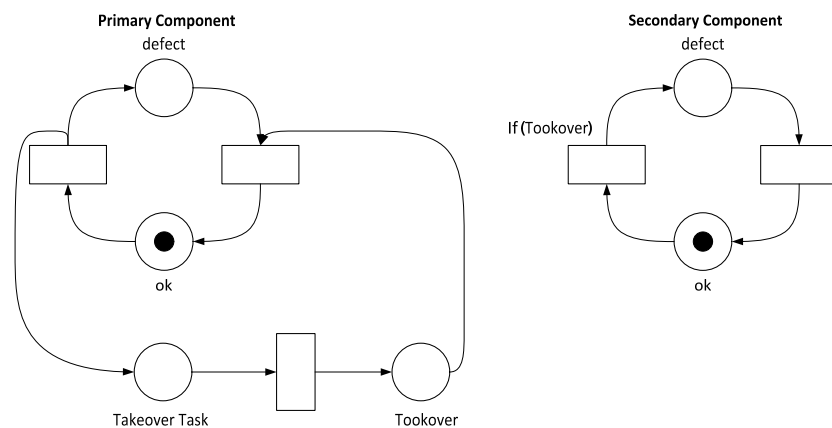


Abb. 4. Modell für Standby-Redundanz zweier Komponenten

Mit diesen Möglichkeiten können nun einzelne Komponenten sowie deren Abhängigkeiten beim Ausfall modelliert werden. Diese Modelle sowie die Beziehungen dieser Komponenten zur Bereitstellung des Services werden dann in einem erweiterten Zuverlässigkeitsblattschaltbild (engl. Reliability Block Diagram - RBD) abgebildet.

Das *Reliability Block Diagram* ist ein Modell, das entwickelt wurde, um militärischen Missionserfolg in Teilerfolge zu zerlegen [30]. Später entwickelte es sich zu einem Standardmodell bei der Verfügbarkeitsanalyse von zusammengesetzten Systemen [26]. Komponenten werden dabei in einem Graphen seriell oder parallel verbunden. Dabei ist das System online, wenn auf einem möglichen Pfad vom Wurzelknoten zu einem Blattknoten kein Fehler vorhanden ist. Das *Extended Reliability Block Diagram* (ERBD) enthält zusätzlich Ausfallabhängigkeiten zwischen den einzelnen Komponenten sowie standby-redundante Komponenten. Ein Beispiel für ein solches Diagramm ist in Abbildung 5 dargestellt.

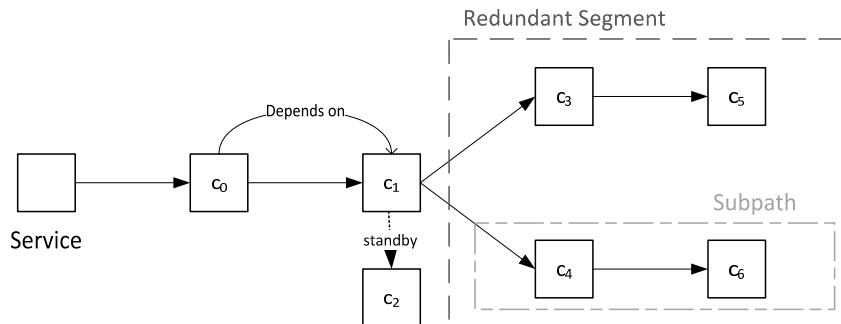


Abb. 5. Beispiel eines Extended Reliability Block Diagrams

Komponenten, die bei alleinigem Ausfall den Erfolg der Bereitstellung verhindern, werden als kritisch bezeichnet (c_0 und c_1). Gibt es zwei oder mehr aktive Komponenten, die die gleiche Funktion erfüllen, werden diese als „(parallel-)redundant“ bezeichnet (c_3 und c_4 sowie c_5 und c_6). Zwischen den Komponenten c_0 und c_1 ist eine Ausfallabhängigkeitsbeziehung dargestellt. Weiterhin übernimmt die standby-redundante Komponente c_2 bei einem Ausfall von c_1 .

Ein Teilgraph, der mehrere seriell verbundene Knoten enthält, wird als Teilpfad (Subpath) bezeichnet. Zwei oder mehrere Teilpfade, die jeweils die gleichen Komponententypen enthalten und unabhängig sind, werden insgesamt als redundantes Segment bezeichnet. Jedes redundante Segment ist somit kritisch.

Ist das ERBD aufgestellt, kann daraus ein Petri-Netz für alle Komponenten und deren Beziehungen erzeugt werden. Sobald sich der Zustand einer Komponente ändert, wird mithilfe des ERBD der Zustand des Gesamtsystems bestimmt. Das System ist ausgefallen, genau dann wenn

- Eine kritische Komponente ausgefallen ist und (noch) keine standby-redundante Komponente übernommen hat oder
- In einem redundanten Segment alle Teilpfade ausgefallen sind.

Ein Teilpfad kann dabei als eigenes System aufgefasst werden. Diese Rekursion endet, sobald ein Teilpfad nur aus (auf ihn bezogen) kritischen Komponenten besteht.

Nachdem der Zustand des Systems nun in einem Modell erfasst ist, muss noch die Nutzung des Systems modelliert werden. Das Nutzungsmodell definiert die Verwendung des im ERBD modellierten Systems als SRN. Es ist in Abbildung 6 dargestellt. Die Zeit zwischen zwei Anfragen wird als exponentialverteilt mit Mittelwert *Mean Time to Request* (MTTReq) angenommen, die Anfragen dann in der Stelle „Service Request“ gesammelt. Die Stellen „SystemUp“ und „SystemDown“ geben den aktuellen Zustand des Systems wider, wobei die zeitlose Transition „ t_{down} “ genau dann aktiviert ist, wenn das System im ERBD nicht verfügbar ist. Andernfalls ist „ t_{up} “ aktiviert.

Ist ein Service Request vorhanden und das System verfügbar, wird die Marke nach einer zufälligen Bearbeitungszeit vernichtet und eine Marke in der Stelle „Request

Served“ erzeugt. Fällt das System während dieser Bearbeitung aus, so werden alle Marken in „Service Request“ vernichtet und entsprechend Marken in der Stelle „Request not Served“ erzeugt.

Verschiedene Rewards können genutzt werden, um unterschiedliche Metriken für die Verfügbarkeit zu berechnen. Wenn die Verfügbarkeit anfrageorientiert geschätzt werden soll, kann ein Impulse Reward bei Eintritt einer Marke in „Request Served“ oder „Request not Served“ wie folgt berechnet werden:

$$\text{Availability}_1 = \# \text{RequestServed} / (\# \text{RequestServed} + \# \text{RequestNotServed}) . \quad (2)$$

Dieser Wert kann dann als Schätzer für die Wahrscheinlichkeit, dass eine Anfrage nicht vom System beantwortet werden kann, angenommen werden.

Eine weitere Möglichkeit ist, die Verfügbarkeit mittels der normierten Betriebsdauer zu berechnen. Dies kann durch einen Rate Reward geschehen:

$$\text{Availability}_2 = t_{\text{SystemUp}} / t_{\text{total}} \quad (3)$$

Dabei wird die Variable t_{SystemUp} mit einer Rate von 1 erhöht, falls „SystemUp“ markiert ist. Die Variable t_{total} gibt die Simulationszeit an.

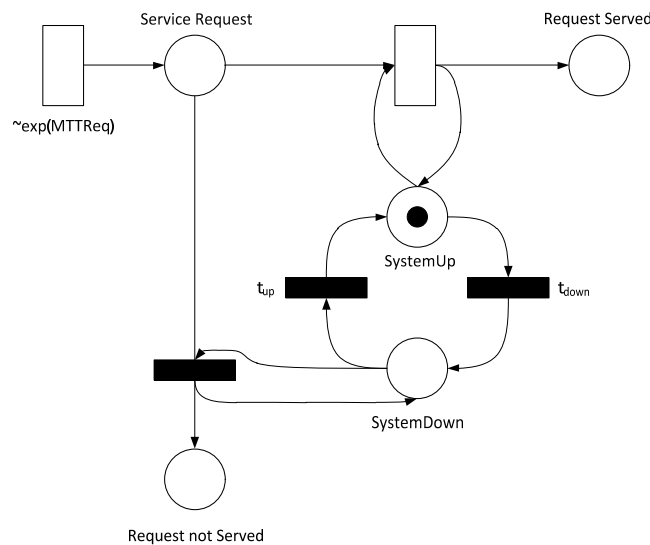


Abb. 6. Nutzungsmodell

5 Verifikation am Beispielszenario eines SAP ASP

Der in Abschnitt 4 vorgestellte Lösungsansatz wird in zwei Schritten verifiziert. Zunächst wird am Beispiel eines international agierenden SAP ASP gezeigt, dass eine reale Systemlandschaft im entwickelten Modell abbildbar ist. Das zugehörige RBD ist in Abbildung 7 dargestellt.

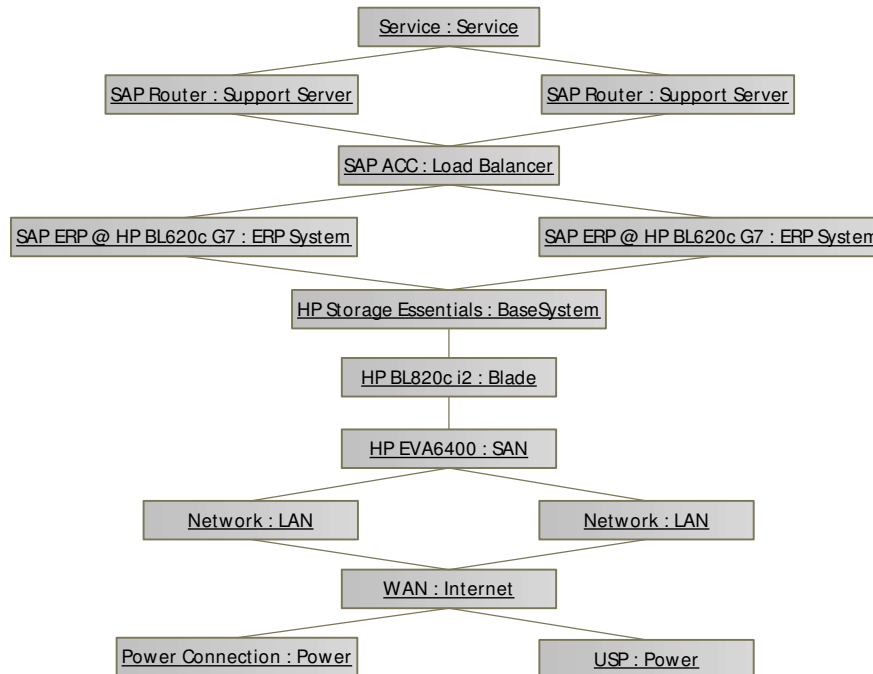


Abb. 7. Beispielszenario als Reliability Block Diagram

Den Kern des Beispielszenarios bildet ein IT-Service, der durch zwei Anwendungssysteme (SAP ERP 6.0) bereitgestellt wird. Eine Anfrage an einen Service wird zunächst von zwei redundanten Routern verarbeitet, welche diese an eines der Anwendungssysteme weiterleitet. Beide ERP-Systeme werden auf je einem Blade betrieben. Beide greifen auf das gleiche Datenbanksystem zu, welches auf einem weiteren Blade betrieben wird und die Datenhaltung auf einem Storage Area Network (SAN) übernimmt.

Neben diesen Hard- und Software-Entitäten sind weitere Komponenten Bestandteil des Beispielszenarios: Die Anwendungen HP Storage Essentials und SAP Adaptive Computing Controller, betrieben auf je einem Blade, verwalten bestimmte Teile der Infrastruktur. Weiterhin sind ein redundantes Netzwerk sowie eine Anbindung an das Internet (nicht redundant) und eine Stromversorgung Teil des Beispielszenarios. Zur Sicherung des Betriebs im Falle eines Stromausfalls ist zudem eine unterbrechungsfreie Stromversorgung (USP) vorhanden.

Im zweiten Schritt wird gezeigt, dass der Lösungsansatz plausible Ergebnisse produziert. Da zum Zeitpunkt der Erstellung dieser Arbeit keine Daten zu den Ausfallzeiten seitens des SAP ASP vorlagen, wurde für die Bestimmung der Größen MTTF und MTTR der jeweiligen Komponenten auf den Datenbestand des Los Alamos National

Lab [29] zugegriffen. Softwarefehler wurden für dieses Beispielszenario nicht berücksichtigt, da Daten über Ausfälle der ERP-Systeme nicht vorlagen.

Um den entwickelten Lösungsvorschlag zu verifizieren, wurde das Beispielszenario zunächst nur als klassisches RBD ohne Ausfallabhängigkeiten oder Standby-Redundanzen modelliert. Weiterhin wurde nur eine Fehlerart betrachtet. In diesem Fall kann die Verfügbarkeit des Gesamtsystems analytisch aus den einzelnen Verfügbarkeiten der Komponenten berechnet werden, welche wiederum aus MTTF und MTTR berechnet werden können [31]:

$$A_c = \text{MTTF} / (\text{MTTF} + \text{MTTR})$$

$$A_{\text{seriell}} = A_1 * A_2 \tag{4}$$

$$A_{\text{parallel}} = 1 - (1 - A_1)(1 - A_2)$$

Nachdem diese Formeln auf das Beispielszenario angewendet wurden, wurde der ermittelte Wert mit den Simulationsergebnissen, basierend auf dem Komponentennetz und dem RBD, verglichen. In Abbildung 8 sind die ermittelten Konfidenzintervalle und Mittelwerte für die Verfügbarkeit nach Formel 3 in Abhängigkeit von der Anzahl der Replikationen dargestellt. Dabei wurden 100.000 Stunden simuliert und ein Konfidenzwert von 0,99 angenommen. Der errechnete Wert 0,99311 ist zur Orientierung als gestrichelte Linie eingezeichnet.

Es ist zu erkennen, dass sich die Simulationsergebnisse bei steigender Replikationsanzahl zusehends dem errechneten Wert annähern. Damit kann mit hoher Wahrscheinlichkeit die Aussage getroffen werden, dass das entwickelte Simulationsmodell die Verfügbarkeit wie analytische Methoden schätzen kann.

Im nächsten Schritt wird für das Beispielszenario bestimmt, welchen Effekt die Einführung von Ausfallabhängigkeiten hat. Dazu wurden verschiedene Abhängigkeiten, z. B. die aller Hardwarekomponenten von der Stromversorgung oder die der Softwarekomponenten von der ausführenden Hardware, modelliert.

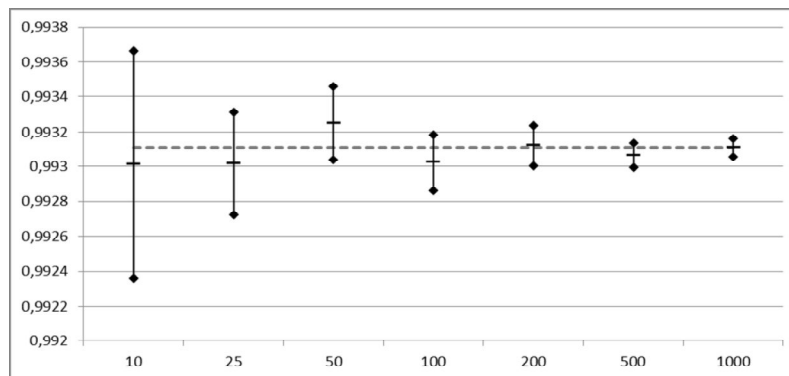


Abb. 8. Konfidenzintervalle der Verfügbarkeit in Relation zur Anzahl der Replikationen

Bei 1000 Replikationen mit 100.000 Stunden Simulationszeit ergibt sich dabei ein 0,99-Konfidenzintervall von [0,99303;0,99313] mit Mittelwert 0,99308. Dies bedeutet, dass die mittlere Verfügbarkeit hier minimal kleiner ist als die mit der Annahme unabhängiger Komponenten. Die Tatsache, dass dieser Unterschied erst in der vierten Stelle nach dem Komma auftritt, zeigt, dass der Effekt von Ausfallabhängigkeiten gering ist. In der Anwendungsdomäne IT-Services mit hohen Verfügbarkeitsgarantien kann jedoch auch ein solch kleiner Unterschied zur Verletzung eines Service Level Agreements führen.

6 Schlussfolgerungen

In diesem Artikel wurde ein Lösungsansatz zur Vorhersage der Verfügbarkeit von IT-Service mittels Diskreter-Ereignis-Simulation vorgestellt. Dafür wurde die Nutzung des IT-Services, dessen Bereitstellungsabhängigkeiten von Komponenten der Anwendungssystemlandschaft sowie der Ausfall einzelner Komponenten modelliert und in eine hierarchische Struktur eingeordnet. Der entwickelte Ansatz ermöglicht die Vorhersage der Verfügbarkeit für Systeme, in denen die Komponenten unabhängig voneinander ausfallen. Diese Fähigkeit wurde mittels eines Vergleiches mit analytischen Verfahren verifiziert. Weiterhin ermöglicht der Lösungsansatz so genannte Ausfallabhängigkeiten zwischen einzelnen Komponenten zu modellieren. Dies bedeutet, dass eine abhängige Komponente nur dann ausfallen kann, wenn die Komponente, von der sie abhängig ist, in Betrieb ist. Auch Standby-Redundanz und verschiedene Fehlerarten sind im Modell abbildbar.

Der Vergleich zwischen dem unabhängigen Ansatz und dem mit Ausfallabhängigkeiten am Beispiel eines Ausschnittes der Anwendungssystemlandschaft des SAP ASP ergab, dass die Unterschiede in der vorhergesagten Verfügbarkeit vorhanden, aber sehr gering sind. Dennoch sind gerade bei der Verfügbarkeit auch hintere Nachkommastellen wichtig, da die garantierten Niveaus oft sehr nah an 100 % liegen.

Damit stellt diese Arbeit ein generisches Modell zur Verfügung, um die Verfügbarkeit von IT-Services vorherzusagen. Durch Modularisierung und durch die freie Wahl des Abstraktionsniveaus können beliebige Anwendungssystemlandschaften abgebildet werden. Die Möglichkeit, dabei Abhängigkeiten beim Ausfall zwischen den einzelnen Komponenten abzubilden, löst das signifikante Problem bisheriger Ansätze, die unabhängige Ausfälle annehmen [23]. Durch die Lösung mit Simulation wird das Problem der Zustandsexplosion vermieden, das die Handhabbarkeit zustandsbasierter Modelle erschwert hat. Weiterhin kann das Modell beliebige Verteilungen für die Ausfallzeiten der Komponenten abbilden. Damit ist eine weitere Einschränkung der zustandsbasierten Verfügbarkeitsmodelle, die in der Regel die Markov-Eigenschaft annehmen, egalisiert. Ein Nachteil des Ansatzes ist, dass wie bei analytischen Verfahren notwendige Kenntnisse über die Struktur des untersuchten Systems vorliegen müssen. Des Weiteren kann die Lösung durch Simulation viel Zeit beanspruchen und liefert nur approximative Ergebnisse.

Die nächsten Forschungsarbeiten auf diesem Gebiet werden die Erweiterung des vorgeschlagenen Lösungsansatzes beinhalten. Bspw. könnten die Abhängigkeiten

zwischen den Komponenten um zusätzliche Relationen erweitert werden. Neben der Ausfallabhängigkeit ist z. B. denkbar, dass Komponenten bei der Wiederinbetriebnahme einer anderen Komponente eine erhöhte Fehlerwahrscheinlichkeit aufweisen (z. B. Stromspitzen nach einem Stromausfall). Auch könnten redundante Komponenten, abhängig von ihrer Auslastung, eine geringe Ausfallwahrscheinlichkeit besitzen.

Parallel dazu wird eine Datensammlung beim SAP ASP durchgeführt. Diese hat das Ziel, das entwickelte Modell und dessen mögliche Erweiterungen an einem realen Anwendungsbeispiel zu validieren und somit die Vorteile gegenüber Verfahren mit der Annahme des unabhängigen Ausfalls von Komponenten zu belegen.

Ist das Vorhersagemodell für die Verfügbarkeit validiert, könnte das Modell auch für die Vorhersage anderer nicht-funktionaler Merkmale eingesetzt werden, bspw. um die Performanz eines Systems abhängig von Ausfällen zu bestimmen. Diese Modelle können dann einen Service-Anbieter bei Entscheidungen betreffend seiner Anwendungssystemlandschaft unterstützen.

Literatur

1. Stahlknecht, P., Hasenkamp, U.: Einführung in die Wirtschaftsinformatik. Springer, Berlin Heidelberg (2002)
2. Lewis, L.: Service Level Management Definition, Architecture and Research Challenges. In: IEEE Global Telecommunications Conference, pp. 1974–1978. IEEE Computing Society (1999)
3. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Kowinski, A., Lee, G., Patterson, D.A., Rabkin, A., Stoica, I., Zaharia, M.: Above the Clouds: A Berkeley View of Cloud Computing. Technical report, Electrical Engineering and Computer Sciences, University of California at Berkeley (2009)
4. Caron, E., Roderer-Merino, L., Desprez, F., Mresan, A.: Auto-Scaling, Load Balancing and Monitoring in Commercial and Open-Source-Clouds. Technical report, Laboratoire de l'Informatique du Parallélisme, University of Lyon (2012)
5. Zhang, Q., Cheng, L., Boutaba, R.: Cloud Computing: state-of-the-art and research challenges. Journal of Internet Services and Applications 1, 7–18 (2010)
6. Hennessy, J.L.: The Future of Systems Research. Computer 32, 27–33 (1999)
7. Aberdeen Group: Datacenter Downtime - How Much Does It Really Cost?. Technical report, www.stratus.com/~media/Stratus/Files/Library/AnalystReports/AberdeenDatacenterDowntimeCost.pdf
8. Büst, R.: CloudUser.de-Artikel über Ausfälle von Amazon Web Services, <http://clouduser.de/news/aws-erneut-mit-ausfall-wieder-ein-stromausfall-wieder-in-north-virginia-schwere-sturme-sind-die-ursache-12547>
9. Günther, R.: Blog-Beitrag über Stromausfall bei Amazon und Microsoft, <http://www.rgblog.de/stromausfall-rechenzentren-amazon-microsoft-cloud-computing-europaweit-lahmgelegt/>
10. Grabski, B., Günther, S., Herden, S., Krüger, L., Rautenstrauch, C., Zwanziger, A.: Very Large Business Applications. Informatik-Spektrum 30, 259–263 (2007)
11. Hamel, G.: Disadvantages of Going Green, eHow Money Online-Magazine www.ehow.com/list_6182126_disadvantages-going-green.html
12. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75–106 (2004)

13. Wilde, T., Hess, T.: Methodenspektrum der Wirtschaftsinformatik – Überblick und Portfoliobildung. Technical report, Institut für Wirtschaftsinformatik und Neue Medien, Ludwig-Maximilians-Universität München (2006)
14. Addicks, J.S., Steffens, U.: Supporting Landscape Dependent Evaluation of Enterprise Applications. In: Multikonferenz Wirtschaftsinformatik, pp. 1815–1825. GITO, Berlin (2008)
15. Matthes, F.: Softwarekartographie. Informatik-Spektrum 31, 527–536 (2008)
16. Siedersleben, J.: SOA revisited. Wirtschaftsinformatik 49, 110–117 (2007)
17. Ahlemann, F., Stettiner, E., Messerschmidt, M., Legner, C.: Strategic Enterprise Architecture Management – Challenges, Best Practises, and Future Developments. Springer, Berlin Heidelberg (2012)
18. Balzert, H.: Lehrbuch der Softwaretechnik – Entwurf, Implementierung, Installation und Betrieb. 3rd Edition, Spektrum, Heidelberg (2012)
19. Ackermann, J., Brinkop, F., Conrad, S., Fettke, P., Frick, A., Glistau, E., Jaekel, H., Kotlar, O., Loos, P., Mrech, H., Ortner, E., Raape, U., Overhage, S., Sahm, S., Schmietendorf, A., Teschke, T., Turowski, K.: Vereinheitlichte Spezifikation von Fachkomponenten. Memorandum des Arbeitskreises 5.10.3 der GI (2002)
20. Berger, T.G.: Konzeption und Management von Service-Level-Agreements für IT-Dienstleistungen. Dissertation, ULB Darmstadt (2005)
21. Sommerville, I.: Software Engineering. 8th Edition, Pearson Education München (2007)
22. Kaiser, T.: Methodik zur Bestimmung der Verfügbarkeit von verteilten anwendungsorientierten Diensten. Herbert Utz, München (1999)
23. Goseva-Popstojanova, K., Trivedi, K.S.: Architecture-based Approach to Reliability Assessment of Software Systems. Performance Evaluation 45, 179–204 (2001)
24. Raftery, A.E.: A Model for High-order Markov Chains. Journal of the Royal Statistical Society B 47, 528–539 (1985)
25. Leangsuksun, C., Shen, L., Liu, T., Song, H., Scott, S.L.: Availability Prediction and Modeling of High Availability OSCAR Cluster. In: 5th IEEE International Conference on Cluster Computing, pp. 380–386. IEEE Computer Society (2003)
26. Muppala, J.K., Ciado, G., Trivedi, K.S.: Stochastic Reward Nets for Reliability Prediction. Communications in Reliability, Maintainability and Serviceability 1, 9–20 (1994)
27. Ciado, G., Muppala, J.K., Trivedi, K.S.: SPNP: Stochastic Petri Net Package. In: Proceedings of the 3rd International Workshop PNPM, pp. 142–151. IEEE Computer Society (1989)
28. Malek, M., Hoffmann, G.A., Milanovic, N., Brüning, S., Meyer, R., Milic, B.: Methoden und Werkzeuge zur Verfügbarkeitsermittlung. Informatik-Berichte, Humboldt-Universität Berlin (2007)
29. Los Alamos National Lab: Records of cluster node outages, workload logs and error logs, <http://cfd.usr.org/data.html> (2012)
30. Military Standard: Reliability Modeling and Prediction (MIL-STD-756B). U.S. Department of Defence (1981)
31. Shooman, M.L.: Reliability of Computer Systems and Networks – Fault Tolerance, Analysis, and Design. John Wiley & Sons, New York (2002)

Measuring IT System Value with Entity-Specific Factors Influencing Process Runtime Efficiency

Jan Muntermann¹, Moritz C. Weber², and Carola Wondrak²

¹ Georg-August-University, Göttingen, Germany
muntermann@wiwi.uni-goettingen.de

² Goethe University, Frankfurt am Main, Germany
{moweber, cwondrak}@wiwi.uni-frankfurt.de

Abstract. This empirical study explores the different factors influencing the runtime of a booking process in the middle office of a major European investment company. The study is based on mixed methods research that combines qualitative research data collection techniques with a quantitative research approach. On the basis of semi-structured interviews, we first develop and present a model of the business process and control points to assess booking runtime. We further identify relevant and IT-related factors such as process automation that may affect booking runtime. Then, we analyze a unique dataset of historical process cycles providing details on booking runtime. Our quantitative analysis of this data provides new insights into how to identify most relevant factors affecting runtime of transaction processes in the financial industry. From a methodological perspective, our study illustrates how to conduct mixed methods research in the field of process performance evaluations.

Keywords: Business Processes, Business Value of IT, Investment Management, Mixed Methods Research

1 Introduction

Dealing with information and risk is the crucial core task in the business processes of financial institutions [1]. These institutions are embedded in financial markets that are highly competitive and also often called to be the most dynamic ones [2]. That is why management of financial business processes is identified as a key component to competitive survival by understanding, analyzing and redesigning processes [3]. Brahe and Schmidt provide a qualitative analysis related to efficiency advantages by the introduction of a business process based workflow management system in a financial institution [4]. These approaches often focus on single aspects like strategy or process efficiency [5] but neglect other factors of process performance [6] including process standardization. Weitzel et al. find that efficiency along the transaction processes is influenced by the institutional ability of automated straight through processing (STP) and how involved parties interact at a communicational level [7].

In this paper, we present an empirical performance analysis of a trade booking process within a so-called ‘middle office’, a department of a major European investment company with more than €150 bn. assets under management. The observed process starts when a trade actually took place at the front office and the so-called ‘trading desk’ as well as the associated broker (i.e. a financial market intermediary) submits the trade (T) and the broker confirmation (BC) to the middle office that aggregates all information about the trade, including its fees. Then, the middle office delegates trades to the responsible departments in the back office (B). The trade booking at the middle office terminates when the trade is booked to the back office systems with all relevant fee information (SC) (s. Figure 1).

The delegation of trade information is time-critical as fund customers need to be updated about their investments and values continuously. Further, book keeping and risk management need to be informed about recent trades. If the middle office processes and the information flow between the front office and the back office departments of the investment company are interrupted, inefficiencies arise due to additional costs (e.g. because of manual processing) and risks in the trading process.

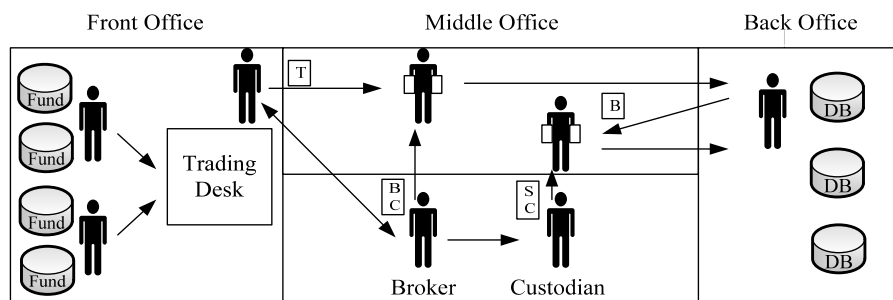


Fig. 1. Process flow in Front, Middle and Back Office of an investment company

Given the complexity of the process described, the automation efforts that were made by related investments into IT and given the diverse factors that influence process performance, we aim to address the following research questions in this paper:

How do technical and organizational factors influence process efficiency in general?

How do IT systems and personnel factors influence process efficiency in particular?

In addition, how do domain-specific and external factors influence process efficiency?

To do so, we first develop a process model of trade bookings and define control points on the basis of experts interviews. In order to explore the impact of the described factors on process efficiency, we present an empirical analysis of a unique dataset of actual trade bookings along with their processing times.

The remainder of the paper is structured as follows: Related work on modeling and measurement of IT Business Value on a process level (2) motivates our research setup and develops hypotheses on influencing factors (3). Descriptive and empirical results on factor process efficiency are presented (4), discussed (5) and concluded (6).

2 Theory and Related Work

2.1 IT Business Value

The subject of business value of IT evaluation is to analyze the contribution of Information Technology to business success. It has been an ongoing issue of Information Systems (IS) research since the 1980s [8] until today [9-10]. Different theoretical foundations, including those of production theory, competitive strategy and consumer value, have found their way in this literature stream [11]. Further, the assessment of IT business value has been explored on different aggregation levels: on higher aggregation levels such as the industry or branch or enterprise level [12-13], or on lower aggregation levels such as the process and business activity level [14-15].

As suggested by Mukhopadhyay et al. [14], these lower aggregation levels of analysis mitigate the difficulty of isolating the true effect of IT on business value, a problem that has been debated extensively in the context of the so-called 'Productivity Paradox' [16]. Consequently, various models and measurement criteria [17] in diverse contexts [18] are described to identify potential process improvement. That is why Marisio [19] argues: 'Measures are the only way to prove improvements in a process' and enable direct empirical evidence compared to indirect measurement by process simulation [20]. Resulting insights may also provide the basis to improve process efficiency [21]. Against this background, we frame our research study at the process and business activity level by collecting and analyzing enterprise process data. To do so, we start our analysis by modeling the business process and by identifying suitable control points within this process.

2.2 Modeling and Measurement of IT Value on a Process Level

'Quantifying the IT investments payoff has always been subject to debate because measurements and real value delivered some time cannot be expressed in typical IT metrics like response time' [22]. Therefore, identifying the relevant factors and the definition of control points is crucial in process management and helps 'to intervene and control processes' [23]. As highlighted by Brynjolfsson and Hitt [11] modeling and measuring the shortcomings in business environments are crucial for empirical research on IT business value. Thus, empirical analyzes being based on the modeling and measuring of problems can contribute to a better understanding of how IT affects business processes and therefore, represent an approach to assess process capabilities and shortcomings that can lead to higher productivity and higher competitiveness [24].

In the context of business processes, existing studies describe how an effective process analysis can contribute to an understanding of relevant factors as well as to the identification of shortcomings. Amadi [25] outlines how to measure the relation of process analysis and strategic alignment to business success, while Heier [26] illustrates how to setup interviews so that IT-related business value can be determined.

Niedermann [27] describes business process optimization as 'design, execution and analysis'. In addition Nissen gives insights on how to measure processes by identify-

ing influencing factors and relevant measures. After modeling a process, the measurement configuration, the process measures and influencing factors must be defined [28]. This enables us to examine the factors that may contribute to both process efficiencies and inefficiencies [27].

Referring to the business value modeling of process views, Shen and Liu [29] connect the perspectives of entity-based activities and the superior organizational structures. In this way, two advantages can be derived: It enables to integrate multiple perspectives of a process and also concludes more generalized results [30]. In consequence, enterprise-deep views allow identifying process risks and potentials, especially in hierarchical organized structures [31].

Performance-centric monitoring of processes makes continuous process improvements possible [32-33]. Thus, models are a precondition to find control points in processes. Braunwarth et al. [34] present a simulation based approach that aims at identifying comparative advantages of manual and automated executions in business processes. In that way, advantages can be identified even in dynamic environments. Therefore, process modeling and the definition of control points can support the flexible alignment of processes, if business requirements change over time [35].

3 Methodology

3.1 General Research Setup

In our study, we followed mixed methods research by combining qualitative research data collection techniques with a quantitative research approach. The advantage of mixed methods research has been highlighted in the literature, as it provides means ‘to deal effectively with the full richness of the real world’ [36]. Therefore, we employ a sequential methodology by first applying qualitative data collection technique (expert interviews) to gain rich insights into the process model followed by a quantitative data collection and analysis of actual process cycles. By doing so, we are able to explore a phenomenon of interest ‘that cannot be fully understood using only a quantitative or a qualitative method’ [37].

To identify and explore the relevant factors that affect the middle office process of trade bookings and the total processing time of an incoming trade, we first studied the history of process models and performed a series of semi-structured interviews with the investment company’s IT department as well as process and domain-specialists. The interviews were conducted by two researchers, both taking notes during the interviews. After each interview, notes were discussed and condensed, which resulted in a detailed process description.

On this basis, we first developed a cross-system process model that discloses involved systems, participants and other relevant external factors. To be able to analyze the total processing time of the process, we defined reference control points within this process, providing the basis for our empirical analysis.

We then conducted a quantitative analysis of historical transaction data that we received from the IT department of the investment company. The transaction data pro-

vided information on a high number of actual process cycles, along with total processing times and information on the relevant factors we have identified earlier.

3.2 Process Description and Model

After the occurrence of a trade and a broker confirmation, both are compared by SYSTEM A (a Middle Office Document Management System). If the confirmations are equal, they are delegated to SYSTEM B (a Document Enrichment System). If they are not equal or another problem occurs then one (and in some cases two) expert(s) has to cross-check the documents and to decide whether to re-request one of the documents or to confirm correctness and equality. The trade information is enriched by expected fees from a data warehouse. SYSTEM B delegates the needed information parts of the enriched trade to the corresponding departments before the trade is stored in the backend database. Parallel to the first booking, the corresponding custodian bank (that stores the asset testament) sends a settlement confirmation with the actual fees that is prestored asynchronously at SYSTEM A. The preprocessed trade is reloaded from the backend database and compared to the information of the settlement confirmation. If all information equals then the trade does not have to be rebooked. If it is unequal then the prebooked trade can be corrected automatically within given ranges, otherwise it is manually adjusted by one or two experts, before resending it to SYSTEM B and storing it in the backend database. Additionally, following corrections can trigger another rebooking cycle (s. Figure 2).

The predefined control points help to validate the analysis from two different perspectives. First, some of the operations in the process are triggered asynchronously (like the arrival of a settlement confirmation), so their influences overlay the effects of the total processing time within the middle office process chain. Second, we focus on process efficiency of the superordinated process. That way measuring the operational efficiency of single steps and activities within the process would give more interesting insights to the process, but limit the generalizability to these specific process steps. That is why we use these additional control points for validation and better generalizability of the results only. Focusing on our research questions we investigate influences to the middle office process and its total processing time in general.

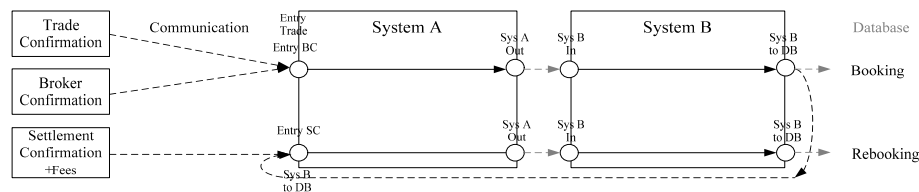


Fig. 2. Trade Booking Model

3.3 Influencing Factors and Hypotheses

We build upon previous research and aim at measuring influencing factors directly. So, we align our research model with Subramanyam and Krishnan [38] who analyze

the resolution times in a call center within an academic organization on a process level. They identified four factor groups that influence the time needed by call center agent need to solve a customer problem. These are: (1) the IT system usage, (2) the personnel factors of call center agents, (3) the problem set and (4) the customer describing the problem. In the following we built upon these factor categories and adjust them to the context of the middle office of the investment company. Consequently, and compared to the original study of Subramanyam and Krishnan [38], our study explores process efficiency within a different organization (financial industry compared to academic organization) in a different application domain (transaction processing compared to customer service).

(1) IT System Usage / Automation. In general, it is assumed that IT and information systems can have positive influences on process efficiency [39]. Here, IT systems are designed to support business and information requirements, help to enhance data consistency and to reduce redundancies, which make processes more stable and reliable [40]. IT systems can enhance supportive processes for the core businesses (like production processes) if they are backed by IT systems [41]. This requires that automation and supporting systems are maintained to fit the process requirements for efficiency [42]. *That is why we expect that IT-enabled automation can decrease the trade booking time, which contributes to general process efficiency.*

(2) Personnel Factor / Brokers&Custodians. Banker and Slaughter [43] and Agrell et al. [44] find that the capabilities and the personnel experience of domain experts positively influence task performances. *That is why we expect that brokers and custodians can decrease the trade booking time and the general process efficiency.*

(3) Problem Set / Subassets. Schiefer [45] states that depending on the problem area of the process there is an additional problem-specific coordination needed within the process chain. Also dynamic approaches that cause more organizational overhead can solve such problem-specific tasks [46]. Especially in highly controlled process environments [47], detection and solving of problems may significantly increase quality [48]. So it might result in longer overall total processing times. Results strongly differ according to the type of application and problem set [49], but a strong problem-specific influence can be assumed. *That is why we expect that the problem-set (here: subasset classes) have a shortening impact to the overall trade booking time.*

(4) Customers / Fund Managers. In our case, fund managers at the front office are like customers to the middle office as they make use of their supportive services. Subramanyam and Krishnan [38] criticize in line with other studies [50] that research of customer factors influencing the IT business value is quite scarce. Cameron and Braiden [51] find that customers have a demand for lower costs and reduced processing times, so they might be willing to enhance efficiency as long as the customer needs and objectives are clearly identifiable. In the context of the investment management industry and the relation between the front and middle office, this assumption has been supported by ethnographic studies [52]. Van Zelst et al. [53] found that

employees that are disturbed and interrupted by customers have reduced process efficiency and have to fulfill customer demands. *So we expect for fund managers that they can decrease as well as increase the trade booking time and the general process efficiency.*

3.4 Process Analysis

To measure the total processing time of the trade booking process, we evaluated and aggregated all log events of a single trade into a single transaction log. The total processing time is measured as time period starting when trade and broker confirmation are present and ending when the last rebooking is done. All intermediated control points are used for validation as we are just interested in the total processing time of the process. Settlement confirmations are not evaluated because of asynchronous event preprocessing in SYSTEM A.

Transactions with missing start and end events (logging errors) as well as transactions that seemed to be mixed up compared to the standard process (complex-distributed asynchronous errors) were revised from the processing time sample.

As we observe process efficiency in terms of total processing time (processing a trade with needed fee information), we select a hazard function model regression [54]. This supports the non-linear behavior of processing times as well as the strict positive characteristics of the model variables and avoids broken assumptions compared to a linear regression [54]. Designed to estimate how long an entity will stay in a certain state, these models have been applied to analyze influences strike duration, divorce rates, length of studies and pensions and mortality expectations in social science [54].

The hazard rate λ is the likelihood at which trade i did not change the state within a given period. The model estimates the likelihood with given influencing factors and allows to estimate direct influences to the processing time. That is why we expect that the processing time is dependent to their influencing factors:

Processing time = f (Automation, Subassets, Custodians, Brokers, Funds, Weekdays)(1)

As the error rates decrease over time (longer processing times are much less likely than shorter ones) we expect a Weibull distribution of the processing time (positive random variables and not normal-distributed) that is also often used in previous research [55] and validate this assumption with the descriptive statistics in the next section.

IT business value is explored by analyzing the influence of process automation, which is measured by four disjunctive dummy variables ((1) full automation, (2) partly automated in SYSTEM A and (3) SYSTEM B or (4) manual). For each occurrence of an influencing factor (assets subclass, brokers, custodians, fund managers), an entity-specific dummy variable is defined. Weekday dummy variables are added regarding assumable weekday effect. The significance of all dummy variables is tested by a Chi-squared test for each factor category as well as for the overall model.

4 Data and Empirical Results

4.1 Dataset and Descriptive Statistics

Our dataset consists of 29200 valid trade bookings including the total processing time and variables for 4 automation states, 23 subasset classes, 38 custodians, 91 broker, 561 funds and control variables for 5 weekdays. All time variables are measured in milliseconds and the average processing time is $1.001 \cdot 10^8$ ms (27.7 hours). The median is $7.3815 \cdot 10^6$ ms (20 hours). 82.14% of bookings are completed within two days. With one day more, namely three days, 88.61% of all transactions are booked. It takes 6.158 days to complete 99% of all transactions. On the one hand, there is the fastest transaction measured with 3.25 min (195000 ms), while on the other hand, the slowest booking took 103 days ($8.89761 \cdot 10^9$ ms). The total processing time has a standard deviation of $1.47681 \cdot 10^8$ ms. The difference between average and median indicates a right-skewed distribution. The histogram (5 outliers $> 2.224 \cdot 10^9$ ms have been discarded) depicted in figure 3 indicates a Weibull distribution that approximates the distribution of processing times best compared to other hazard models.

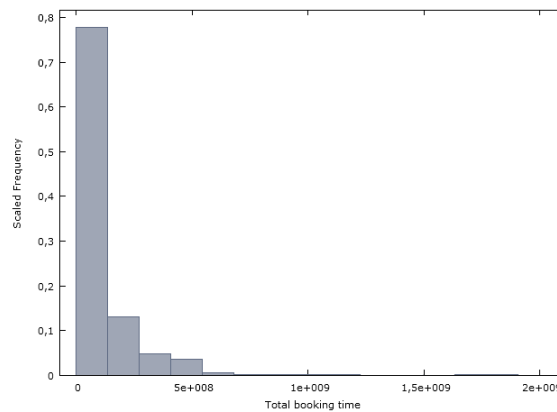


Fig. 3. Frequency distribution of total processing times measured in milliseconds

4.2 Empirical Results

The previous section illustrated that the distribution of trade bookings shows declining processing times and that times are positive, randomly ordinary and not normal-distributed. That is why we regress all 29200 trade observations as a cross-sectional dataset and investigate the influences using a Weibull-distributed hazard function model. Results of the regression analysis explain the influence of each individual entity within the four factor categories to the processing time.

Due to the complexity of the dataset, its 717 variables and the entity specific perspective to project influence, we ease the readability of the regression results by a bar chart (s. Figure 4) that highlights positive and negative influences (above/below y-axes) as well as the significance level of the results (1-3 units). Bars above the hori-

zontal line show numerical negative influence (shorter processing times), bars below highlight numerical positive influence (longer processing times). The longest bars with 3 units length indicate high levels of significance ($< 1\%$), mid length bars with 2 units indicate significances ($< 5\%$) and bars with one unit length indicate significance ($< 10\%$), but not at the conventional 5% level. Grey fields indicate no significance. An extract of the numerical results is given in Table 1.

System-specific automation variables show high significance as the p -values are below the 1% of significance. The rate at which the automation reduces the total processing time of the trade bookings is between -88.61% and -47.54%. Interestingly the influence with partly automated processing is even higher than with full automation (significance). Full automation has also a negative influence, but the influence is not significant compared to the other factors in the regression. These findings provide evidence for positive IT business value since the related IT investments facilitated automated processing resulting in decreased process runtimes on the process level.

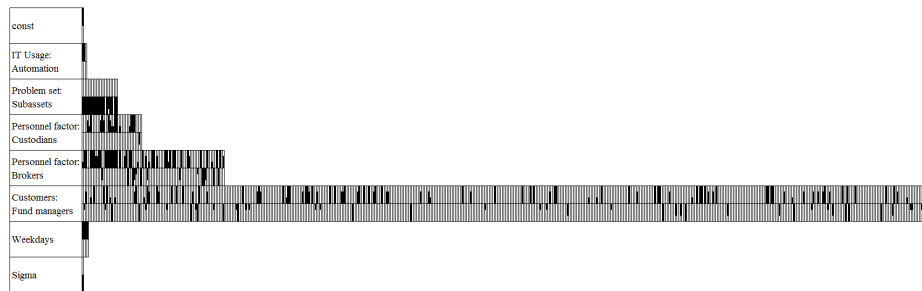


Fig. 4. Plot of directions of influences and levels of significance of each entity in the dataset

All subasset class variables are estimated with positive coefficients by the hazard model regression. That gives evidence that most assets have a prolonging influence to the total processing time of the middle office process. Significant coefficients lie in the ranges of 74.79% up to 276.15%. Except for two coefficients that are not significant (Subasset 15 and 20) all other asset sub classes show a high level of significance.

Most custodian and broker variables show significant influences to the total processing time in the middle office with levels of significance at 1, 5 and 10%. The bar chart easily highlights positive bars above the horizontal line that indicate that the influences of brokers and custodian banks are numerical negative (shortening processing times in the middle office). The best custodian bank (Custodian 33) influences processing times that way that it is 2.59 times faster than the average, while the worst custodian bank (Custodian 36) has a slowing effect that is 110% longer than the average. Broker coefficient ranges lie between factors of 1.13 faster (Broker 42) and 1.58 (Broker 61) slower than the average. These outliers decelerating the process are also detectable as 'bars below the horizontal line'. Compared to all other broker and custodian banks, these negative influences are occasionally.

Some funds specific variables show significant influences, but those significances were observed less frequently than in the other factor groups. In addition, the directions of the influences vary: Some fund managers have an accelerating influence to

the trade booking, while others are decelerating the process. The range of coefficients lies between -2.58 and 2.56, which means that some fund managers are two and a half times faster or slower than the average. Compared to other factor groups in Figure 4, it is easy to see that the fund-specific factor have the lowest relative significance.

Control variables on weekdays show the expected highly significant influences. While Fridays show a positive influence as processes might block or stop during the weekend, all other weekdays have a negative influence. These weekdays show coefficients that indicated a decrease in processing times during the work week. Starting with Mondays that are up to 70% faster than Fridays, the processing time decreases to being 50% faster on Thursdays. A Chi-squared test indicates overall model validity.

Table 1. Extract of regression results for total processing time with the first two significant variables of each factor category

	Coefficient	Std. Error	z	p-Value	
Constant	18.86060	0.46416	40.63350	<0.00001	***
Just Sys A Auto	- 0.88163	0.17460	- 5.04930	<0.00001	***
Just Sys B Auto	- 0.47541	0.18145	- 2.62010	0.00879	***
All Auto	- 0.28386	0.18882	- 1.50330	0.13276	
Subasset 1	1.05527	0.22396	4.71180	<0.00001	***
Subasset 2	0.83046	0.22427	3.70300	0.00021	***
...
Custodian 4	- 1.04503	0.42224	- 2.47500	0.01332	**
Custodian 5	- 1.13403	0.64595	- 1.75560	0.07916	*
...
Broker 1	- 0.13079	0.07870	- 1.66190	0.09654	*
Broker 2	- 0.33210	0.11916	- 2.78710	0.00532	***
...
Fund 2	2.00083	1.07391	1.86310	0.06244	*
Fund 3	- 0.75065	0.36130	- 2.07760	0.03774	**
...
Monday	- 0.70196	0.02515	-27.91270	<0.00001	***
Tuesday	- 0.62772	0.02386	-26.30890	<0.00001	***
Wednesday	- 0.58743	0.02348	-25.01860	<0.00001	***
Thursday	- 0.50146	0.02536	-19.77730	<0.00001	***
sigma	1.20606	0.00559	215.67620	<0.00001	***
Chi-square (679)	7,165.40		p-Value	0.00	***

*/**/*** = significant at a 10%/5%/1% level.

Given the above statistics, we would like to highlight that it is not our main research objective to identify individual Custodian or Broker contributions to process performance but to identify the general influences and direction of each entities group.

5 Discussion

5.1 Implications on Theory

The empirical results give evidence that the hypotheses extracted from Subramanyam and Krishnan's call center setup can also be supported with processes from other domains like the financial industry and in other contexts like transaction processing processes. Additionally, we enhanced the quality of research by building upon quali-

tative data collection techniques and by adding a control variable for week day influences. In general this measurement setup is an extension to well-established process analysis methods. Descriptive research (like literature reviews) and qualitative research results (like case studies) can be cross-validated with precisely measured real world data. This approach is scalable. In the given setup we focused on a central cross-system process and included the involved entities like systems, employees, customers (fund managers) and products (subasset classes) as possible multiple influencing factors. Additionally, we manage the complexity of the process and its efficiency so that results stay generalizable. Additional information about subprocesses is included to validate the correctness of the main process. That way we measure the direct influence of each entity and groups of entities to the overall processing time and results are not blurred by a subprocess-specific measurement. As a result, process models from previous research can be used to define control points for each involved entity and evaluated the specific influence of each entity to the overall process efficiency. This approach is not restricted to process runtimes, but can also be extended to error or cost measures and helps closing the deduction-induction research cycle for process modeling.

5.2 Practical Implications

Reflecting the research questions from the introduction our results show that processes and their efficiency in the dimension of runtime are not only influenced by technical and personnel factors. Especially domain-specific influences, like the processed subasset class, show that problem-specific influence has a clearly negative impact on duration of the trade booking. It is also showing that external factors, like fund managers that are customers of this service, have mixed influences to the process. Some fund managers decelerate the process, while others accelerate it. In general there is no consistency in the influence of external factors as some entities are significant and others are not. So evidence is given for the third research question, i.e. how domain-specific and external factors influence process efficiency: domain and problem-specific influences have a negative effect on efficiency, which results in longer processing times (hypothesis 1), while the influence of external factors depends on the specific entity (like a specific customer / fund managers in hypothesis 4).

After filtering external factors and problem-specific influences from our data by explaining their influence to efficiency of processing times, it becomes possible to discuss the influences from the technical and personnel influences within the measured process. Results show high significance for system-specific influences. Interestingly this influence is not statistical significant with regard to full-automation processes. One interpretation of the results is that influences from personnel factors cannot be neglected. As coefficients show that automation has always a negative (accelerating) effect to processing time, results also indicate that the full-automation effect is less important (significant) than other influences from the model. This is confirmed by other human/ personnel factors like the corresponding broker and the staff at the custodian banks. Their domain-specific knowledge seems to have also negative (accelerating) effects that counteract the external and problem-specific influences from

the previous section. In general, we answer the second research question on IT systems (hypothesis 2) and personnel factors (hypothesis 3) by confirming a negative (accelerating) influence on the efficiency of processing time.

In general, we see for the first research question that technical and organizational factors have significant influence on the time efficiency of the process. We also see that there are positive influences by technical and personnel factors, negative influence by domain and problem-specific factors and mixed influences by external factors like fund managers/customers. Our measurement setup gives not only insights into the direction and the significance level of the analyzed factor groups. Individual and specific influences give detailed information on which organizational and which technical entities can and should be improved. That way it supports decision making, modeling of future process restructuring and measurement for potentially implemented improvements.

5.3 Limitations

While our dataset covers a period of 12 months with multi-dimensional influences like fund manager, subasset-specific influences, broker and custodians as well as the influence of technological support, our research perspective is limited to a single case. This case has a data-intensive setup with extensive automation and support potential within the financial domain, but on the other hand, our case might have a domain-specific bias. Our dataset shows data for different subasset classes, but because of legal issues, our current perspective is restricted to the booking process of bonds and lacks of less standardized financial assets. In addition, we had to remove incomplete process cycles and those where the logging chronology did not fit the assumed process model. Due to the nature of asynchronous event processing, this is not surprising, but it discards events that do not fit the assumed process definition, which may limit the results, too. Multiple control and validation methods have been applied to check the robustness of the results and to avoid a bias resulting from dataset characteristics.

6 Conclusion

Our three research questions concerning process efficiency, technical, personnel, domain-specific and external factors are analyzed following as mixed methods research that combines qualitative research data collection techniques with a quantitative research approach. Using interview and modeling techniques we identified measurement points and influencing entities within a middle office process of a major European investment company. We measured the trade booking times of 29200 transactions and the influence of entities that are related to these specific transactions. As a result, we found evidence that IT system usage and personnel factors like brokers and custodians have a shorting influence to processing, while problem-specific influences like subasset classes extend the processing time. We found mixed influences by process customers like fund managers. From a methodological perspective our study

illustrates how to conduct mixed methods research in the field of business process performance evaluation.

For further research, we plan to address the identified limitations of the current research setup and aim to extend the dataset. A larger sample size and may help to evaluate our findings. To upgrade our research, it is also possible to validate our findings in other domains. This would enable to compare the influences to transaction processing time in different domains. During the research, there is also the opportunity to take other aspects of influence into account, such as motivational constraints, team and communication structures as well as the clarity of the process improvement purposes and the expectations of participants. We thankfully acknowledge the support of the E-Finance Lab, Frankfurt and its research partners for this work.

References

1. Weiss, B., Winkelmann, A.: Developing a Process-Oriented Notation for Modeling Operational Risks - A Conceptual Metamodel Approach to Operational Risk Management in Knowledge Intensive Business Processes within the Financial Industry. *System Sciences (HICSS)*, pp. 1-10 (2011)
2. Vykoukal, J., Setzer, M., Beck, R.: Grid Architecture for Risk Management: A Case Study in a Financial Institution. *PACIS 2008 Proceedings*, Paper 211 (2008)
3. Aguilar, M., Rautert, T., Pater, A. J. G.: Business process simulation: A fundamental step supporting process centered management. In: Farrington, P.A., Nembhard, H.B., Sturrock, D.T., Evans, G.W. (eds.): *Proceedings of the 1999 Winter Simulation Conference*, Phoenix/AZ, pp. 1383-1392. ACM Press (1999)
4. Brahe, S., Schmidt, K.: The story of a working workflow management system. In Gross, T. et al. (eds.): *Group 2007: Proceedings of the International Conference on Supporting Group Work*, pp. 249-258 (2007)
5. Reitbauer, S., Kohlmann, F., Eckert, C., Mansfeldt, K., Alt, R.: Redesigning business networks: reference process, network and service map. In: *Proceedings of the 2008 ACM symposium on Appl. computing (SAC '08)*, pp. 540-547. ACM, New York, (2008)
6. Münstermann, B., Eckhardt, A., Weitzel, T.: The performance impact of business process standardization – An empirical evaluation of the recruitment process. In: *Business Process Management Journal (BPMJ)* 16 (1), 29-56 (2010)
7. Weitzel, T., Martin, S.V., König, W.: Straight Through Processing auf XML-Basis im Wertpapiergeschäft. In: *Wirtschaftsinformatik* 45 (4), 409-420 (2003)
8. Solow, R.: We'd better watch out. In: *New York Times Book Review*, p. 36 (1987)
9. Lee, D.H.-D.: Contextual IT Business Value and Barriers: An E-Government and E-Business Perspective. *Proceedings of the 38th Ann. Hawaii Int. Conference* (2005)
10. Kohli, R., Grover, V.: Business value of IT: An essay on expanding research directions to keep up with the times. *Jour. of the Ass. for Information Systems* 9 (1), 23-39 (2008)
11. Brynjolfsson, E., Hitt, L.: Paradox Lost? Firm-level Evidence on the Returns to Information Systems Spending. *Management Science* 42 (4), 541-558 (1996)
12. Loveman, G.W.: An Assessment of the Productivity Impact on Information Technologies. In: *Information Technology and the Corporation of the 1990s: Research Studies*, pp. 84-110. MIT Press, Cambridge, MA (1994)

13. Morrison, C.J., Berndt, E.R.: Assessing the Productivity of Information Technology Equipment in U.S. Manufacturing Industries. NBER Working Papers 3582, National Bureau of Economic Research, Inc. (1991)
14. Mukhopadhyay, T., Kekre, S., Kalathur, S.: Business Value of Information Technology: A Study of Electronic Data Interchange. *MIS Quarterly* 19 (2), 137-156 (1995)
15. Francis, A., MacIntosh, R.: The market, technological and industry context of business process re-engineering in the UK. *International Journal of Operations & Production Management* 17 (4), 344-364 (1997)
16. Dewan, S., Kraemer, K.L.: International dimensions of the productivity paradox. *Communications of the ACM* 41 (8), 56-62 (1998)
17. Tan, W.A., Shen, W.M., Xu, L.D., Zhou, B.S., Li, L.: A Business Process Intelligence System for Enterprise Process Performance Management. *IEEE Transactions on Systems Man and Cybernetics Part C: Applications and Reviews* 38 (6), 745-756 (2008)
18. Alfaro, J.J., Rodriguez-Rodriguez, R., Verdecho, M.J., Ortiz, A.: Business process interoperability and collaborative performance measurement. *International Journal of Computer Integrated Manufacturing* 22 (9), 877-889 (1998)
19. Marisio, M.: Measurement Processes are Software, Too. *Journal of Systems and Software* 49 (1), 17-31 (1999)
20. Villeta, M., Rubio, E.M., Sebastian, M.A., Sanz, A.: New criterion for evaluating the aptitude of measurement systems in process capability determination. *International Journal of Advanced Manufacturing Technology* 50 (5-8), 689-697 (2010)
21. Garvin, M.J., Wooldridge, S.C., Miller, J.B., McGlynn, M.J.: Capital planning system applied to municipal infrastructure. *Jour. of Management in Eng.* 16 (5), 41-50 (2000)
22. Gupta, J., Sharma, S.: Measuring Business Value of IT Investments of E-Commerce Using Porter's Framework. *AMCIS 2003 Proceedings*, Paper 38 (2003)
23. Buhl, H.U., Winter, R.: Full Virtualization - BISE's Contribution to a Vision. *BISE Edition* 1 (2), 133-136 (2009)
24. Lesáková, L.: The Process of Forming the Regional Innovation Strategy. *Acta Polytechnica Hungarica* 8 (1), 5-22 (2011)
25. Amadi, A.: Tying Strategic Alignment and IT Value to Business Success Using Business Process Analysis And Redesign (BPAR). *AMCIS Proceedings*, Paper 402 (1998)
26. Heier, H., Borgman, H.P., Maistry, M.G.: Examining the Relationship between IT Governance Software and Business Value of IT: Evidence from Four Case Studies. *System Sciences, 40th Annual Hawaii International Conference* (2007)
27. Niedermann, F., Radeschütz, S., Mitschang, B.: Deep business optimization: A platform for automated process optimization. *LNI, Vol. P-177*, pp. 168-180. GI, Bonn (2010)
28. Nissen, M.E., Sengupta, K.: Incorporating software agents into supply chains: Experimental investigation with a procurement task. *MIS Quarterly* 30 (1), 145-166 (2010)
29. Shen, M.X., Liu, D.R.: Coordinating interorganizational workflows based on process-views. *Database and Expert Systems Applications Book Series: LNCS, Vol. 2113*, pp. 274-283. Springer, Berlin Heidelberg (2001)
30. Choi, Y., Kang, D., Chae, H., Kim, K.: An enterprise architecture framework for collaboration of virtual enterprise chains. *International Journal of Advanced Manufacturing* 35 (11-12), 1065-1078 (2008)
31. Kmec, P.: Temporal hierarchy in enterprise risk identification. *Management Decision* 49 (9), 1489-1509 (2011)
32. Han, K.H., Kang, J.G., Song, M.: Two-stage process analysis using the process-based performance measurement framework and business process simulation. *Expert Systems with Applications* 36 (3), 7080-7086 (2009)

33. Han, K.H., Kang, J.G., Lee, G.: Performance-centric business activity monitoring framework for continuous process improvement. Proceedings 9th WSEAS International Conference on Artificial intelligence, knowledge engineering and data bases, pp. 40-45 (2010)
34. Braunwarth, K.S., Kaiser, M., Müller, A.-L.: Economic Evaluation and Optimization of the Degree of Automation in Insurance Processes. BISE Edition 2 (1), 29-39 (2010)
35. Zhang, Q.Y., Vonderembse, M.A., Cao, M.: Achieving flexible manufacturing competence - The roles of advanced manufacturing technology and operations improvement practices. *Int. Journal of Operations and Production Manag.* 26 (5-6), 580-599 (2006)
36. Mingers, J.: Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research* 12 (3), 240-259 (2001)
37. Venkatesh, V., Brown, S.A., Bala, H.: Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly* 37 (1), 21-54 (2013)
38. Subramanyam, R., Krishnan, M.S.: Business value of IT-enabled call centers: An empirical analysis. *Proc. 22nd Internat. Conf. Inform. Systems, New Orleans*, pp. 55-64 (2001)
39. Mukhopadhyay, T., Rajiv, S., Srinivasan, K.: Information Technology Impact on Process Output and Quality. *Management Science* 43 (12), 1645-1659 (1997)
40. Xu, H., Koh, L., Parker, D.: Business processes inter-operation for supply network coordination. *International Journal of Production Economics* 122 (1), 188-199 (2009)
41. Aybar, M., Potti, K., LeBaron, T.: Modeling methodology: using simulation to understand capacity constraints and improve efficiency on process tools. WSC '02 Proceedings of the 34th Conference on Winter simulation: exploring new frontiers, pp. 1431-1435 (2002)
42. Zeltyn, S., Tarr, P., Cantor, M., Delmonico, R., Kannegala, S., Keren, M., Kumar, A.P., Wasserkrug, S.: Improving efficiency in software maintenance. *Proceeding MSR '11 Proceedings of the 8th Working Conf. on Mining Software Repositories*, pp. 215-218 (2010)
43. Banker, R., Slaughter, S.A.: Field Study of Scale Economies in Software Maintenance. *Management Science* 43 (12), 1709-1725 (1997)
44. Agrell, P.J., Bogetoft, P., Tind, J.: Incentive plans for productive efficiency, innovation and learning. *International Journal of Production Economics* 78 (1), 1-11 (2002)
45. Schiefer, G.: Environmental control for process improvement and process efficiency in supply chain management the case of the meat chain. *International Journal of Production Economics* 78 (2), 197-206 (2002)
46. Wibe, S.: Efficiency: A dynamic approach. *International Journal of Production Economics* 115 (1), 86-91 (2008)
47. Bordoloia, S., Guerrero, H.H.: Design for control: A new perspective on process and product innovation. *International Journal of Production Economics* 113(1), 346-358 (2008)
48. Pool, A., Wijngaard, J., van der Zee, D.-J.: Lean planning in the semi-process industry, a case study. *International Journal of Production Economics* 131 (1), 194-230 (2011)
49. Romero, C.A., Rodriguez, D. R.: E-commerce and efficiency at the firm level. *International Journal of Production Economics* 126 (2), 299-305 (2010)
50. Kim, S., James E.W., Jr.: How long did it take to fixbugs?. *MSR 06 proceedings* (2006)
51. Cameron, N.S., Braiden, P.M.: Using business process re-engineering for the development of production efficiency in companies making engineered to order products. *International Journal of Production Economics* 89 (3), 261-273 (2003)
52. Ho, K.: *Liquidated: An Ethnography of Wall Street*. Duke University, Durham (2009)
53. van Zelst, S., van Donselaar, K., van Woensel, T., Broekmeulen, P., Fransoo, J.: Logistics drivers for shelf stacking in grocery retail stores: Potential for efficiency improvement. *International Journal of Production Economics* 121 (2), 620-632 (2009)
54. Greene, W.H.: *Econometric Analysis*. Prentice Hall (2007)

Agilität von Geschäftsprozessen trotz Compliance

Kai Kittel

Martin-Luther-Universität Halle-Wittenberg, Lehrstuhl für Wirtschaftsinformatik, insb.
Betriebliches Informationsmanagement, Halle, Germany
kai.kittel@wiwi.uni-halle.de

Abstract. Die Unterstützung der Agilität von Geschäftsprozessen, um bspw. auf veränderte Marktsituationen schnell reagieren zu können, ist eine der aktuellen Anforderungen an Workflowmanagementsysteme. Eine Agilität der Geschäftsprozesse bedeutet, Ad-hoc-Änderungen an dessen Aktivitäten ausführen zu können, was im schlechtesten Fall vorher integrierte Kontrollen für Compliance unbrauchbar macht. Diesen scheinbaren trade-off zwischen Agilität und Compliance zu verringern, ist Ziel des vorliegenden Artikels, welcher eine neue modell-basierte Methode zur Ad-hoc-Integration von Compliance-Kontrollen in Workflows vorstellt. Insbesondere werden in diesem Artikel die Eigenheiten von Agilität und Compliance diskutiert, die neue Integrationsmethode erläutert, und deren praktische Umsetzung mit Hilfe eines Prototyps demonstriert.

Keywords: Compliance, Geschäftsprozessmanagement, WfMS, Controls, Agilität

1 Agilität und Compliance bei Workflows

Compliance, d. h. die Konformität (eines Unternehmens) mit gesetzlichen, aber auch unternehmensinternen Vorgaben, ist eine der wesentlichen Herausforderungen, denen sich ein Unternehmen aktuell stellen muss [1]. Dies bedeutet, dass ein Unternehmen mit Compliance-Anforderungen (CA) im Einklang stehen muss [2] und deren Einhaltung zu validieren hat [3]. Insbesondere die Ausgestaltung von Geschäftsprozessen, welche die Grundlage jeder unternehmerischen Tätigkeit darstellen sowie die Berücksichtigung entsprechender Kontrollmaßnahmen bei deren Ausführung, sind von besonderer Bedeutung für die Erreichung von Compliance [4].

Unternehmen nutzen die fortschreitenden Entwicklungen im Bereich der Workflow-Management-Systeme (WfMS), um ihre Geschäftsprozesse zu automatisieren oder IT-gestützt auszuführen [5]. Hierbei erfordern effektive und effiziente Kontrollen für Compliance, dass Kontrollaktivitäten ebenfalls automatisiert bzw. mit Hilfe von WfMS unterstützt ausgeführt und durchgesetzt werden. Die Sicherstellung der Compliance in automatisierten Geschäftsprozessen erfolgt bislang überwiegend durch „festverdrahtete“ Kontrollaktivitäten, welche in die Geschäftsprozesse bei deren Entwurf integriert werden [6-7]. Zum Ausführungszeitpunkt werden damit die Kontrollaktivitäten erzwungen, indem sie durch das WfMS automatisch ausgeführt

werden. Eine solch manuelle Integration von Kontrollen zum Entwurfszeitpunkt ist jedoch personal- und kostenintensiv.

Im Fall von agilen Geschäftsprozessen, welche viele Unternehmen heutzutage aus Wettbewerbsgründen implementieren [8], stößt ein solch manueller Ansatz zunehmend an Grenzen. Mit jeder Änderung des Ablaufs wird eine Überprüfung und/oder Anpassung der Kontrollen erforderlich, um deren Effektivität sicherzustellen. Dies führt in Unternehmen mit einer Vielzahl an Geschäftsprozessen und unterschiedlichen CA schnell zu einer steigenden Komplexität, die das Management von Compliance bei agilen Geschäftsprozessen, respektive eine manuelle Integration von Kontrollen nahezu unmöglich macht.

Um die Problematik zwischen Agilität und Compliance der Geschäftsprozesse zu reduzieren, wird im Folgenden eine neue Methode vorgestellt, die während der Ausführung von Workflows automatisiert Kontrollaktivitäten integriert und somit diesem scheinbaren Trade-off zwischen Agilität und Compliance entgegenwirkt. Im nächsten Kapitel werden hierfür der aktuelle Stand der Forschung für die Integration von Kontrollen in Geschäftsprozesse herausgearbeitet und bekannte Ansätze vorgestellt bzw. deren Probleme untersucht. Im nachfolgenden Kapitel drei wird dann als neuer Lösungsansatz die neue Methode „FlexCom“ zur flexiblen Integration von Kontrollen in automatisierte Geschäftsprozesse (Workflows) zum Zeitpunkt ihrer Instanzierung vorgestellt. Der design-science Forschungsmethodik [9] folgend, stellt die neue Methode das zu entwickelnde Artefakt dar, welches durch die Erstellung eines Prototypen evaluiert wird. Dieser Prototyp „KitCom“ wird im vierten Kapitel beschrieben. Abschließend werden, neben einer kurzen Zusammenfassung, insbesondere die Lessons learned beschrieben und ein Ausblick auf noch offene Forschungsfragen gegeben.

2 Agilität von Geschäftsprozessen und Methoden zur Integration von Kontrollen in Workflows

In dynamischen Märkten lassen Kundenanforderungen oder auch das Verhalten der Konkurrenz die Agilität von Unternehmen zum relevanten Wettbewerbsfaktor werden [8]. Agile Geschäftsprozesse ermöglichen es, auf geänderte Marktsituationen schnell und kostengünstig zu reagieren [10]. Eine Grundlage für die Agilität von automatisierten oder IT-gestützten Geschäftsprozessen bildet Flexibilität sowohl auf der Konstruktionsebene als auch auf der operativen und technischen Ebene [11].

Auf der *Konstruktionsebene* der Geschäftsprozesse wird Flexibilität bzgl. der Abläufe (inkl. der verwendeten Daten und Ressourcen) benötigt [12]. Hierfür stehen unterschiedliche Geschäftsprozessmodellierungssprachen zur Verfügung, bspw. UML [13], EPK [14] oder BPMN [15]. Während sich die einzelnen Modellierungssprachen an unterschiedliche Adressaten wenden und sich in ihrer Mächtigkeit und ontologischen Klarheit unterscheiden [16], bieten sie auf der Konstruktionsebene prinzipiell die benötigte Flexibilität bzgl. der Ablaufgestaltung.

Ein großer Teil der Geschäftsprozesse in Unternehmen ist bereits teil- oder vollautomatisiert, mindestens jedoch die Ausführung durch IT unterstützt [17-18]. Diese

starke Abhängigkeit von der IT macht eine adäquate Berücksichtigung der *technischen Ebene* notwendig. Einen vielversprechenden Ansatz, um Flexibilität auf der technischen Ebene zu erreichen, bilden so genannte service-orientierte Architekturen (SOA) [19-20]. Hierbei werden betriebliche Funktionalitäten als Dienste gekapselt, in Form von internen oder externen Webservices implementiert und nach Bedarf in die Ausführung der Geschäftsprozesse flexibel eingebunden [21]. Flexibilität auf der Konstruktionsebene kann somit prinzipiell auch durch die technische Ebene unterstützt werden [22].

Konstruktionsebene und technische Ebene sind über die operative Ebene mit Hilfe von WfMS verbunden [23]. Hierfür werden, noch auf der Konstruktionsebene, die modellierten Geschäftsprozesse zu Workflows konkretisiert. Die Zusammenführung zwischen der Konstruktionsebene und der technischen Ebene ist im Status quo jedoch mit Einschränkungen der Flexibilität verbunden: Ad-hoc-Änderungen einer Workflow-Instanz werden von aktuellen WfMS nicht ausreichend unterstützt und führen zu fehlender Flexibilität und/oder manuellen Workarounds außerhalb des WfMS [24-25]. Erste WfMS, die das Flexibilitätspotential auch während der Ausführung von Workflows unterstützten, wurden zwar bereits entwickelt [12] und Agilität von Geschäftsprozessen kann mit solchen Verfahren zumindest auf syntaktischer Ebene unterstützt werden, jedoch bislang ohne Berücksichtigung semantischer Aspekte wie der Einhaltung von Compliance.

Zur Einhaltung von Compliance im Kontext automatisierter Geschäftsprozesse wird die Adaption von CA von jedem betroffenen Unternehmen auf seinen jeweiligen Unternehmenskontext vorgenommen und als Regeln abgelegt [26]. Zentrales Werkzeug zur Sicherstellung der Einhaltung eines Teils dieser Regeln sind ablaufbezogene Kontrollprozesse, die Kontrollen als Vergleich zwischen Ist- und Sollzustand sowie den Umgang mit Abweichungen beinhalten [27-28]. Bei ablaufbezogenen Kontrollen für Geschäftsprozesse lassen sich zwei mögliche Ansätze unterscheiden, welche der Überwachung dienen und die Wahrscheinlichkeiten eines Regelverstößes reduzieren sollen [29]: Einerseits gibt es Kontrollprozesse, welche für sich alleine stehen und unabhängig vom Geschäftsprozess ausgeführt werden. Diese sogenannten Prüfungen können auch in flexiblen und automatisierten Geschäftsprozessen ohne zusätzliche Maßnahmen erfolgen und verhalten sich wie Geschäftsprozesse nur mit anderer Zielstellung. Andererseits gibt es Kontrollaktivitäten, welche abhängig vom Ablauf der Geschäftsprozesse sind und mit diesen verbunden werden müssen. Diese Abhängigkeit bedarf besonderer Beachtung und lässt sich wiederum in zwei Arten unterscheiden [30]:

Compliance by detection: Bei diesen Ansätzen wird nach der eigentlichen Ausführung der Workflow-Instanzen eine Überprüfung im Sinne eines Audits durchgeführt (bspw. [31-32]). Die Einhaltung der CA wird hierbei lediglich validiert und Verletzungen ggf. im Nachhinein sanktioniert. Aus Sicht des Geschäftsprozessmanagements haben by design-Ansätze den Vorteil, dass die Flexibilität nicht eingeschränkt wird und die originären Geschäftsprozesse bei ihrer Ausführung nicht durch Kontrollprozesse beeinflusst oder gar gestört werden [33]. Aus Sicht des Compliance-Managements sind by detection-Ansätze jedoch mit erheblichen Nachteilen und unter Umständen mit betriebswirtschaftlich hohen Risiken verbunden, da sie Verstöße ge-

gen die CA nicht verhindern können [34]. Aufgrund dieses schwerwiegenden Nachteils werden by detection-Ansätze im Folgenden nicht weiter betrachtet.

Compliance by design: Hierzu zählen alle Ansätze, die vor der eigentlichen operativen Ausführung der Workflows den Nachweis erbringen, dass eine Verletzung der Regeln nicht möglich ist. Diese Ansätze wirken präventiv, können jedoch nur für einen Teil der Kategorien von CA eine Einhaltung sicherstellen [35]. Eine Überprüfung auf „compliance by design“ erfolgt vor allem in Form von formalen Methoden, wie Modell-Checking [36]. Hierbei werden die Modelle der Geschäftsprozesse adressiert und daraufhin überprüft, ob eine Sicherstellung der CA erfolgt. Sogenannte Lebenseigenschaften [37] können aufgrund ihrer Nichtentscheidbarkeit („non-observable obligation“) zum Zeitpunkt der Datenspeicherung nicht überprüft werden [38]. Eine Überführung von Lebenseigenschaften in entscheidbare Eigenschaften ist nur durch weitere restriktive Einschränkungen möglich, da während des Prozessentwurfs (Konstruktionsebene) die erforderlichen Kontextdaten des tatsächlichen Ablaufs einer Workflow-Instanz noch nicht zur Verfügung stehen [39].

Aus Sicht des Compliance-Managements müssten für ein „compliance by design“ alle Lebenseigenschaften durch ergänzende Restriktionen in entscheidbare Eigenschaften (observable obligation) überführt werden (conservative refinement) [38]. Kontrollprozesse werden dann auf Konstruktionsebene, d. h. während des Prozessentwurfs und vor der eigentlichen operativen Ausführung in die Geschäftsprozesse integriert, wie bspw. in [6]. Durch diese „hartverdrahtete“ Verbindung von Kontroll- und Geschäftsprozessen soll sichergestellt werden, dass bei der nachgelagerten Ausführung der Workflow-Instanzen Regeln nicht verletzt werden können [40].

Der Entwurf Compliance-konformer Geschäftsprozesse erfordert in der Regel die Identifikation, Auswahl sowie Integration umfangreicher Kontrollprozesse. Dies ist bis heute eine personalintensive Tätigkeit, die sowohl unternehmensspezifisches Expertenwissen als auch eine Koordination über Abteilungsgrenzen hinweg erfordert [41-42]. Prinzipiell lässt sich damit die Einhaltung von CA zwar durchsetzen, den Vorteilen eines geringen Risikos aus Sicht des Compliance-Managements stehen jedoch erhebliche Nachteile aus Sicht des Geschäftsprozessmanagements gegenüber: Durch die Integration einer Vielzahl von Kontrollprozessen in die einzelnen Geschäftsprozesse wird zum einen die Effizienz der originären Geschäftsprozesse negativ verändert und zum anderen die Fähigkeit, Workflows flexibel an veränderte Umweltbedingungen anzupassen, erheblich eingeschränkt [33].

Um die betrieblich geforderte Flexibilität in einem gewissen Maße dennoch zu erreichen, werden in der Praxis unterschiedliche Workflow-Schemata vorgehalten, die für alle vom Unternehmen erwarteten Situationen vorab mit Kontrollen versehen und auf Compliance geprüft sind [4], [33]. Die Definition und Pflege solcher Schemata ist jedoch ebenfalls mit hohem personellem Aufwand verbunden. In einem dynamischen Unternehmensumfeld, geprägt durch vielschichtige, sich häufig ändernde CA und Geschäftsprozesse, stößt ein solches Vorgehen jedoch sowohl aus wirtschaftlicher als auch bezüglich der Beherrschbarkeit der entstehenden Komplexität schnell an Grenzen.

Weiterhin wird mit regelbasierten Systemen [43] die Sicherstellung von CA verfolgt. Dabei werden im Geschäftsprozess Punkte bestimmt, an denen erforderliche

Regeln identifiziert und erzwungen werden sollen. Basierend auf Regel-Graphen erlaubt bspw. der Ansatz von [44], Compliance-Verstöße zu identifizieren. Das vorgeschlagene System enthält bislang jedoch keine Mechanismen, um Compliance-Verstöße zu verhindern. Weiterhin ist die vorgeschlagene Modellierung der Regeln für Compliance-Verantwortliche nicht so aufbereitet, dass diese die Regel selbst definieren könnten. Somit wird die IT-Abteilung eines Unternehmens benötigt, um neue oder geänderte Compliance-Anforderungen zu integrieren. Eine Flexibilität der Geschäftsprozesse wird mit einem solchen Vorgehen stark eingeschränkt und wird den betrieblichen Anforderungen in Bezug auf Flexibilität und der damit einhergehenden Änderungshäufigkeit bislang nicht gerecht.

Weitere Ansätze, die Abweichungen von den Abläufen im Sinne einer Risikobetrachtung erfassen (z. B. [6], [45-46]), bieten bislang nur die Möglichkeit, identifizierte Schwachstellen im Geschäftsprozessmodell zu kennzeichnen und Ausweichaktivitäten auf einer fachkonzeptionellen Ebene zu definieren. Andere Ansätze, die eine modellbasierte Identifikation und Entwicklung von erforderlichen Kontrollprozessen fokussieren, nutzen wiederkehrende Kontrollprinzipien [47]. Die Auswahl und Integration der Kontrollprinzipien in Geschäftsprozesse wird allerdings in keinem der Ansätze näher thematisiert. Der Ansatz von [4] geht diesbezüglich einen Schritt weiter. So werden für bestimmte Klassen von CA entsprechende Muster für Kontrollen identifiziert und integriert. Bei diesem Ansatz wird jedoch eine Integration rein technisch vorgenommen und es bleibt unberücksichtigt, ob dies zu Prozessstillständen oder -störungen führt. Auch der Ansatz von [48] bleibt ausschließlich auf der technischen Ebene, erläutert jedoch welche Kontrollflussregeln prinzipiell vorkommen können und bildet damit eine Grundlage, um Kontrollprozesse definieren zu können.

Alle diese Ansätze betrachten Kontrollen und Geschäftsprozesse bislang entweder als technische Herausforderung oder berücksichtigen nur einen der beiden Aspekte Agilität der Geschäftsprozesse oder Einhaltung der CA mit Hilfe von Kontrollprozessen. Somit stellen die bekannten Ansätze nur eine Zwischenlösung dar, die es erlaubt, ein gewisses Maß an Flexibilität bei gleichzeitiger Einhaltung von Compliance-Regeln sicherzustellen. Je agiler Geschäftsprozesse in einem Unternehmen mit IT-Unterstützung ablaufen sollen, desto schwieriger ist es, durch heute bekannte Ansätze, eine effiziente Kontrolle zu implementieren und damit die Compliance sicherzustellen [49].

Ein allgemeiner methodischer Ansatz zur flexiblen Integration von Kontrollen in Workflows fehlt somit bislang und wird im nachfolgenden Kapitel als Ziel des Artikels präsentiert, um diesen scheinbaren Trade-off zwischen Agilität und Compliance zu reduzieren.

3 „FlexCom“ - Eine Methode zur flexiblen Integration von Kontrollen in Workflows

Um einen wesentlichen Teil der benötigten Agilität der Geschäftsprozesse bei „Compliance by design“-Ansätzen zu erhalten, bietet die Optimierung der Verbindung von Kontroll- und Geschäftsprozessen für jede einzelne Instanz eine vielver-

sprechende Möglichkeit [50]. Ein solches Vorgehen bietet die Chance, Kontextinformationen jeder einzelnen Instanz des Kontroll- und Geschäftsprozesses verwenden zu können, ohne grundsätzlich neue Theorien für die Auswahl und Integration von Kontrollprozessen entwickeln zu müssen.

Eine automatisierte Integration von Kontrollprozessen in Geschäftsprozesse erfordert es, beide Prozessarten formal definieren zu können. Für Geschäftsprozesse gibt es, wie zuvor beschrieben, Geschäftsprozessmodellierungssprachen, welche eine formale Darstellung erlauben. Kontrollprozesse können prinzipiell mit denselben Methoden und Werkzeugen wie Geschäftsprozesse modelliert und umgesetzt werden [51]. Die Realisierung der Kontrollaktivitäten, deren automatisierte Auswahl und Verbindung mit den Geschäftsprozessen wird jedoch bislang kaum diskutiert [52] und soll im Folgendem besonders im Fokus stehen.

Aufbauend auf den grundsätzlichen Überlegungen in [49] und der konzeptionellen Diskussion in [53] wird eine Ad-hoc-Integration von Kontrollprozessen in Geschäftsprozesse während der Ausführung der Workflow-Instanzen verfolgt und in ein anderes Anwendungsszenario eingebettet. Die neue modell-basierte Methode „FlexCom“ wird dabei anhand von drei Schritten erläutert und ist grob in Abbildung 1 skizziert:

1. Die Modellierung und Definition von sogenannten Referenzkontrollprozessen, welche die Art und Weise, wie Kontrollaktivitäten ausgeführt werden, repräsentieren.
2. Kontrollintegrationsparameter, welche genutzt werden, um Referenzkontrollprozesse zu identifizieren und den effektiven Einsatz basierend auf den individuellen Workflow-Instanzen zu bestimmen.
3. Eine Methode, um Kontrollaktivitäten mit Hilfe der Kontrollintegrationsparameter situationsabhängig in konkrete Instanzen von Workflows zu integrieren.

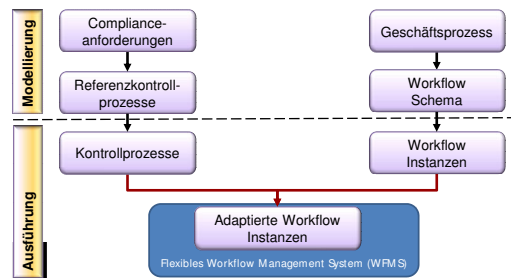


Abb. 1. Abhängigkeiten bei der Ad-hoc-Integration von Kontrollen in den Geschäftsprozess [53]

3.1 Modellierung und Definition von Referenzkontrollprozessen

Für jede CA kann mindestens ein Referenzkontrollprozess definiert werden. Dabei besteht der Referenzkontrollprozess aus verbundenen Kontrollaktivitäten, bspw. kann das Vieraugenprinzip, ein oft verwendetes Prinzip um CA abzubilden, sowohl sequentiell als auch parallel ausgeführt werden. Auch können die Kontrollaktivitäten direkt nebeneinander bzw. hintereinander ausgeführt werden oder mit weiteren Ge-

schäftsprozessaktivitäten dazwischen. Jeder dieser Referenzkontrollprozesse ist mit gängigen Geschäftsprozessmodellierungssprachen darstellbar und kann dementsprechend mit den passenden Tools modelliert werden (siehe bspw. Abbildung 3). Während die Definition und Identifikation von generellen Referenzkontrollprozessen eine offene Forschungsfrage ist und letztlich von jedem Unternehmen selbst auf dessen Bedürfnisse angepasst werden muss, kann mit einem solchen Vorgehen eine neue, eigenständige Sicht auf Kontrollprozesse erzeugt werden. Dies stellt eine prinzipielle Voraussetzung dar, um eine Automatisierung von Compliance bei Erhaltung der Agilität der Geschäftsprozesse zu ermöglichen.

3.2 Kontrollintegrationsparameter

Jede CA kann mindestens durch eine, in der Regel jedoch durch unterschiedliche Kontrollaktivitäten bzw. Referenzkontrollprozesse verfolgt werden. Dabei sind sowohl die Kontrollaktivitäten selbst als auch deren Kombination an verschiedenen Stellen des Geschäftsprozesses möglich. Wenn bspw. im Einkauf die Freigabe von Rechnungen kontrolliert werden soll, kann dies sowohl bei Eingang der Rechnung, bei der Bearbeitung der Rechnung oder bei der Verbuchung der Rechnung kontrolliert werden. Die Kontrolle ist dabei durch verschiedene Verfahren möglich, bspw. manuelle Prüfung vs. automatisierter Abgleich mit der Bestellung. Je nach gewählter Kombination differieren Effektivität und Kosten. Die Entscheidung, welche Kontrollaktivität oder Bündel von Kontrollaktivitäten verwendet werden soll, ist dabei von jedem Unternehmen individuell zu bestimmen. Auf einer allgemeinen Ebene hingegen können sogenannte Kontrollintegrationsparameter verwendet werden, um bestimmen zu können, an welchen Punkten innerhalb des Geschäftsprozesses eine Integration von Referenzkontrollprozessen und somit Kontrollaktivitäten oder Bündel von Kontrollaktivitäten grundsätzlich möglich sind.

Als Kontrollintegrationsparameter können letztlich alle Informationen des WfMS genutzt werden, wie bspw. Datenelemente, welche über Werte oder Wertebereich eine Eingrenzung vornehmen oder Bearbeiter, die eine Rolle innehaben und für die Ausführung von Aktivitäten verantwortlich sind. Insbesondere sind hierbei auch Abhängigkeiten zu beachten und abzubilden, bspw. der Bearbeiter, welcher Aktivität A durchgeführt hat, darf nicht auch Aktivität D durchführen. Weiterhin sind zeitliche Restriktionen zu beachten; auch hierbei können einzelne Zeitpunkte, wie auch Zeitintervalle von Interesse sein und schließlich die Aktivität des Workflows selbst, welche vor oder nach einem Referenzkontrollprozess durchgeführt werden müssen.

Konzeptionell bestimmen die Kontrollintegrationsparameter somit auf der syntaktischen Ebene, welche Referenzkontrollprozesse an welchen Stellen im Geschäftsprozess integrierbar sind. Damit wird nicht nur eine Stelle einzelner Referenzkontrollprozesse bestimmbar, sondern alle realisierbaren alternativen Integrationsstellen identifizierbar. Die Kontrollintegrationsparameter können somit genutzt werden, um einen modell-basierten Überblick über alle technisch möglichen Integrationspunkte im Geschäftsprozess zu erlangen.

3.3 Situationsabhängige Integration von Kontrollaktivitäten in Workflow-Instanzen

Nachdem Referenzkontrollprozess und Integrationsparameter festgelegt wurden, erfolgt nun die eigentliche Integration. Der Ansatz, Kontrollen erst während der Ausführung oder zum Zeitpunkt der Instanziierung des Workflows zu integrieren, erlaubt, mehr vorhandene Informationen einzubeziehen als zum Entwurfszeitpunkt des Workflow-Schemas zur Verfügung stehen. Die Methode FlexCom wählt in einem ersten Schritt einen Referenzkontrollprozess mit seinen festgelegten Kontrollintegrationsparametern aus und sucht nach Punkten in der Workflow-Instanz, an denen eine Kontrollintegration möglich wäre. Hierbei wird jede einzelne Aktivität inkl. aller In- und Outputs der Workflow-Instanz gelesen und auf eine Übereinstimmung mit den Kontrollintegrationsparametern untersucht (siehe Abbildung 2). Dabei können drei mögliche Situationen auftreten:

1. Es wird kein Integrationspunkt gefunden. Der Workflow erlaubt keinerlei Integration und kann somit nicht kontrolliert werden, jedoch ist der daraus resultierende Compliance-Verstoß aufzeichenbar (logbar) und könnte zu einem späteren Zeitpunkt ausgewertet und sanktioniert werden (siehe „compliance by detection“). Diese Information könnte auch genutzt werden, um einerseits daraufhinzuweisen, dass in einem solchen Workflow entweder keine compliance-relevanten Aktivitäten durchgeführt werden, der Prozess selbst fehlerhaft ist oder Referenzkontrollprozesse noch fehlen.
2. Im Workflow wird genau ein Integrationspunkt gefunden. Wird exakt ein Integrationspunkt gefunden, so kann nur diese Stelle für die Integration der Kontrolle genutzt werden und muss als verpflichtende Kontrolle integriert werden.
3. Werden mehrere Integrationspunkte gefunden, kann mit Hilfe einer Vielzahl von bekannten Heuristiken, wie bspw. der Optimierung des kritischen Pfads [54] oder einem Vorsichtsprinzip folgend, so früh wie möglich Referenzkontrollprozesse zu integrieren [55], eine Integration erfolgen. Die vorgestellte Methode lässt hierbei auch absichtlich Raum für den Einsatz anderer Verfahren oder Heuristiken, die jeweils vom betrachteten Unternehmen bestimmt werden müssen. Ohne eine Änderung an der prinzipiellen Methode ermöglicht dieser modulare Aufbau eine Anpassung an das jeweilige Unternehmen und seine strategische Ausrichtung und ist somit auch offen, um zukünftige, neue Verfahren bei Bedarf einzusetzen.

Die Integration der Referenzkontrollprozesse wird, wie in Abb. 2 beschrieben, iterativ vorgenommen. Jede Workflow-Instanz (wfi_k) wird dabei analysiert und mögliche Integrationspunkte werden identifiziert. Wird in der ersten Schleife kein Integrationspunkt mit Hilfe der Kontrollintegrationsparameter gefunden, so wird der nächste Referenzkontrollprozess verwendet und über alle Kontrollintegrationsparameter gesucht. So setzt sich der Algorithmus fort bis alle Referenzkontrollprozesse (r_i) inklusive aller ihrer Kontrollintegrationsparameter (ki_j) abgearbeitet wurden.

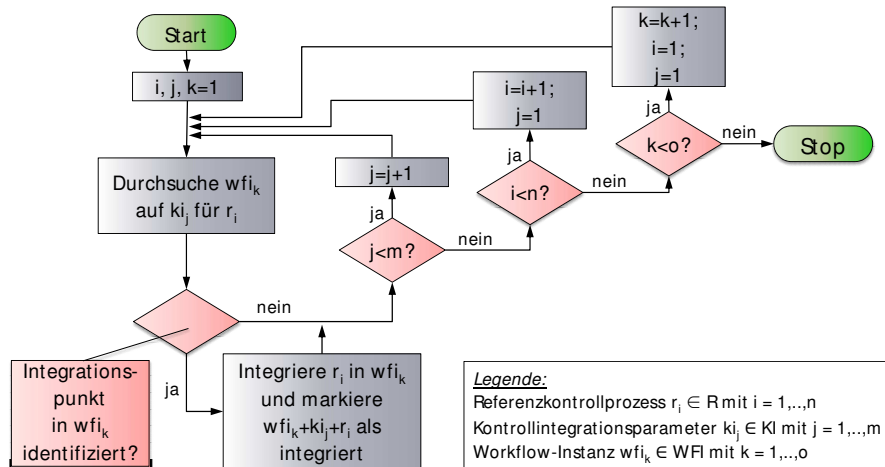


Abb. 2. Programmablaufplan zur automatischen Integration von Referenzkontrollprozessen in Workflows

Sollte allerdings eine Übereinstimmung gefunden werden, so wird dieser Integrationspunkt in die Menge aller möglichen Integrationspunkte für den Referenzkontrollprozess aufgenommen. Wurden alle möglichen Integrationspunkte bestimmt, gilt es, sich für einen Integrationspunkt zu entscheiden. Aus betriebswirtschaftlicher Sicht ist es i. d. R. nicht sinnvoll, alle möglichen Kontrollen redundant an allen theoretisch möglichen Stellen zu integrieren. Eine Auswahl, resp. Einschränkung des Zustandsraumes ist daher erforderlich. Diese erfolgt aktuell durch die Integration von Kontrollaktivitäten möglichst früh im Geschäftsprozess (siehe Abbildung 2). Das mit der Auswahl verbundene betriebswirtschaftliche Optimierungspotential stellt eine offene Forschungsfrage dar, die in diesem Beitrag jedoch nicht weiter adressiert wird. Eine Nutzung fortgeschrittener Auswahlverfahren in der entwickelten Methode sollte problemlos möglich sein. Die Integration selbst entspricht dann dem Einfügen eines Subprozesses und erfolgt, wenn alle syntaktischen Bedingungen erfüllt sind, welche zu keiner Verletzung des gesamten Geschäftsprozessmodells führen wie in [56] ausführlich beschrieben.

4 Der Prototyp KitCom zur flexiblen Integration von Kontrollen in Workflows

Der Design-Science Forschungsmethodik [9] folgend wird mit dem Prototyp KitCom die prinzipielle Funktionsfähigkeit des Artefakts, der in Kapitel drei beschriebenen Methode, evaluiert. Die Funktionsweise des Prototyps basiert dabei auf dem WfMS Aristaflo [57] und erweitert dieses um die Klassen der Methode FlexCom. Ein besonderes Merkmal dieses WfMS stellt die Berücksichtigung von Ad-hoc-Änderungen dar [58]. Jede einzelne Instanz lässt sich auch noch während der Ausführung ändern und die strukturelle Konsistenz wird von dem WfMS sichergestellt. Diese Besonder-

heit von AristaFlow wurde ergänzt um bspw. eine Datenbank für die Referenzkontrollprozesse oder einer Modellierungsoberfläche, um die Kontrollen zu modellieren und deren Kontrollintegrationsparameter festlegen zu können.

Die Arbeitsweise von KitCom wird im Folgenden anhand eines exemplarischen und aus didaktischen Gründen einfach gehaltenen Beispiels erläutert: Ein Geschäftsprozess im Einkauf, resp. der Teilprozess zur Freigabe von Rechnungen, wird dabei verwendet und als Kontrolle sollen Rechnungen mit einem Wert über 5000€ zusätzlich überprüft werden. In einem ersten Schritt hat der Prozessdesigner den Geschäftsprozess zu definieren und in Aristaflow anzulegen, wie dies in Abbildung 3 dargestellt ist:

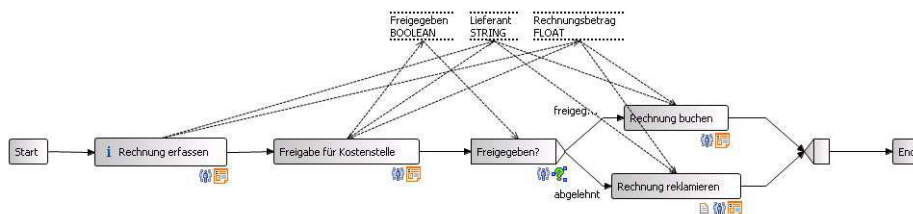


Abb. 3. Beispielhafter Workflow mit dem AristaFlow Process Template Editor generiert

Im nächsten Schritt hat der Compliance-Beauftragte den Referenzkontrollprozess zu definieren (Abbildung 4). Hierfür wurde eine extra Sicht in den AristaFlow Process Template Editor integriert, welche es neben der Definition von Referenzkontrollprozessen erlaubt, einen Überblick über alle im WfMS befindlichen Referenzkontrollprozesse zu erhalten.

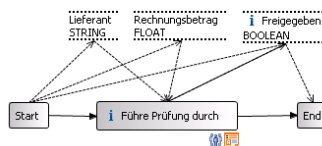


Abb. 4. Beispielhafter Referenzkontrollprozess in KitCom

Anschließend werden die Kontrollintegrationsparameter festgelegt, welche für das Auslösen einer Instanz des Referenzkontrollprozesses zuständig sind (Abbildung 5). Die Auswahl eines Kontrollintegrationsparameters kann dabei über ein neues Eingabefeld erfolgen und um beliebige weitere Integrationsparameter ergänzt werden, wie bspw. eine Einschränkung das bestimmte Sachbearbeiter die Kontrollaktivitäten nicht ausführen oder das Kontrollaktivitäten nur in einem bestimmten Zeitraum ausgeführt werden dürfen.

Funktion	Name	Operation	Wert
Daten	Rechnungsbetrag	>	5000

Abb. 5. Kontrollintegrationsparameter in KitCom

Nachdem alle drei Teile definiert wurden funktioniert KitCom wie folgt: Als erstes startet die Instanz eines Workflows ohne Kontrolle (Abbildung 6).

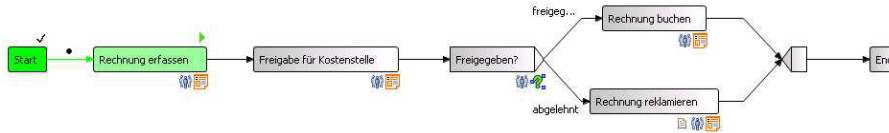


Abb. 6. Gewöhnliche Ausführung einer Workflow Instanz in der Software AristaFlow

Übersteigt der Betrag 5000€ wird eine Instanz der Referenzkontrolle, wie in Abbildung 7 zu sehen, als Subprozess Ad-hoc integriert. Die Integration erfolgt dabei durch eine neue Aktivität „Kontrolle“, welche eine Instanz des Referenzkontrollprozesses (Abbildung 4) automatisch in Form eines Subprozessaufrufs integriert. Der Workflow wird intern durch KitCom markiert, so dass erkenntlich ist, welcher Referenzkontrollprozess bereits integriert wurde. Somit können Mehrfachintegrationen verhindert werden und eine Rückverfolgbarkeit ist gegeben, welche Referenzkontrollprozesse, an welchen Stellen und in welchen Workflow-Instanzen vorgenommen wurden.

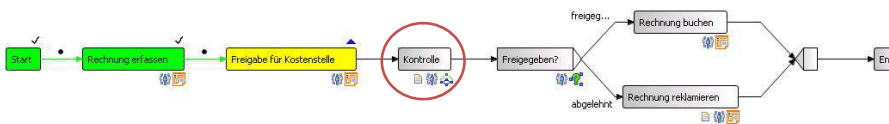


Abb. 7. Automatische Integration von einer Instanz der Referenzkontrolle in die Workflow-Instanz durch KitCom

In einem nächsten Schritt werden die Kontrollen und Geschäftsprozesse aus [59] als Referenzgröße verwendet und mit Hilfe des Prototypen modelliert und getestet. Sollten, wie zu erwarten, dabei alle Referenzkontrollprozesse in die Workflows automatisch integriert werden, ist die Funktionsfähigkeit in Bezug auf statische Workflow-Instanzen evaluiert. In einem weiteren Schritt werden die Geschäftsprozesse verändert und erneut wird wieder die Anzahl der integrierbaren Kontrollprozesse untersucht und damit die Evaluierung des Artefakts in Bezug auf dynamische Workflow-Instanzen abgeschlossen. Obwohl diese zwei letzten Schritte der generellen Evaluierung des Artefakts bislang noch ausstehen, konnte bereits die generelle Funktionsfähigkeit des Prototypen KitCom gezeigt werden. Beispielhaft wurden mehrere Workflowschemata, Kontrollprozesse und Kontrollintegrationsparametern aus dem Bereich der Rechnungsverarbeitung simultan getestet und automatisch bei Bedarf integriert. Die damit erreichten positiven Ergebnisse bieten ein valides Indiz für eine erfolgreiche Weiterentwicklung der Methode.

5 Offene Forschungsfragen und weitere Schritte

Der Artikel adressiert das Problem der Compliance-Konformität von agilen Geschäftsprozessen. Gängige Methoden basieren auf Veränderungen des Workflow-Schemas und können somit nur eingeschränkt die eigentlich gewünschte Agilität unterstützen. Die vorgestellte Methode FlexCom verringert diese Lücke, indem die je-

weiligen Workflow-Instanzen sozusagen „on-the-fly“ mit Kontrollen angereichert werden. Die Besonderheit jeder einzelnen Instanz wird berücksichtigt und die notwendigen Kontrollaktivitäten lassen sich so für einzelne Instanzen optimieren. Die neu vorgestellte Methode FlexCom wurde dabei nicht nur theoretisch beschrieben, sondern bereits praktisch implementiert und die generelle Funktionsfähigkeit mit Hilfe des Prototypen KitCom positiv evaluiert.

Die Lesson Learned bei der Findung der neue Methode lassen sich wie folgt zusammenfassen: Die Integration von Kontrollprozessen in Workflow-Instanzen erlauben es, mehr Informationen einzubeziehen als dies bei der klassischen Integration in Workflow-Schemata der Fall ist. Dennoch sind auch bei einer solch automatisierten Integrationsmethode weiterhin Abstimmungsprozesse zwischen den Compliance- und Geschäftsprozessverantwortlichen unumgänglich – allerdings nicht mehr bei jeder Änderung: Der beschriebene Ansatz erlaubt in soweit eine Automatisierung, wie die Definition der Referenzkontrollprozesse und insbesondere der Kontrollintegrationsparameter den Compliance-Anforderungen des Unternehmens entsprechen und modelliert wurden. Gibt es generelle Änderungen, sowohl durch neue Gesetze oder grundsätzlich geänderte Workflowschemata, erfordert dies insbesondere eine Anpassung der Kontrollintegrationsparameter durch den Compliance-Verantwortlichen. Der Abstimmungsprozess zwischen Compliance- und Geschäftsprozessverantwortlichen wird auch auf absehbare Zeit eine manuelle Tätigkeit der beteiligten Abteilungen bleiben, welche jedoch dank der neuen Methode nicht mehr bei jeder Änderung von einzelnen Workflow-Instanzen notwendig ist, sondern generell für Workflowschemata nur noch einmal überprüft werden muss. Anschließend können Änderungen an den Workflow-Instanzen ohne Umgehung der Kontrollaktivitäten vorgenommen werden und behalten somit einen gewissen Grad der geforderten betriebswirtschaftlichen Flexibilität.

Zur vollständigen Realisierung der Vorteile agiler Geschäftsprozesse bei gleichzeitiger Compliance-Konformität ist eine Automatisierung von Kontrollen notwendig und wirft noch weitere Fragen auf: So ist bspw. insbesondere die Definition von allgemeingültigen Referenzkontrollprozessen bislang ungelöst. Auch die Auswahlproblematik für effektive und effiziente Kontrollen, insbesondere bei einer automatisierten Integration, gilt es zu untersuchen. Weiterhin sollte, auf dem Weg zu einer optimalen Unterstützung agiler Geschäftsprozesse, auch die Ableitung auf das Unternehmen angepasster Kontrollen aus gesetzlichen Vorgaben Ziel weiterer Untersuchungen sein. Mit der vorgestellten Methode ist jedoch eine technische Grundlage gelegt und ein Schritt hin zur Automatisierung von Compliance vollzogen.

Literatur

1. Teubner, A., Feller, T.: Informationstechnologie, Governance und Compliance. *Wirtschaftsinformatik* 50, 400-407 (2008)
2. Rüter, A.: *IT Governance in der Praxis*. Springer, Berlin Heidelberg (2010)
3. Cannon, J., Byers, M.: Compliance deconstructed. *Queue* 4 (7), 30-37 (2006)
4. Sadiq, S., Governatori, G., Namiri, K.: Modeling control objectives for business process compliance. In: *LNCS*, Vol. 4714, pp. 149-164. Springer, Berlin Heidelberg (2007)

5. van der Aalst, W., van Hee, K.M.: Workflow management: models, methods, and systems. MIT press (2004)
6. Lu, R., Sadiq, S., Governatori, G.: Compliance aware business process design. In: Proceedings of the 2007 International Conference on Business process management, pp. 120-131. Springer, Berlin Heidelberg (2007)
7. Künzle, V., Reichert, M.: Herausforderungen auf dem Weg zu datenorientierten Prozess-Management-Systemen. EMISA Forum 29 (2), 9-24 (2009)
8. Horstmann, J.: Operationalisierung und Unternehmensflexibilität: Entwicklung einer umwelt- und unternehmensbezogenen Flexibilitätsanalyse. DUV, Wiesbaden (2007)
9. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems research. MIS Quarterly 28 (1), 75-105 (2004)
10. Schonberg, H., Mans, R., Russell, N., Mulyar, N., van der Aalst, W.: Process flexibility: A survey of contemporary approaches. In: LNBIP, Vol. 10, pp. 16-30. Springer, Berlin Heidelberg (2008)
11. Heinrich, B., Klier, M., Zimmermann, S.: Automatisierte Modellierung, Umsetzung und Ausführung von Prozessen – Ein Web Service-basiertes Konzept. 10th International Conference on Wirtschaftsinformatik, pp. 88-98 (2011)
12. Rinderle, S., Reichert, M., Dadam, P.: Correctness criteria for dynamic changes in workflow systems--a survey. Data Knowledge Engineering 50, 9-34 (2004)
13. Fowler, M.: UML konzentriert. Pearson Education (2004)
14. Scheer, A.W.: Aris-Business Process Modeling. Springer, Berlin Heidelberg (2000)
15. White, S.A., Miers, D.: BPMN modeling and reference guide: understanding and using BPMN. Future Strategies Inc. (2008)
16. Rosemann, M., Recker, J., Indulska, M., Green, P.: A study of the evolution of the representational capabilities of process modeling grammars. In: LNCS, Vol. 4001, pp. 447-461. Springer, Berlin Heidelberg (2006)
17. Franke, J., Gomber, P., Beimborn, D., König, W., Wagner, H.T., Weitzel, T.: Business-IT-Alignment in deutschen Banken: Die Rolle von IT für die Geschäftsprozesse und die Wettbewerbsposition deutscher Kreditinstitute. Books on Demand, Norderstedt (2008)
18. Sackmann, S., Strüker, J.: Electronic Commerce Enquete 2005: 10 Jahre E-Commerce: Eine stille Revolution in deutschen Unternehmen. Konradin-IT, Leinfelden (2005)
19. Erl, T.: SOA: Principles of Service Design. Prentice Hall, Upper Saddle River, NJ (2008)
20. Krafzig, D., Banke, K., Slama, D.: Enterprise SOA: service-oriented architecture best practices. Prentice Hall PTR, Upper Saddle River, NJ (2005)
21. Richter, J.P., Haller, H., Schrey, P.: Serviceorientierte Architektur. Informatik-Spektrum 28, 413-416 (2005)
22. Khalaf, R.: Supporting business process fragmentation while maintaining operational semantics: a BPEL perspective. dissertation.de, Berlin (2008)
23. Müller, J.: Workflow-based integration: Grundlagen, Technologien, Management. Springer, Berlin Heidelberg (2005)
24. Mahr, E.: www.it-director.de/nc/home/newsdetails/article/dunkle-prozesse-sichtbar-machen.html
25. Dadam, P., Reichert, M., Kuhn, K.: Clinical Workflows - The Killer Application for Process-oriented Information Systems?. In: Proceedings of the BIS'00, pp. 36-59 (2000)
26. Kopp, C.: Vorgehensmodell für die Einführung von IT-Governance in mittelständischen Unternehmen: Vorgehensmodell für die Einführung mit den Referenzmodellen COBIT, ITIL und Val-IT. GRIN (2009)

27. Göbel, R.: *Interne Überwachung mit Hilfe von Auswahlverfahren: Möglichkeiten zur Steigerung der Wirtschaftlichkeit interner Überwachungssysteme*. Duncker & Humblot, Berlin (1990)
28. Baetge, J.: *Überwachung*. In: Bitz, M., Dellmann, K., Domsch, M., Egner, H. (eds.): *Vahlens Kompendium der Betriebswirtschaftslehre*. 3. Auflage, Vahlens, München (1993)
29. Knapp, E.: *Interne Revision und Corporate Governance: Aufgaben und Entwicklungen für die Überwachung*. Erich Schmidt, Berlin (2008)
30. El Kharbili, M., Stein, S., Markovic, I., Pulvermüller, E.: *Towards a framework for semantic business process compliance management*. In: *GRCIS'08 Workshop at CAiSE'08 - Governance, Risk and Compliance: Applications in IS*, pp. 1-15. Citeseer (2008)
31. Agrawal, R., Johnson, C., Kiernan, J., Leymann, F.: *Taming compliance with Sarbanes-Oxley internal controls using database technology*. In: *22nd ICDE*, p. 92. IEEE (2006)
32. Mulo, E., Zdun, U., Dustdar, S.: *Monitoring web service event trails for business compliance*. In: *International Conference on Service-Oriented Computing and Applications (SOCA)*, pp. 1-8. IEEE (2009)
33. Schumm, D., Leymann, F., Ma, Z., Scheibler, T., Strauch, S.: *Integrating Compliance into Business Processes: Process Fragments as Reusable Compliance Controls*. *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI'10)* (2010)
34. van der Aalst, W., de Medeiros, A. K.: *Process mining and security: Detecting anomalous process executions and checking process conformance*. *Electronic Notes in Theoretical Computer Science* 121, 3-21 (2005)
35. Leucker, M., Schallhart, C.: *A brief account of runtime verification*. *Journal of Logic and Algebraic Programming* 78, 293-303 (2009)
36. Schmitz, R., Kriha, W.: *Sichere Systeme - Konzepte, Architekturen und Frameworks*. Springer, Berlin Heidelberg (2009)
37. Schneider, F.B.: *Enforceable security policies*. *ACM Transactions on Information and System Security (TISSEC)* 3, 30-50 (2000)
38. Pretschner, A., Hilty, M., Basin, D.: *Distributed usage control*. *Communications of the ACM* 49 (9), 39-44 (2006)
39. Müller, G., Terzidis, O.: *IT-Compliance und IT-Governance*. *Wirtschaftsinformatik* 50, 341-343 (2008)
40. Rinderle-Ma, S., Ly, L.T., Dadam, P.: *Business Process Compliance*. *EMISA Forum* 28 (2), 24-29 (2008)
41. Eversberg, A.: *Der Einfluß von Electronic Data Interchange auf die Kontrolle von Geschäftsprozessen*. In: Berndt, R. (ed.): *Business reengineering: effizientes Neugestalten von Geschäftsprozessen*. Springer, Berlin Heidelberg (1997)
42. Gehrke, N.: *Zur Automatisierung von Revisionsdienstleistungen zwecks Unternehmensüberwachung—Ein Überblick*. In: *LNI*, Vol. P-154, pp. 3760-3774. GI, Bonn (2009)
43. von Halle, B.: *Business rules applied: building better systems using the business rules approach*. John Wiley & Sons, New York (2002)
44. Ly, L., Rinderle-Ma, S., Knuplesch, D., Dadam, P.: *Monitoring business process compliance using compliance rule graphs*. In: *On the Move to Meaningful Internet Systems: OTM 2011, LNCS*, Vol. 7044, pp. 82-99. Springer, Berlin Heidelberg (2011)
45. Brabaender, E., Ochs, H.: *Analyse und Gestaltung prozessorientierter Risikomanagementsysteme mit Ereignisgesteuerten Prozessketten*. In: *Proceedings of the 1st GI-Workshop on Business Process Management with Event-Driven Process Chains (EPK 2002)*, pp. 17-35 (2002)
46. zur Muehlen, M., Rosemann, M.: *Integrating risks in business process models*. In: *Proceedings of the ACIS 2005* (2005)

47. Schaad, A., Moffett, J.D.: A framework for organisational control principles. In: Proceedings of the 18th Annual Computer Security Applications Conference, pp. 229-238. IEEE (2002)
48. Awad, A., Weske, M.: Visualization of compliance violation in business process models. LNBIP, Vol. 43, pp. 182-193. Springer, Berlin Heidelberg (2010)
49. Kittel, K., Sackmann, S.: Gaining Flexibility and Compliance in Rescue Processes with BPM. In: ARES Conference - Workshop on "Resilience and IT-Risk in Social Infrastructures" (RISI 2011), pp. 639-644. IEEE (2011)
50. Sackmann, S., Hofmann, M., Kühnel, S.: Return on Controls Invest – Ein Ansatz zur wirtschaftlichen Spezifizierung von Internen Kontrollsystemen. HMD - Praxis der Wirtschaftsinformatik 50, (2013)
51. Pitthan, J., Philipp, M.: Einsatz von Petri-Netzen für die Aufnahme, Dokumentation und Analyse Interner Kontrollsysteme im Rahmen der Jahresabschlußprüfung. In: Petri-Netze zur Modellierung verteilter DV-Systeme–Erfahrungen im Rahmen des DFG-Schwerpunktprogramms "Verteilte DV-Systeme in der Betriebswirtschaft", pp. 87–104 (1997)
52. Abdullah, S.N., Indulska, M., Sadiq, S.: A study of compliance management in information systems research. In: 17th European Conference on Information Systems, pp. 1-10. ECIS Standing Committee (2009)
53. Kittel, K., Sackmann, S.: Flexible Controls for Compliance in Catastrophe Management Processes. In: Tagungsband der Multikonferenz Wirtschaftsinformatik (MKWI'12), pp. 1675-1687. GITO, Berlin (2012)
54. Verworn, B.: Projektplanung während der frühen Phasen. In: Herstatt, C., Verworn, B. (eds.): Management der frühen Innovationsphasen: Grundlagen - Methoden - neue Ansätze. Gabler, Wiesbaden (2003)
55. Flieder, K.: Geschäftsprozessmanagement – Geschäftsprozesse mit dem virtuellen Reißbrett steuern. WING-business 39 (3), 35-38 (2006)
56. Rinderle, S., Reichert, M., Dadam, P.: On Dealing with Structural Conflicts between Process Type and Instance Changes. In: Desel, J., Pernici, B., Weske, M. (eds.): Business Process Management, LNCS, Vol. 3080, pp. 274-289. Springer, Berlin Heidelberg (2004)
57. Aristaflow GmbH, <http://www.aristaflow.com/>
58. Dadam, P., Reichert, M., Rinderle, S., Jurisch, M., Acker, H., Göser, K., Kreher, U., Lauer, M.: ADEPT2 - Next Generation Process Management Technology. In: Proceedings Fourth Heidelberg Innovation Forum. D.punkt (2007)
59. Namiri, K.: Model-driven management of internal controls for business process compliance. PhD thesis, Universität Fridericiana zu Karlsruhe (2008)

Risikomodellierung in strategischen Liefernetzwerken für hybride Wertschöpfung

Holger Schrödl¹, Laura Geier², Matthias Geier², und Paulina Simkin²

¹ Otto-von-Guericke-Universität Magdeburg, Lehrstuhl für Wirtschaftsinformatik I,
Magdeburg, Germany
holger.schroedl@ovgu.de

² Universität Augsburg, Lehrstuhl für Wirtschaftsinformatik und Systems Engineering,
Augsburg, Germany
{laura.geier.de,mayutamano}@googlemail.com, paulina.simkin@gmx.de

Abstract. Im Sachgütermarkt zeichnet sich zunehmend der Trend von der einzelnen Individualleistung hin zur individualisierten Massenfertigung ab. Gegenläufig hierzu verläuft die Ausweitung von hybriden Produkten, welche gerade in der Individualisierung ihre Stärke zeigen und sich aus Sach- und Dienstleistungen zusammensetzen, die kundenspezifische Problemlösungen verfolgen. Dabei können diese Teilleistungen von verschiedenen Lieferanten stammen. Diese werden bei einem Produzenten gebündelt und von dort aus in Form eines einzelnen Produktes an den Kunden geliefert. Durch die große Anzahl heterogener Lieferanten innerhalb des Zulieferungsnetzwerkes sowie den spezifischen Merkmalen hybrider Leistungsbündel ist es für den fokalen Lieferanten eine Herausforderung, seine Lieferanten nach Risikoeigenschaften zu bewerten und daraufhin eine optimale Auswahl zu treffen. Wie eine an die Situation angepasste Risikobewertung zustande kommt, wird im vorliegenden Beitrag anhand eines Modellierungsansatzes der Risikobewertung im Liefernetzwerk hybrider Leistungsbündel dargestellt. Der Modellierungsansatz wird mittels eines dafür entwickelten Prototypen simuliert.

Keywords: hybride Wertschöpfung, Risikomodellierung, Liefernetzwerke, Supply Chain Management

1 Einführung

Als hybrides Produkt wird ein Leistungsbündel bezeichnet, welches aus aufeinander abgestimmten Produkten und Dienstleistungen besteht und als besonderes Kennzeichen die Integration der Teilleistungen aufweist, um ein Kundenproblem zu lösen [1]. Dadurch kann der Wert hybrider Produkte die Summe der Werte der einzelnen Teilleistungen übersteigen [2]. Mit einem solchen integrierten Portfolio ist es Unternehmen möglich, sich aus dem Markt hervor zu heben [2], höhere Margen zu generieren [3] und den Aufbau von längerfristigen und intensiveren Kundenbindungen zu fördern [4]. Außerdem kann die Leistungsfähigkeit des Produktes durch die individuelle

Anpassung an die Kundenbedürfnisse gesteigert [5] und eine höhere Wertschöpfung sowohl für den Kunden als auch für den Produzenten generiert werden [6]. An einem hybriden Leistungs Bündel ist allerdings meist nicht nur ein Unternehmen als anbietendes Unternehmen beteiligt, sondern oft ein ganzes Unternehmensnetzwerk aus autonomen Unternehmen, die einen Beitrag zu dem hybriden Leistungs Bündel leisten. Dieses Netzwerk kann sich bereits im Rahmen des Entwicklungsprozesses der hybriden Produkte bilden. Die Entwicklung des hybriden Leistungs Bündels kann aber auch von einem anbietenden Unternehmen alleine durchgeführt und das Netzwerk erst zum Beginn des Herstellungsprozesses aufgebaut werden [7]. Unabhängig von der Entstehung des Netzwerkes ist meist eine große Anzahl an Lieferanten und Vorlieferanten involviert. Jeder dieser Teilnehmer bringt ein Risiko mit in das Netzwerk, wodurch sich die Risikobewertung der einzelnen Lieferketten verändert. Je größer und verzweigter das Netzwerk ist, desto komplexer ist auch das damit verbundene Risikomanagement. Ein weiterer Aspekt sind die spezifischen Eigenschaften hybrider Leistungs Bündel. Die enge, oft untrennbare Integration der Teilkomponenten im Leistungs Bündel sowie die enge Verzahnung zwischen dem hybriden Leistungs Bündel und Teilprozessen auf Kundenseite bieten neue Herausforderungen für etablierte Managementprozesse. So konnte beispielsweise gezeigt werden, dass klassische Risikomanagementmethoden in der hybriden Wertschöpfung nur eingeschränkt eingesetzt werden können [8]. Die zentrale Frage dieses Beitrags ist: wie kann dieses Risiko systematisch erfasst werden, um dem Kunden gegenüber eine größtmögliche Sicherheit bei der Leistungserbringung zu bieten? Zur Berechnung dieses Risikos wurde ein Modell zur Risikobewertung aufgestellt, welches in einem Software-Prototypen realisiert wurde. Dieser Prototyp wurde für eine Simulation unterschiedlicher Risikoszenarien verwendet. Der Beitrag ist wie folgt aufgebaut: In Kapitel 2 wird der Forschungshintergrund zur hybriden Wertschöpfung und dem Lieferantenmanagement in strategischen Wertschöpfungsnetzwerken dargestellt. Kapitel 3 beschäftigt sich mit den Methoden der Lieferantenbewertung und dem Aufstellen eines angepassten Entscheidungskataloges. In Kapitel 4 erfolgen die Modellierung der Risikobewertung und ihre beispielhafte Anwendung. Kapitel 5 umfasst Erläuterungen zum Softwareprogramm sowie das zu Grunde liegende Datenmodell. In Kapitel 6 wird die Arbeit zusammengefasst und weiterer Forschungsbedarf aufgezeigt.

2 Forschungshintergrund

2.1 Hybride Wertschöpfung

Vom Brocke spricht von einem hybriden Sachverhalt, der als “zielgerichtetes System beschrieben werden [kann], dessen Systemzweck nach spezifischen Regeln durch alternative Teilsysteme auf unterschiedliche Art erfüllt wird“ [9]. Der hybride Sachverhalt manifestiert sich über die drei Merkmale Heterogenität (es lassen sich heterogene Teilsysteme unterschiedlicher Art unterscheiden), Konkurrenz (ein Zweck des Gesamtsystems kann durch konkurrierende Teilsysteme erfüllt werden) und Koexistenz (heterogene, um die Erfüllung eines Zwecks konkurrierenden Teilsysteme blei-

ben im hybriden System erhalten; siehe auch [5]). Mit dem Begriff „hybrides Leistungsbündel“ sind nun im Folgenden aufeinander abgestimmte Produkte und Dienstleistungen gemeint, welche die Lösung eines Kundenproblems zum Ziel haben [2], [10-11]. So können beispielweise klassische Produkte mit intangiblen Leistungen wie Service Level Agreements, Verfügbarkeitsgarantien, Finanzierungen oder Leistungsgarantien gebündelt werden [12] und damit zu hybriden Leistungsbündeln wie beispielsweise ein Performance-Contracting zur Erbringung einer spezifischen industriellen Leistung führen. Hybride Wertschöpfung in der Folge umfasst alle Maßnahmen und Tätigkeiten im Zusammenhang mit der Entwicklung, Erstellung und Konsumierung des hybriden Leistungsbündels [5]. Die Ausprägungen hybrider Wertschöpfung werden sehr heterogen betrachtet. Sawhney et. al. klassifizieren die Bestandteile hybrider Wertschöpfung in die vier Dimensionen Angebote (offerings), Kunden (customers), Prozesse (processes) und Wertschöpfungsbereich (presence) [13]. Ein Beispiel für ein hybrides Leistungsbündel ist das iPhone. Hier wird im Normalfall nicht nur das Produkt in Form eines Mobiltelefons verkauft, sondern auch der Vertrag, die Beratung und der Service. Aber auch die bei Sawhney vorkommenden Kriterien wie die Marke, das Netzwerk und die vom Hersteller angebotenen Softwarelösungen sind Teil des hybriden Leistungsbündels.

2.2 Risikobewertung im strategischen Liefernetz für hybride Wertschöpfung

Risikomanagement in Liefernetzen ist seit einigen Jahren ein sehr intensiv diskutiertes Thema [14-15]. Auch wurde dem Thema Risikomanagement in Liefernetzen bereits ein umfassendes Instrumentarium zur Seite gestellt [16]. Schwerpunkt all dieser Bemühungen ist die klassische Produktindustrie. Ansätze, das Thema Risikomanagement auch für Dienstleistungen zu adressieren, sind teilweise vorhanden [17-18]. Für die Integration von Sach- und Dienstleistungen in hybriden Leistungsbündeln fehlt eine solche Diskussion sowohl in der wissenschaftlichen Literatur als auch in der Praxis. Durch die steigende Bedeutung hybrider Leistungsbündel in einer Vielzahl von Märkten als Differenzierungsmerkmal gegenüber Mitbewerbern [19] bei gleichzeitiger zunehmender Fragmentierung von Teilleistungsanbietern ist die Risikobetrachtung in der hybriden Wertschöpfung ein zentraler Bestandteil einer nachhaltigen, langfristigen Unternehmensstrategie.

Für den Begriff Risiko existieren mehrere Definitionen. Eine davon betitelt Risiko als „Gefahr eines Verlustes oder eines Schadens[, welcher] durch das Misslingen der Leistungen, das auf nicht beeinflussbare oder erwartete Ereignisse zurückzuführen ist[, entsteht]“ [20]. Risiko kann als Wahrscheinlichkeit gesehen werden, dass ein bestimmtes ungünstiges Ereignis während einer festgelegten Zeit eintritt, oder aus einer Herausforderung heraus resultiert. Dies kann aufgeteilt werden in zwei Risikotypen (siehe Tabelle 1):

Tabelle 1. Risikotypen

Arten von Risiko	Diese können zum Beispiel strategisch, operational, finanziell, gesetzlich oder beschaffungsabhängig sein [20].
Arten von Verlusten	Diese können Auswirkungen wie beispielsweise finanziell, leistungsbezogen, sozial oder zeitlicher Art haben [21].

Im Fall von Liefernetzwerken umfasst eine Risikobetrachtung eine unbeschränkte Anzahl an Faktoren, welche die liefernden Unternehmen betreffen. Um diese Mannigfaltigkeit an Kriterien systematisch nutzbar zu machen, bedarf es eines entsprechenden Verfahrens. Hierzu bedient man sich im konkreten Fall der Bewertung von Lieferanten(-ketten) einer Reihe von Kriterien, welche relevante Risikoelemente operationalisiert und einem eindeutig definierten Kalkulationsschema zugänglich macht.

Es gibt eine Vielzahl an etablierten Verfahren für die Lieferantenauswahl und -bewertung wie beispielsweise lexikografische Regeln [22], Clusterbildung [23], Data Envelopment Analysis (DEA) [24] oder Min-Max Ansätze [25]. Allerdings wurde für die hybride Wertschöpfung bisher noch keine umfängliche Eignungsprüfung der Anwendbarkeit der bekannten Methoden unternommen. Eine Überprüfung für die Methoden Preis-/Entscheidungsanalyse und Punktbewertungsverfahren wurde in Schrödl & Geier [8] durchgeführt mit dem Ergebnis, dass das Punktbewertungsverfahren eingeschränkt geeignet für die Anwendung in der hybriden Wertschöpfung ist. Für die Sicherstellung von Standards bei der Lieferantenbewertung, ist eine Methode nötig, die Lieferanten separiert, welche die Auswahlkriterien, beziehungsweise einen bestimmten Schwellenwert für diese, nicht erfüllen [15].

Die Herausforderung in der hybriden Wertschöpfung umfasst die Identifikation der relevanten Risikoelemente sowie deren spezifische Operationalisierung. Während sich in der Wertschöpfung klassischer Produkte oder isolierter Dienstleistungen die Risikoelemente klar definieren lassen und sich in qualitative und quantitative Faktoren aufteilt, sind in der hybriden Wertschöpfung die Charakteristika des hybriden Leistungsbündels von zentraler Bedeutung. Es ist von elementarer Bedeutung, dass die Auswahl der relevanten Risikoelemente auf die jeweilige Beschaffungssituation optimal angepasst ist. So zeigen beispielsweise Aissaoui et al. [15] angepasste Auswahlkriterien für unterschiedliche Beschaffungsszenarien. Ein weiteres Beispiel eines solchen angepassten Kriterienkatalog findet sich in Weber et al. für die Just-in-Time-Produktion [26]. Darum gilt es, die spezifischen Charakteristika hybrider Leistungsbündel in Risikofaktoren zu transferieren und diese dann als Berechnungsgrundlage für ein Risikomodell quantifizierbar zu machen. Ansätze hierzu liefert zum einen die Klassifikation des Nutzungsangebots hybrider Leistungsbündel. Das Nutzungsangebot kann dabei in drei unterschiedliche Leistungserbringungsformen unterteilt werden: Funktionsorientierung (der Anbieter gewährleistet eine bestimmten Funktionsfähigkeit), Nutzungsorientierung (der Anbieter gewährleistet eine bestimmte Verfügbarkeit unter Integration von Kundenprozesses in das Leistungsangebot) und Ergebnisorientierung (der Anbieter gewährleistet ein bestimmtes Produktionsergebnis und übernimmt dabei weitere Risiken wie beispielsweise Bedienungsrisiken) [27]. Ergänzend sind sieben Kriterien identifizierbar, die charakteristisch für hybride Leistungs-

bündel sind und daher einen weiteren Ansatz zur Operationalisierung von Risiken bietet: Art des Kundennutzens, Umfang des Leistungsangebots, Anzahl/Heterogenität der Teilleistungen, Grad der technischen Integration, Grad der Integration in die Wertschöpfungsdomäne des Kunden, Grad der Individualisierung und zeitliche Dynamik/Veränderlichkeit der Leistungserbringung [27]. Weitere Beispiele für solche Entscheidungskriterien finden sich bei Heyder et. al [28], Müssigmann [29] und Pousttchi et. al [30]. Welche Kriterien nun im Detail für ein Unternehmen von Nutzen und entscheidungsrelevant sind, ist eine Einzelfallentscheidung und kann nicht per se festgelegt werden. Zum gegenwärtigen Zeitpunkt sind keine Arbeiten bekannt, die eine gesicherte Aussage über eine relevante Auswahl von Kriterien im Bereich der hybriden Wertschöpfung zulässt. Die aktuell bekannte umfangreichste, konsolidierte Darstellung spezifischer Kriterien für die Risikobewertungen findet sich bei Schrödl & Geier [8]. Für die Lieferantenbewertung bei hybriden Leistungsbündeln empfiehlt sich eine Kriterienliste, die Dimensionen wie Preis, Qualität oder Zuverlässigkeit beinhaltet und darüber hinaus Kriterien zur Klassifizierung des hybriden Leistungsbündels umfasst. Dadurch lassen sich Eigenschaften des Leistungsbündels manifestieren und damit eine Auswahl und Gewichtung entsprechender Kriterien erzielen.

3 Lieferantenbewertung

Die Lieferantenbewertung und die Lieferantenauswahl gehören zu den wichtigsten Aufgaben im Lieferantenmanagement [31]. Die Ziele der Evaluierung sind dabei Qualitätssicherung, Identifikation von zuverlässigen Lieferanten und Identifikation von Lieferanten mit niedrigen Preisen [32]. Darüber hinaus spielt die Bereitstellung einer Rangfolge von geeigneten Lieferanten, die dem fokalen Lieferanten erlaubt, eine entsprechende Auswahl dieser zu treffen [29], eine große Rolle.

Damit der sehr wichtigen Aufgabe der Lieferantenauswahl Rechnung getragen werden kann, braucht die Beschaffungsfunktion im Unternehmen geeignete Verfahren zur Entscheidungsunterstützung, die es ermöglichen, Kriterien flexibel und unternehmensspezifisch zu kombinieren [29]. Die Lieferantenauswahl trägt sowohl strategischen, als auch operativen Charakter. Im Mittelpunkt der strategischen Auswahl stehen Erfolgspotenziale der Lieferanten, was sich vor allem in der Wahl der Kriterien nieder schlägt [33], während im Rahmen der operativen Auswahl konkrete Aufträge vergeben werden [32].

3.1 Methoden zur Risikobewertung von Lieferanten

Zur Risikobewertung von Lieferanten haben sich im Sachgütermarkt mehrere Methoden etabliert. Das Verfahren der Punktebewertung als numerisches Darstellungsverfahren, welches eine Ausprägung der qualitativen Verfahren ist [34] dominiert in der Literatur [35]. Auch für den Fall der Risikobewertung in Liefernetzwerken mit hybriden Leistungsbündeln ist es am geeignetsten [8]. Punktebewertungsverfahren werden in der wissenschaftlichen Literatur noch einmal unterteilt in 100-Punkte-Bewertungsverfahren, Prozentbewertungsverfahren und Scoring-Modelle. Für das zu

entwickelnde Modell wird das Scoring-Modell verwendet, um eine optimale Gewichtung der Einflussfaktoren zu erreichen [32].

3.2 Beispielhafte Berechnung anhand von Entscheidungskriterien

Um das gesamte Liefernetz zu bewerten, ist es zunächst erforderlich, die einzelnen Knoten zu bewerten [29]. Ein Knoten entspricht in unserem Modell einem Lieferanten, Vorlieferanten oder dem fokalen Lieferanten selbst.

Aufstellen des Kriterienkatalogs. Der fokale Lieferant erstellt hierfür zunächst einen Entscheidungskatalog anhand von Erfahrungswerten und den für ihn entscheidungsrelevanten Kriterien. Diese Liste an Entscheidungskriterien muss nun mit einem Punktbewertungsverfahren beurteilt werden. Dazu wird jedem Kriterium ein Punktwert zugewiesen, welcher aussagt, wie viel dem fokalen Lieferanten die Kenntnis dieses Kriteriums von einem Lieferanten oder Vorlieferanten wert wäre. Ein Beispiel hierfür sind die ersten beiden Spalten in Tabelle 2. In diesem Fall wäre dem fokalen Lieferanten die Kenntnis über die Liefertreue des zu bewertenden Lieferanten sehr viel wert, das Wissen über den Ort des Firmensitzes allerdings verhältnismäßig wenig.

Tabelle 2. Bewertung von Entscheidungskriterien inkl. codierter und uncodierter Lieferantendaten

Kriterium	Gewichtungspunkte	Lieferantendaten (uc)	Lieferantendaten (c)
Liefertreue	20	0,9	0,9
Lieferqualität	17	-	-1
Leistungsgarantien	16	-	-1
Service-Kontrakte	14	3 Stück	0,85
Qualität der Arbeiter	13	-	-1
Liquidität	13	0,8	0,8
Anzahl Mitarbeiter	8	10000	0,7
Finanzierung	6	Leasing	0,3
Firmensitz	3	Augsburg	0,9

Bewertung durch die Lieferanten. Nun muss diese Liste mit den entsprechenden Informationen von den zu bewertenden Lieferanten, Vorlieferanten und Rohstoffhersteller gefüllt werden. Ein mögliches Ergebnis ist in der dritten Spalte der Tabelle 2 dargestellt.

Codierung der Lieferantendaten. Nicht alle Lieferantendaten entsprechen dabei einem Zahlenwert, was für die weiteren Berechnungen zwingend nötig ist. Deswegen müssen die Werte noch codiert und als Zahl zwischen 0 und 1 dargestellt werden, wobei 0 als der schlechteste Wert angesehen wird und 1 als der optimale. Das bedeu-

tet, dass nicht-quantifizierbare Informationen wie Firmensitz oder Anzahl Mitarbeiter so weit klassifiziert werden müssen, dass sie die Bedeutung innerhalb des Kriterienkatalogs widerspiegeln. Den Wert 0,9 des Firmensitzes kann beispielsweise so interpretiert werden, dass der fokale Lieferant selbst auch in Augsburg seinen Firmensitz hat und somit die Transportstrecke minimal ist, wodurch auch das Risiko, dass während des Transports ein Ausfall eintritt, minimiert wird. Der Wert „-1“ wird als Co-dierung für nicht angegebene Daten genommen, um diese Information in die Berechnung einfließen lassen zu können.

Informationsunvollständigkeit. Solch eine Tabelle kann in der Praxis mehrere hundert Entscheidungskriterien beinhalten. Lieferanten können und wollen in der Regel nicht zu allen Kriterien explizit Stellung nehmen. Allerdings möchte das fokale Unternehmen weitestgehende Informationen zur optimalen Auswahl. Daher ist eine bewusste Rahmensetzung von Kriterien notwendig, die vom Lieferanten anzugebenden sind. Eine Möglichkeit hierfür wäre eine Mindestanzahl der Summe der Gewichtungspunkte festzusetzen, für deren Kriterien Daten von den zu bewertenden Lieferanten erhalten werden müssen. Der fokale Lieferant würde hierfür bestimmen, dass von jedem, in Frage kommenden Lieferanten Angaben zu machen sind, die in Summe mindestens beispielsweise 60 Gewichtungspunkte haben. Dies hätte den Vorteil, dass der Lieferant entweder sehr viele Angaben zu Kriterien machen kann, die für den fokalen Lieferanten zwar einzeln keinen großen Wert haben, aber in Summe eine gute Einschätzung des Lieferanten erlauben, oder der Lieferant macht nur wenige Angaben, dafür aber zu Kriterien, die für den fokalen Lieferanten höchst entscheidungsrelevant sind. Würde nun der fokale Lieferant beispielsweise festlegen, dass er Daten in Höhe von 60 Gewichtungspunkten benötigt, so wäre dies im oben genannten Beispiel erfüllt. Hier bietet der Lieferant Daten in Höhe von 20 Punkten + 14 Punkte + 13 Punkte + 8 Punkte + 6 Punkte + 3 Punkte = 64 Punkte an.

4 Modellierung der Risikobewertung im Liefernetzwerk hybrider Wertschöpfung

Für die Modellierung der Risikobewertung wird folgende Situation zugrunde gelegt: Der fokale Lieferant hat ein hybrides Leistungsbündel an einen einzelnen Kunden angeboten. Das hybride Leistungsbündel stellt dabei eine integrierte Lösung tangibler und intangibler Produkte und Lösungen dar, die an die spezifischen Anforderungen des Kunden angepasst sind. Der Kunde hat dem Angebot zugestimmt. Nun stellt der fokale Lieferant nicht alle Elemente des hybriden Leistungsbündels selbst her. Er muss daher, bezogen auf die Bestellung des Kunden, die fehlenden Komponenten in seinem Liefernetzwerk beziehen. Für diesen Beschaffungsvorgang möchte der fokale Lieferant das Risiko minimieren, einen ungeeigneten Lieferanten auszuwählen, dessen Angebot im Rahmen des Integrationsprozess der Teilkomponenten zur Kundenlösung eventuell zu Problemen führen könnte.

Die Grundlage der Risikomodellierung stellt hierbei das Punktbewertungsverfahren dar. Dieses wird erweitert durch die Gestaltungselemente, die in Kapitel 3 disku-

tiert wurden. Zunächst stellt der fokale Lieferant einen für das gesamte Liefernetzwerk gültigen Kriterienkatalog auf. Dieser Kriterienkatalog umfasst die spezifischen Risikoelemente, wobei darüber hinaus jedes Entscheidungskriterium mit einem Gewichtungsfaktor verbunden ist, der die Relevanz des Kriteriums innerhalb der Menge der Entscheidungskriterien widerspiegelt. Im zweiten Schritt wird dieser Kriterienkatalog von den Lieferanten befüllt im Rahmen der Beschaffungsanfrage seitens des fokalen Lieferantens. Diese Lieferantenantworten werden anhand vorgegebener Umrechnungstabellen konvertiert, um auch nicht-numerische Antworten in die Risikobewertung einfließen lassen zu können.

Die durch die Konvertierung und Gewichtung entstandenen Werte der einzelnen Knoten müssen nun zu einem Risikowert komprimiert werden, welcher eine konkrete Lieferkette umfasst. Hierbei muss insbesondere beachtet werden, dass speziell für dienstleistungsorientierte Komponenten des hybriden Leistungsbündels die Antworten der jeweiligen Lieferanten nicht immer vollständig sein müssen. Das zu entwickelte Modell muss daher in der Lage sein, diese Unvollständigkeit in der Antwort so in die Risikobewertung zu integrieren, dass das Verfahren nicht zum Abbruch kommt. Im Folgenden soll nun der Wert von einer speziellen Lieferkette für den fokalen Lieferanten berechnet werden.

4.1 Entscheidungskriterium

Dieser Wert der i -ten Lieferkette l_i soll im Folgenden Ω_{l_i} heißen. Der Wert liegt im Intervall $[0;1]$, wobei er umso besser ist, je mehr er sich 1 annähert. Die Eingrenzung auf das Intervall $[0;1]$ ist dabei lediglich als Normierung zu verstehen. Grundsätzlich könnte jedes andere Intervall verwendet werden. Ein besserer Wert für eine Lieferkette heißt in diesem Fall, dass diese Lieferkette weniger Risiko aufweist als Lieferketten, deren Wert für Ω_{l_i} niedriger ist. Ω_{l_i} ist sowohl abhängig von den Gewichtungspunktzahlen der betroffenen Kriterien als auch den expliziten Ausprägungen der in der zu betrachtenden Lieferkette betroffenen Kriterien.

4.2 Modellformulierung

Als Grundlage für die Modellformulierung und damit für die Berechnung der Risikobewertung Ω_{l_i} der i -ten Lieferkette werden folgende Variablen und Indizes verwendet:

Variablen

$\Omega_{l_i} \in [0;1]$: Wert der Lieferkette l_i für den fokalen Lieferanten

$l_i = \{m_{klmk} \in K; K \text{ Konfiguration}\}$: i -te Lieferkette bestehend aus den Knoten, die ein vom Kunden gewähltes Modul, liefern können

$M = \{m_1; \dots; m_z\}$: Menge aller Lieferanten im gesamten Netzwerk

M_{k, l_i} : Lieferant der Kette l_i mit der ID k

$C = \{c_1; \dots; c_y\}$: Menge aller im Netzwerk vorkommenden Kriterien

$G = \{g_1; \dots; g_y\}$: Menge aller Gewichtungspunkte des fokalen Lieferanten

$g_j \in \{1; 2; \dots; o\}$: Gewichtungspunktzahl eines Kriteriums c_j

a_{j,l_i} : Häufigkeit des Kriteriums c_j innerhalb der Lieferkette l_i
 $w_{k,j} \in [0;1]$: konkreter Wert eines Kriteriums c_j für den Lieferanten m_k

Indizes

y : Anzahl der im Netzwerk vorkommenden Kriterien
 o : höchstmögliche Gewichtungspunktzahl in der Kriterienliste
 z : Anzahl der Lieferanten im gesamten Netzwerk
 k : ID eines Lieferanten im gesamten Netzwerk
 i : ID einer der möglichen Lieferketten für ein vom Kunden gewähltes Modul
 j : ID eines im Netzwerk vorkommenden Kriteriums

Daraus ergibt sich für die Berechnung des Wertes einer Lieferkette für den fokalen Lieferanten folgende Formel:

$$\Omega_{l_i} = \frac{\sum_{k=1}^{m_z} \sum_{j=1}^{c_y} f(w_{k,j}) * g_j}{\sum_{j=1}^{c_y} (g_j * f(a_{j,l_i}))} \quad (1)$$

Der Zähler von Ω_{l_i} stellt die gewichtete Lieferantenantwort dar und setzt sich somit zusammen aus dem Produkt von den jeweils konkreten Werten der Kriterien $w_{k,j}$ und den entsprechenden Gewichtungspunktzahlen g_j über alle an der jeweiligen Lieferkette beteiligten Lieferanten M und alle von diesen veröffentlichten Kriterien C . Der Nenner von Ω_{l_i} beschreibt den Umfang der Antworten, das heißt, die Summe der Gewichte der Kriterien, zu denen der Lieferant eine Angabe gemacht hat.

Die Teilfunktion $f(w_{k,j})$ extrahiert aus der Menge der Antworten eines Lieferanten die Antworten, die in die Berechnung mit einfließen sollen. Antwortmöglichkeiten, die vom Lieferanten nicht gefüllt wurden, werden über diese Funktion eliminiert. Der Kriterienkatalog ist dabei so gestaltet, dass die Variable $w_{k,j}$ den Wert -1 annimmt, falls diese Daten in der Tabelle nicht angegeben wurden. Die Formel zur Berechnung der Teilfunktion $f(w_{k,j})$ lautet dabei:

$$f(w_{k,j}) = \begin{cases} w_{k,j} & \text{für } w_{k,j} \neq -1 \\ 0 & \text{sonst} \end{cases} \quad (2)$$

Die Teilfunktion $f(a_{j,l_i})$ extrahiert aus der Menge der Kriterien, die seitens der Lieferanten beantwortbar wären, diejenigen, für die auch eine Antwort erfolgt ist. Wurde ein Kriterium durch den Lieferanten beantwortet, wird das Gewicht des Kriteriums durch die Funktion ermittelt. Die Formel zur Berechnung der Teilfunktion $f(a_{j,l_i})$ lautet dabei:

$$f(a_{j,l_i}) = \begin{cases} a_{j,l_i} & \text{für } w_{k,j} \neq -1 \\ -1 & \text{sonst} \end{cases} \quad (3)$$

Zusammengefasst wird die Risikobewertung Ω_{l_i} und damit der Wert einer bestimmten Lieferkette l_i berechnet als das Verhältnis der gewichteten Lieferantenantworten zu dem Umfang der abgegebenen Antworten des spezifischen Lieferanten.

4.3 Beispielhafte Anwendung des Modells

Das entwickelte Modell soll nun an einem exemplarischen Beispiel verdeutlicht werden, das in der Komplexität reduziert ist, aber alle relevanten Aspekte der Problematik des Risikomanagements für die Beschaffung hybrider Leistungsbündel umfasst. Angenommen, der fokale Lieferant besäße nur einen einzigen Lieferanten, der für das in diesem Beispiel betrachtete hybride Leistungsbündel relevant ist. Die Kriterien für dieses Beispiel sollen dabei die Ausprägungen gemäß Tabelle 2 haben. Die Gewichtungspunkte entsprechen den Einträgen in der zweiten Spalte g_j und die Daten des Lieferanten entsprechen den Einträgen in der dritten Spalte $w_{k,j}$. Somit ergibt sich als Ergebnis für die Risikokalkulation nach Formel (1) der Wert $\frac{50,4}{107} = 0,471$.

5 Software-Prototyp zur Simulation des Risikomodells

Das vorgestellte Risikomodell wurde in Form eines Software-Prototypen realisiert. Dieser Prototyp stellt eine Implementierung des Forschungsartefakts dar und ist daher geeignet, als Basis für eine Evaluation des Artefakt in Form eines strukturellen Tests herangezogen zu werden [36]. Der Prototyp wurde in modularer 3-Schicht-Architektur implementiert, um einen hohen Grad an Flexibilität bei der Abbildung unterschiedlicher Risikoszenarien zu ermöglichen.

5.1 Datenmodell

Um eine für dieses Modell passende Darstellung der hybriden Leistungsbündel und eine effiziente Implementierung zu generieren, wurde das semantische Datenmodell von Schrödl et. al [37] entsprechend angepasst (siehe Abbildung 1).

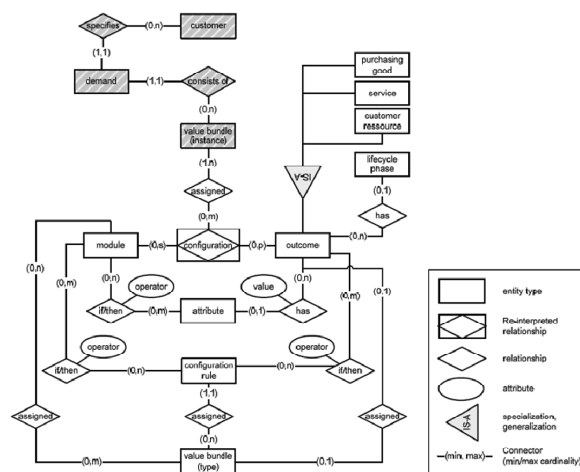


Abb. 1. Semantisches Datenmodell für hybride Leistungsbündel [37]

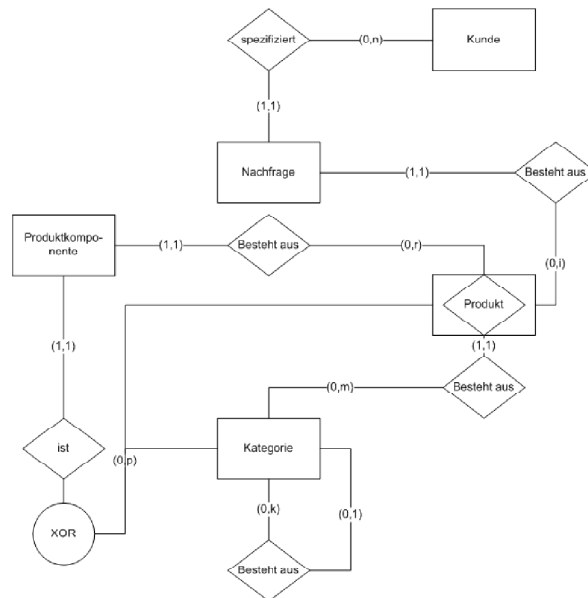


Abb. 2. Einordnung von Kategorien und Produktkomponenten in Anlehnung an Schrödl et. al [37]

An erster Stelle steht der Kunde mit einer Nachfrage nach einem Produkt, d. h. einer Instanz eines Leistungsbündeltypen. Im gesamten Liefernetzwerk fragen Kunden Instanzen bei ihren Lieferanten an, welche wiederum als Kunde Instanzen bei ihren Lieferanten erfragen. Der Typ des Leistungsbündels besteht nun aus einer Reihe von Modulen, welche „Outcomes“ zugewiesen bekommen. In der Sicht auf das gesamte Liefernetzwerk sind aber „Outcomes“ ebenfalls (hybride) Leistungsbündel von Lieferanten. Zur Realisierung von Modulen wurden Kategorien und Produktkomponente eingeführt, die wie folgt in das Modell eingegliedert werden (siehe auch Abbildung 2): Ein Produkt besitzt keine, oder eine Reihe von Komponenten. Eine Komponente besteht entweder aus einem Produkt oder aus einer Kategorie. Eine Kategorie besteht aus weiteren Kategorien oder Produkten. Eine Kategorie kann maximal eine übergeordnete Kategorie besitzen, ein Produkt ist immer genau einer Kategorie zugeordnet.

5.2 Funktionsweise des Prototypen

Das entwickelte Java-Programm bietet sowohl die Funktionalität zur Bewertung und Auswahl von Lieferanten als auch grafische Elemente zur Verwaltung der Daten, auf denen die Berechnung basiert. Darüber hinaus sind Komponenten zum Verwalten von Kriterien, Knoten, Produkten und Kategorien realisiert. Zuletzt kann der Benutzer noch durch die Visualizer-Komponente Berechnungen auf der von ihm einzugebenden Datenbasis durchführen. Der Kriterien-Editor erlaubt das Erstellen und Löschen von Einträgen. Der Knoten-Editor erstellt Knoten und weist deren Zuordnungen untereinander, die Produktion von Produkten und deren Kriterienausprägungen zu. Der

Graph, der das Lieferantennetzwerk darstellt, wird ausschließlich durch Erstellen von eingehenden Kanten definiert. Ausgehende Kanten für betroffene Knoten werden automatisch hinzugefügt. Kreise und unerwünschte Konstruktionen im Lieferantennetzwerk werden vom System nicht als Erstellungsoption angeboten.

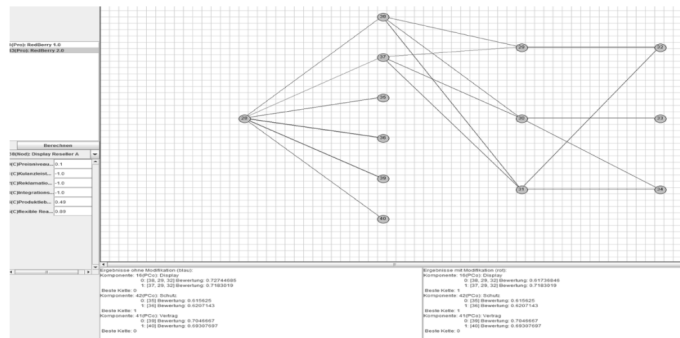


Fig. 3. Visualizer

Die Berechnung der Risiken der einzelnen Lieferketten wird durch den Visualizer sichtbar gemacht. Ausgehend vom ersten Knoten ohne ausgehende Kanten, im Normalfall der fokale Lieferant, wird das Netzwerk der erreichbaren Knoten erstellt. Nicht erreichbare Knoten werden vom Programm vernachlässigt. Links im Fenster findet sich eine Auswahl der Produkte, die der fokale Lieferant produziert. Anpassungen der Parameter führen im Regelfall zu einer anderen Berechnung und werden im Graphen in einer anderen Farbe präsentiert.

5.3 Praktische Relevanz

Der vorgestellte Prototyp wurde mit mehreren Experten aus dem Supply Chain Management diskutiert. Hierbei waren Experten aus Unternehmen vertreten, die hybride Leistungsbündel anbieten und entsprechende Komponenten der Leistungsbündel beschaffen. Eine Analyse ergab, dass im praktischen Umfeld eine strukturierte Betrachtung des Beschaffungsrisikos in der hybriden Wertschöpfung nur ansatzweise stattfinden. Im Wesentlichen werden dabei die klassischen Risiken der Beschaffung von Sachleistungen herangezogen. Darüber hinaus hat die Präsentation des Prototypen mit spezifisch angepassten Leistungsbündeln ergeben, dass solch ein Instrument einen Mehrwert in aktuellen und strategischen Beschaffungsvorgängen darstellen würde. So bietet die systematische Bestimmung des Risikos die Möglichkeit, mehr kleine und hochspezialisierte Lieferanten in die Lieferkette zu integrieren, um die Angebotsdifferenzierung weiter zu optimieren, ohne den Servicegrad gegenüber dem Kunden zu gefährden. Weiterhin kann das Modell genutzt werden, um die Lieferantenauswahl mit einem höheren Automatisierungsgrad zu implementieren, was einen positiven Effekt auf Prozesskosten und Laufzeiten hätte.

6 Zusammenfassung und Ausblick

Das Ziel dieses Beitrags ist die Entwicklung eines Modells zur Risikobewertung von Lieferanten für hybride Leistungsbündel. Hierzu wurden anhand einer Literaturstudie Merkmale hybrider Leistungsbündel identifiziert sowie unterschiedliche Verfahren zur Risikobewertungen in Lieferantenbeziehungen ausgewertet. Basierend auf diesen Erkenntnissen wurde ein neues Modell für eine Risikobewertung von Lieferanten für hybride Leistungsbündel entwickelt. Dieses Modell beruht auf einem Punktbewertungsverfahren mit angepasstem Entscheidungskatalog in Ergänzung mit einem gleichenden Mittelwertverfahren. Insbesondere bietet dieses Modell die Möglichkeit, mit unvollständigen Informationen der Lieferanten dennoch eine Risikooptimierung durchführen zu können. Damit unterscheidet es sich signifikant von etablierten Modellen des Supply Chain Management für klassische Produkte. Zur Simulation dieses Modells wurde ein Prototyp entwickelt, der die unterschiedlichen Einsatzmöglichkeiten des Risikomodells demonstriert. Anhand dieser Simulationsmöglichkeit können unterschiedliche Risikoszenarien untersucht werden und die für den jeweiligen Beschaffungsbedarf optimalste Lieferkette ermittelt werden. Es zeigt sich, dass das Modell nachvollziehbare und gut interpretierbare Ergebnisse liefert, die Anbietern von hybriden Leistungsbündeln ermöglichen, ihre Angebot mit einem minimierten Risiko am Markt anzubieten. Dies ermöglicht den Unternehmen eine weitere Fokussierung auf ein innovatives Angebotsportfolio und damit verbunden einer Ausweitung von interessanten Marktpositionen. Das vorgestellte Modell ist variabel in den Kriterien und kann daher auch genutzt werden, um eine optimale Lieferantenstrategie für bestimmte hybride Angebote zu identifizieren. Darüber hinaus kann das Modell genutzt werden, um in der operativen Beschaffung als Entscheidungsgrundlage zu fungieren, falls eines der Risiken in der Beschaffung eintritt und sich die Frage nach einer optimalen Ausweichvariante stellt. Zusammenfassend lässt sich festhalten, dass das vorgestellte Modell eine Optimierung der Lieferantenbeziehungen für die Anbieter hybrider Leistungsbündel darstellt. Das vorgestellte Modell kann als erster Schritt in Richtung eines umfassenden Risikomanagements in Liefernetzwerken für hybride Leistungsbündel gesehen werden. Als weitere Schritte sind mehrere Aspekte denkbar. Hierzu zählt unter anderem die Aufnahme des Faktors Zeit und damit eine Erweiterung in Richtung operativer Beschaffung. Dieser ist besonders relevant bei Änderungen von Lieferungen, die zum Beispiel durch Lieferverzug, Teillieferung oder Lieferausfälle entstehen können. Ein weiterer Faktor ist die Tatsache, dass in der Praxis eventuell nicht alle Lieferanten die vom fokalen Lieferanten geforderten Informationen zur Verfügung stellen werden. In dem in dieser Arbeit aufgestellten Modell würde dies zwar zu einem durchaus akzeptablen Ergebnis führen, bei großer Varianz könnten die Werte aber trotzdem die Interpretation aus Sicht des fokalen Lieferanten erschweren. Es bleibt also auch über diese Arbeit hinaus die Frage nach dem Umgang mit unvollständigen Informationen in einem solchen Modell. Des Weiteren wäre noch ein Maß für die Verzahnung von Dienstleistung und Sachgütern in Form von hybriden Leistungsbündeln als Kriterium für Risiko, und somit als Einflussfaktor auf die Bewertung von Lieferketten, zu evaluieren. Als letzter Punkt ist die Tatsache zu erwähnen, dass der fokale Lieferant bezüglich der Kriterienliste andere Bewertungen

beziehungsweise Schwerpunkte haben kann als ein Vorlieferant. So kann beispielsweise die Lieferwahrscheinlichkeit verschiedene Bedeutungen aufweisen, was zu unterschiedlichen Bewertungen von Lieferketten führen kann. Auch eine unterschiedliche Gewichtung der Lieferanteninformationen je nach deren Position in der Lieferkette wäre denkbar.

Literatur

1. Leimeister, J.M., Glauner, C.: Hybride Produkte - Einordnung und Herausforderung für die Wirtschaftsinformatik. *Wirtschaftsinformatik* 50 (3), 248–251 (2008)
2. Böhm, T., Krcmar, H.: Hybride Produkte: Merkmale und Herausforderungen. In: *Wertschöpfungsprozesse bei Dienstleistungen*. Gabler (2007)
3. Wise, R., Baumgartner, P.: Go Downstream: The New Profit Imperative in Manufacturing. *Harvard business review* 77, 133-141 (1999)
4. Howells, J.: Industrial Consumption and Innovation. In: *The workshop: "Industrial ecology and spaces of innovation"*, pp. 203-219. Edward Elgar Publishing (2003)
5. Becker, J., Beverungen, D., Knackstedt, R.: Wertschöpfungsnetzwerke von Produzenten und Dienstleistern als Option zur Organisation der Erstellung hybrider Leistungsbündel. In: *Wertschöpfungsnetzwerke*. Physica, Heidelberg (2008)
6. Galbraith, J.R.: Organizing to Deliver Solutions. *Org. Dynamics* 31, 194-207 (2002)
7. Kersten, W., Zink, T., Kern, E.-M.: Wertschöpfungsnetzwerke zur Entwicklung und Produktion hybrider Produkte. In: *Wertschöpfungsnetzwerke*. Schmidt (2006)
8. Schrödl, H., Geier, L.: Risikomanagement in der hybriden Wertschöpfung: ein Vergleichsrahmen zur Bewertung von Risikomodellen für die Lieferantenauswahl. In: Thomas, O., Nüttgens, M. (eds.): *Dienstleistungsmodellierung 2012*. Springer Gabler (2012)
9. vom Brocke, J.: *Serviceorientierte Architekturen - SOA*. Management und Controlling von Geschäftsprozessen. Vahlen, München (2008)
10. Johansson, J.E., Krishnamurthy, C., Schliessberg, H.E.: Solving the solutions problem. *McKinsey Quarterly* 3, 116–125 (2003)
11. Foote, N.W., Galbraith, J., Hope, Q., Miller, D.: Making solutions the answer. *McKinsey Quarterly* 3, 84–93 (2001)
12. Scholl, G.: Product Service Systems. In: *Perspectives on Radical Changes to Sustainable Consumption and Production (SCP)*, pp. 25-43 (2006)
13. Sawhney, M., Wolcott, R.C., Arroniz, I.: The 12 Different Ways for Companies to Innovate. *MIT Sloan Management Review* 47 (3), 74-82 (2006)
14. Vahrenkamp, R., Amann, M. (eds.): *Risikomanagement in Supply Chains*. Gefahren abwehren, Chancen nutzen, Erfolg generieren. Schmidt, Berlin (2007)
15. Aissaoui, N., Haouari, M., Hassini, E.: Supplier selection and order lot sizing modeling: A review. *Computers & Operations Research* 34 (12), 3516–3540 (2007)
16. Kajüter, P.: Instrumente zum Risikomanagement in der Supply Chain. In: Stölzle, W. (ed.): *Supply Chain Controlling in Theorie und Praxis*. Gabler, Wiesbaden (2003)
17. Voigt, K.-I., Thiell, M.: Beschaffung wissensintensiver Dienstleistungen. In: Bruhn, M., Stauss, B. (eds.): *Dienstleistungsnetzwerke*. Gabler, Wiesbaden (2003)
18. Thiell, M.: *Strategische Beschaffung von Dienstleistungen*. Univ.-Dissertation (2006)
19. Burr, W.: *Service-Engineering bei technischen Dienstleistungen*. DUV, Wiesbaden (2002)
20. Götze, U., Betz, S., Götze-Henselmann-Mikus: *Risikomanagement*. Physica, Heidelberg (2001)

21. Harland, C., Brenchley, R., Walker, H.: Risk in supply networks. *Journal of Purchasing & Supply Management* 9 (2), 51-62 (2003)
22. Wright, P.: Consumer Choice Strategies: Simplifying Vs. Optimizing. *Journal of Marketing Research* 12 (1), 60-67 (1975)
23. Holt, G.D.: Which contractor selection methodology?. *International Journal of Project Management* 16 (3), 153-164 (1998)
24. Liu, J., Ding, F.-Y., Lall, V.: Using DEA to compare suppliers for supplier selection and performance improvement. *SCM: An International Journal* 5, 143-150 (2000)
25. Talluri, S., Narasimhan, R.: Vendor evaluation with performance variability: A max-min approach. *European Journal of Operational Research* 146, 543-552 (2003)
26. Weber, C.A., Current, J.R., Desai, A.: Vendor: A Structured Approach to Vendor Selection and Negotiation. *Journal of Business Logistics* 21, 135-167 (2000)
27. Burianek, F., Ihl, C., Bonnemeier, S., Reichwald, R.: Typologisierung hybrider Produkte TUM Lehrstuhl für BWL - Information Organisation u. Management, München (2007)
28. Heyder, M., Fahrtmann, K., Theuvsen, L.: Lieferantenbewertung in der Lebensmittelindustrie: Eine empirische Analyse. In: *Jahrbuch der Österreichischen Ges. für Agrarökonomie*, Band 18 (1). Facultas, Wien (2009)
29. Müssigmann, N.: Evaluierung und Auswahl von strategischen Liefernetzen unter Berücksichtigung kritischer Knoten. Univ. Dissertation, Augsburg (2006)
30. Pousttchi, K., Schrödl, H., Turowski, K.: Characteristics of Value Bundles in RFID-enabled Supply Networks. In: *9th ICEB 2009*, pp. 886-893 (2009)
31. Muralidharan, C., Anantharaman, N., Deshmukh, S.G.: A Multi-Criteria Group Decision-making Model for Supplier Rating. *J. of Supply Chain Management* 38, 22-33 (2002)
32. Janker, C.G.: Multivariate Lieferantenbewertung. Gabler / GWV, Wiesbaden, (2008)
33. Sarkis, J., Talluri, S.: A Model for Strategic Supplier Selection. *Journal of Supply Chain Management* 38, 18-28 (2002)
34. Beucker, S.: Ein Verfahren zur Bewertung von Lieferanten auf der Grundlage von Umweltwirkungen unter Berücksichtigung von Prozesskosten. Univ. Dissertation, Stuttgart (2005)
35. Koppelmann, U.: *Beschaffungsmarketing*. Springer, Berlin Heidelberg (2004)
36. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. *MIS Quarterly* 28, 75-105 (2004)
37. Schrödl, H., Gugel, P., Turowski, K.: Modellierung strategischer Liefernetze für hybride Leistungsbündel. In: Thomas, O., Nüttgens, M. (eds.): *DLM 2010. CEUR*, Vol. 577, pp. 1-18. CEUR-WS.org (2010)

A Holistic Approach for Integrating Methods in Quality Management

Florian Johannsen

University of Regensburg, Department of Management Information Systems, Regensburg,
Germany
Florian.Johannsen@wiwi.uni-regensburg.de

Abstract. Quality management has become more and more important in the last couple of years. The user is faced with a multitude of methods such as Six Sigma, TQM or Theory of Constraints. Therefore quality managers are more and more engaged in the selection of an appropriate approach for achieving the quality goals as they have been defined. Due to the necessity of coordinating heterogeneous quality methods, employees usually oppose the use of more than one approach within an enterprise. However, guidelines on how to integrate several methods, while considering the strengths of the original approaches, are still missing. This problem is being dealt with in the paper at hand. The paper introduces an integration approach, supporting the user in establishing an integrated quality management method. The variety of quality management approaches within an enterprise can thus be influenced by the user.

Keywords: integration, quality management, quality technique

1 Introduction

According to Gartner Research “improving business processes” has been a business strategy of major priority in recent years and will be pursued by CIOs in the years to come (see [1]). In that context, process-oriented quality management has gained considerable attention [2], providing the process manager with a multitude of methods such as Six Sigma, Total Quality Management (TQM), EFQM or KAIZEN for example (see e.g. [3-5]). While many quality management methods have its origin in the production industry, they are increasingly used in service industries as well (see [4]). Quality management methods are usually characterized by specific strengths (see [4]). Lean Management (see [6]) for example enables the elimination of “non-value” adding activities. Regarding the multitude of existing methods, quality managers are often overstrained selecting an appropriate method to achieve a company’s quality goals. As a consequence, enterprises usually use more than just one method in parallel, e.g. Six Sigma and EFQM. On the one hand this seems reasonable, since synergies between quality management methods exist (see e.g. [7-9]). On the other hand the parallel use of more than just one quality management method needs proper coordination [10-11]. If the use of the methods (e.g. Six Sigma, EFQM, TQM, etc.) is not

coordinated, results achieved by one initiative may be redone by a subsequent project using another quality management method [11]. Nevertheless, most employees do not have the time to become acquainted with the functionality of several different quality management methods [10]. Integrating quality management methods is a means for handling the variety of methods and generating a “single process improvement-based approach” [12]. Nevertheless guidelines on how to systematically integrate methods in quality management in a value-adding way do not yet exist. The paper at hand deals with that problem and introduces a holistic approach for combining methods in quality management following a design science approach (see [13]). The structure of the paper is as follows: In the following section, basics and challenges of quality management are described. Integration scenarios for specifying the application range of the integration approach are shown in section 3. Afterwards (section 4) requirements on an approach for integrating methods in quality management are derived. In section 5, the approach for integration is introduced, and applied at an automotive bank in section 6. The paper concludes with limitations of the paper and an outlook on further research.

2 Basics and Challenges

2.1 Elements of a Quality Management Method

According to Klefsjö et al. [14], quality management is lacking a theoretical foundation. A precise definition of the term “quality management method” cannot be found in literature. In that context, de Mast [15] introduces elements (steps, rules, concepts, tools) for comparing alternative methods in quality management. Andersson et al. [4] mention objectives, theories or strategies to describe a method from a strategic perspective. A more precise definition of a method can be found in method engineering. Consequently a method consists of activities (determining a procedure model), result documents, techniques (tools), roles, and a meta model (see [16]). Transferring these elements to the context at hand, a quality management method is characterized by activities, result documents, quality techniques and roles.¹ The activities of a quality management method make up its procedure model (e.g. DMAIC-cycle). The activities are performed by roles (e.g. team member) for producing result documents (e.g. performance data). Quality techniques (e.g. fishbone diagram) support the user in establishing the result documents. These method elements enable a structured description of a quality management method from an operational view, pointing out its functionality.

¹ We leave out the meta model as a constitute element (see [16]) for the research at hand since the quality management methods do not produce conceptual models as a main result.

2.2 Related Work

In literature, several combinations of quality management methods are described. To bring clarity to the topic a literature review (see [17]) has been conducted in a previous work (see [18]) to analyze integration efforts in quality management. The main focus was to identify those combinations of methods that are often dealt with in literature (e.g. Six Sigma and Lean Management), the underlying motivation for the integration effort as well as the steps taken for performing the integration. It became obvious that usually the result of the integration is presented. A systematic procedure or even guidelines on how the methods have been integrated are not given. Table 1 presents an overview regarding existing works, dealing with the integration of certain quality management methods. The first column shows the general steps taken by the authors to integrate the methods on an abstract level. These have been derived from the implicit description within the corresponding papers since a systematic and profound process of integration is not described. In the second and third column, the methods being integrated that way as well as the sources are shown.

Table 1. Examples for the integration of quality management methods

Steps taken for integration	Integrated methods	Authors
Search for common core concepts in quality management methods to derive synergies	<ul style="list-style-type: none"> • Six Sigma and ISO 9000 • Six Sigma and 5-S • ISO 9000, EFQM, BSC and Six Sigma 	[7], [19], [20], [9], [21]
Analysis of weaknesses and strengths and derivation of synergies	<ul style="list-style-type: none"> • Six Sigma and ISO 9000 	[22]
Creation of a framework in which quality management methods fulfill specific tasks	<ul style="list-style-type: none"> • Six Sigma, BSC and EFQM • TQM and ISO 9000 • TQM and TPM (Total Productive Maintenance) 	[8], [23], [24]
The procedure model (e.g. PDCA-cycle) of a method is enhanced by activities of another method	<ul style="list-style-type: none"> • Six Sigma and Theory of Constraints (TOC) • Six Sigma and CQI (Continuous Quality Improvement) 	[25], [26]
A completely new procedure model is derived from the activities of the methods considered	<ul style="list-style-type: none"> • Six Sigma and Lean Management 	[27]
A specific method is enhanced by certain quality techniques from another method	<ul style="list-style-type: none"> • Six Sigma and Lean Management 	[9]

Analyzing the quality management literature (see Table 1), it becomes obvious that integrating methods is performed in an ad-hoc manner, a fact that also holds true for related areas of application (e.g. systems engineering) (see [39]). While several integration efforts can be found, quality management is missing a profound and established theory for integration in general. A reason for that are specific challenges that need to be considered, making integration in quality management a complex discipline (see section 2.3). As a result, well-established and holistic approaches that both guide users in integrating quality management methods and, at the same time cope with these challenges, do not exist yet.

2.3 Challenges of Integration in Quality Management

Integration in quality management is a demanding discipline. However, the challenges associated with the process of integration are not externalized in literature. Therefore the challenges have to be derived from the authors' description of integration efforts in quality management (see [18]). In the following, we concentrate on those challenges that are generally valid and are independent of the combination of specific methods (e.g. Six Sigma and ISO 9000). A major problem in quality management is that *different interpretations* can be found. Magnusson et al. [28] introduce different perspectives on Six Sigma for example, which comprise its interpretation as a holistic companywide strategy, an improvement method as well as a mere toolbox of quality techniques. In literature (see e.g. Table 1), attention has thus been given to an author's interpretation. While some authors reflect upon quality management methods from a strategic perspective (see e.g. [7], [20]), others take a more operational view instead (see e.g. [25-26]). Thus a classification scheme for interpreting quality management methods can be derived which is shown in Fig. 1. On each level different aspects of a method are focused (see [18]).

Level 1: On level 1, the authors focus on aspects such as organizational concepts, core values (e.g. process orientation), underlying philosophies, or quality goals for example (see e.g. [7], [20]). A strategic view on quality management is given.

Level 2: In a further category (level 2), authors interpret quality management methods as constructs compound of method elements (see section 2.1) that can be found in method engineering (see [16]). A quality management method is seen as a means for improving a business process, and not a philosophy or business strategy. Consequently, in most cases procedure models (e.g. DMAIC, PDCA, etc.) and result documents of the quality management methods (see e.g. [26-27]) are focused. Integration is usually reached by merging procedure models or deriving a new procedure model from the methods considered on that level.

Level 3: Some authors only consider quality techniques (e.g. fishbone diagram, quality function deployment, etc.). This perspective is similar to the "toolbox" perspective on Six Sigma as introduced by Magnusson et al. [28]. Thus a quality management method is considered as a collection of quality techniques (see [29]).

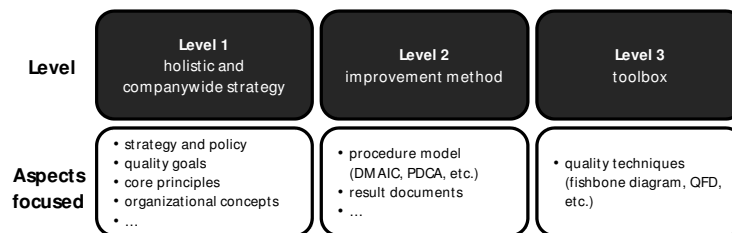


Fig. 1: Perspectives on quality management methods

A further challenge when integrating quality management methods are *naming conflicts* (see e.g. [30]). Due to the missing theoretical foundation of quality management (see [14]) the user is confronted with heterogeneous terms during integration which

hamper the identification of common concepts and differences between quality management methods. Synonyms may cause redundancies in the final method [31]. For example, the activities “definition of key performance indicators” (from quality management method A) and “definition of process metrics” (from quality management method B) may be transferred to the procedure model of the integrated method, because the user expects differences between these activities which do not exist. Also homonyms may cause problems. In quality management, the term “Six Sigma” is not only used to address the “improvement method” (see [28]) but also the sixfold standard deviation of a variable for example (see [3]).

Furthermore Bruhn [32] mentions “competing interdependencies” that exist between quality techniques. For example Lean Management techniques (e.g. value-stream-map) emphasize cost reduction, while Six Sigma and its techniques have a strong customer focus (e.g. VOC-/CTQ-matrix). According to Bruhn [32], this different focus may result in conflicts if corresponding techniques are used in combination.

3 Selection of an Integration Scenario

The sections above show that integration efforts in quality management can be diversified. For developing an approach for integration it is thus necessary to find an interpretation of a quality management method to work with. In the paper at hand, an operational perspective is taken (level 2 in Figure 1) and a quality management method is seen as a means for improving business processes. In section 2.1, the elements characterizing a quality management method have been introduced. Strategic aspects (e.g. organizational aspects) of a quality management method (see level 1 in Figure 1) are reflected in these method elements. Quality goals or core principles, such as reduction of waste, are realized by specific quality techniques (e.g. value-stream-map) for example. Organizational aspects on the other hand are managed by a corresponding role concept.

To structure integration efforts in quality management, integration scenarios can be established. Scenarios help in reducing complexity and are characterized by criteria [33]. The integration scenarios for quality management have been derived from a prior literature review (see [18]) analyzing case studies concerning integration efforts in quality management: At first, there is a certain motivation for integration efforts. In many cases, synergies between methods are searched for to eliminate weaknesses of certain quality management methods (e.g. missing activities for measuring process performance) (see e.g. [25]). Additionally, it becomes obvious that not only quality management methods (e.g. Six Sigma and Lean Management) are being integrated in literature, but also quality techniques (e.g. KANO-model and QFD) as well as quality management methods and quality techniques (e.g. FMEA and Six Sigma). Furthermore the integration can be performed regarding a specific improvement project or independent from a specific project constellation. If the integration is performed regarding an improvement project, project characteristics may require the availability of

certain activities or quality techniques in the resulting method.² In addition quality management methods can be “merged” or “joined” (form of integration) (see [34]). When merging methods one integrated quality management method results. Joining methods means that the original methods exist further on, but their parallel application is to be coordinated properly. This can be reached by exchanging result documents between the methods for example. The integration strategy determines whether two or more methods are integrated at a particular time (see [30]). Table 2 summarizes these characteristics.

Table 2. Morphological box to categorize integration scenarios

Characteristic	Value		
<i>Motivation for integration</i>	New methods are to be combined with existing methods	Synergies between methods / weaknesses are mitigated by integration	One method is the prerequisite for introducing another one
<i>Objects for integration</i>	Integration of quality management methods	Integration of quality techniques	Integration of quality management methods and quality techniques
	<i>e.g. Six Sigma and Work-Out</i>	<i>e.g. KANO-model and FMEA</i>	<i>e.g. FMEA and EFQM</i>
<i>Situativity of integration</i>	Project-related integration		Project-independent integration
	<i>Integration regarding a specific project constellation</i>		<i>Integration independent from a specific project constellation</i>
<i>Form of integration</i>	Merging		Joining
<i>Integration strategy</i>	Binary integration		n-ary integration

To manage the complexity of integration in quality management and to guarantee a decent level of detail a specific scenario is focused in the following. Thus this paper emphasizes the scenario most of the integration efforts found in literature (see [18]) can be ascribed to. Regarding the variety of integration efforts it is impossible to develop an integration approach that is suitable for all integration scenarios in an equal manner (see Table 2). In the scenario considered at hand, quality management methods (e.g. Lean Management and Six Sigma) are to be integrated, while synergies between methods are strived for. Quality techniques are part of quality management methods, so that solely focusing on techniques would restrict the range of application for the integration approach too much. The quality management methods are to be integrated independent from a specific project constellation. By that it can be guaranteed, that the integrated method can be used for different improvement projects, while it can be adapted for different contexts after the integration has been performed. The methods are to be merged, which means that the integration results in one integrated method. The problems in coordinating different quality management methods have been described in literature (see [10-11]). In addition employees appreciate using one method for quality management while no problem of selecting a method is given. The integration approach to be developed focuses the integration of two methods at a particular time which reduces complexity (see [30]). The coloring in Table 2 highlights the scenario considered in the following.

² No activities or quality techniques supporting the collection of data are required within the integrated method for example, if no data is to be collected in an improvement project.

4 Requirements on an Integration Approach

Following the design science paradigm (see [13]) requirements on the integration approach are defined in the following. In neighboring disciplines, such as method engineering, requirements on methods are defined which can be transferred to the context at hand. Greiffenberg [35] summarizes these requirements and categorizes them into “completeness”, “consistency”, and “intended purpose”. Since these requirements stem from a comprising literature review they are referred to and specified accordingly. In addition, there are requirements that arise from the challenges of integration in quality management as described in section 2.3. These are summarized in the category “challenges in quality management” (see Table 3). The requirements are formulated in such a way that their fulfillment can unambiguously be judged. A detailed description of all requirements in this paper is not possible. Instead the requirement “consistency in the procedure model” is exemplarily shown.

Consistency in the procedure model: Consistency in the procedure model focuses on the logical sequence of activities within the procedure model (see [35]). It has to be assured that the activities (see [16]) are performed in an order in which no contradictions arise regarding their sequence (see [35]). For the integration approach at hand, the “integration potential” has to be determined first before the quality management methods can actually be integrated for example. Table 3 provides an overview of the requirements on the integration approach, which have been derived from Greiffenberg [35] and adapted accordingly, as well as of the challenges in quality management (section 2.3).

Table 3. Requirements on the integration approach

Category	Requirement	Summary
<i>Completeness</i>	Input-/Output-completeness	All input required by specific activities is produced as output by other activities of the approach.
	Completeness regarding the method elements	Completeness of the approach is given, if a procedure model, result documents and corresponding techniques can be found.
	Completeness regarding the procedure model	Each result document can be unambiguously assigned to one or more activities.
<i>Consistency</i>	Consistency in the procedure model	<i>See explanation above.</i>
	Consistency of the result documents to be produced	The approach must assure that the result documents a user produces are definitely consistent with each other. For example a modeling notation used for describing methods A and B must be clearly defined to enable a comparison.
<i>Intended purpose</i>	Construction adequacy	The integration approach must enable a proper integration of quality management methods.
	Efficiency	No redundant activities are given and no result documents are produced redundantly.
	Ease of learning	A rapid understanding should be given for all employees of the company, while the approach should be easy to use as well.
	Flexibility	An adaption to the needs of the user, for example by skipping specific activities, should be possible.
<i>Challenges in quality management</i>	Support of consistent method perception	The approach must support and build on a consistent and unambiguous interpretation (method elements) of a quality management method.
	Assurance of consistency of the terms used	The approach must consider means to avoid naming conflicts in the resulting quality management method.
	Assurance of consistency of quality techniques	There must be means to avoid “competing interdependencies” (see [32]) between quality techniques in the integrated method.

5 A Holistic Approach for Integrating Methods

5.1 Approaches for Integration in Literature

In many disciplines such as data modeling (see [30], [36]), process modeling (see [37]), software development (see [38-39]) or even IT-Governance (see [40]), integration is a widely established concept. Batini et al. [30] introduce a four-step approach for schema integration in database management for example. Also the work of Hars [36] deals with data models and their integration, while a corresponding procedure is introduced. Rosemann [37] develops an approach for integrating business process models, while Kronl f and Ryan [39] present a general procedure for integrating methods in systems engineering. Van Hillegersberg and Kumar [38] integrate concepts for object-oriented systems development using meta models of the methods considered. Alter and Goeken [40] describe the integration of reference models in IT-Governance using a four-step approach and meta models. Table 4 shows the reflection of the these approaches against the design requirements as introduced in section 4.

Table 4. Reflection of the approaches against the design requirements

Requirements	Integration approach					
	[30]	[36]	[37]	[38]	[39]	[40]
Input-/Output-completeness	✓	✓	✓	✓	✓	✓
Completeness regarding the method elements	⊙	⊙	⊙	⊙	⊙	⊙
Completeness regarding the procedure model	✓	✓	✓	✓	✓	✓
Consistency in the procedure model	✓	✓	✓	✓	✓	✓
Consistency of the result documents to be produced	⊙	⊙	✓	✓	⊙	⊙
Construction adequacy	⊙	⊙	⊙	⊙	⊙	⊙
Efficiency	✓	✓	✓	✓	✓	✓
Ease of learning	⊙	⊙	⊙	⊙	⊙	⊙
Flexibility	⊙	⊙	⊙	⊙	⊙	⊙
Support of consistent method perception	⊙	⊙	⊙	⊙	⊙	⊙
Assurance of consistency of the terms used	✓	✓	✓	✓	⊙	✓
Assurance of consistency of quality techniques	⊙	⊙	⊙	⊙	⊙	⊙

(Legend: ✓: given and explicitly emphasized; ⊙: partly or implicitly given; ○: not given)

While the approaches are all based on a clear and consistent procedure model there are drawbacks in dealing with requirements that stem from the challenges of integration in quality management (see section 2.3). This especially concerns the support of a consistent method perception as well as a thorough consideration of quality techniques and their interdependencies (see e.g. [32]). An ill-conceived transfer to the domain of quality management is thus not possible, since the approaches have not been developed for that particular field of application. Therefore we introduce an approach for integration in quality management in the next section.

5.2 The Integration Approach for Quality Management

While the integration approaches introduced (see section 5.1) have problems in dealing with certain requirements stemming from quality management in particular, they

are all characterized by a consistent procedure model. Focusing the underlying procedures of the integration approaches it becomes obvious that (despite the different areas of application) there are many similarities. The following steps can thus be derived from the approaches (see [30], [36-41]):

1. Graphical visualization of the methods (see e.g. [40], [38])
2. Comparison and identification of conflicts (see [30], [36-37], [41])
3. Determination of an integration strategy (see e.g. [30], [36-37], [40])
4. Resolution of conflicts (see e.g. [30], [36-41])
5. Integration (see e.g. [30], [36-41])
6. Validation and Restructuring (see e.g. [30], [36-37], [40])
7. Selection of the methods (see e.g. [38])

These steps are generally valid and can therefore guide the integration of quality management methods as well. However, to fulfill the design requirements as defined in section 4, a proper specification of each step for the field of quality management is necessary (see section 5.3).

Figure 2 shows the integration approach on a generic level. It is based on the assumption that the quality management methods to be integrated have already been selected. Process managers are often expected to analyze a possible integration of newly developed methods with those already existing within an enterprise (see e.g. [42]). The introduction of new methods is usually expected by customers or dictated by management (see e.g. [42]). The methods to be integrated are already given in that case. In addition, literature presents a variety of value-creating combinations of quality management methods (see Table 1) which supports a practitioner searching for promising combinations. Thus our approach starts when the practitioner has already chosen the quality management methods to be integrated. The “graphical visualization of the quality management methods” (step 1) is helpful to illustrate the functionality of the methods to the employees. The comparison of the visualizations for methods A and B enables the identification of similarities and differences.

In the second step (“identification of integration potential and conflicts”) the quality management methods are compared. This way, integration potential is being recognized and the question as to which degree two methods complement each other is answered. Within a project for example two quality management methods A and B are given. By comparing both methods it becomes obvious that only some of the quality techniques from method B enhance method A in a value-adding way, for example. This insight is important for the later integration since in that case it does not make sense to derive a new procedure model from both methods. Much more the toolbox of quality techniques of method A should be extended by the corresponding quality techniques of method B. At the same time, conflicts (e.g. naming conflicts, competing interdependencies) have to be recognized to arrive at a consistent method during integration. The insights gained from step 2 are necessary for deriving an appropriate integration strategy (step 3 – “determination of an integration strategy”). The notion of “integration strategy” does not only address the differentiation between a binary and a n-ary integration but also a procedure for performing the integration. The quality management method A can be declared as a base method for example which is

selectively enhanced by activities, quality techniques, and roles from method B. This approach is advisable, if the method A produces the desired result documents to achieve the project goals (e.g. improving customer satisfaction). Nevertheless, enhancing method A (e.g. by the activity “collecting data”) leads to more precise results, which may be important for a project (e.g. “verified process performance”). An overview of integration strategies is given in a previous study (see [18]).

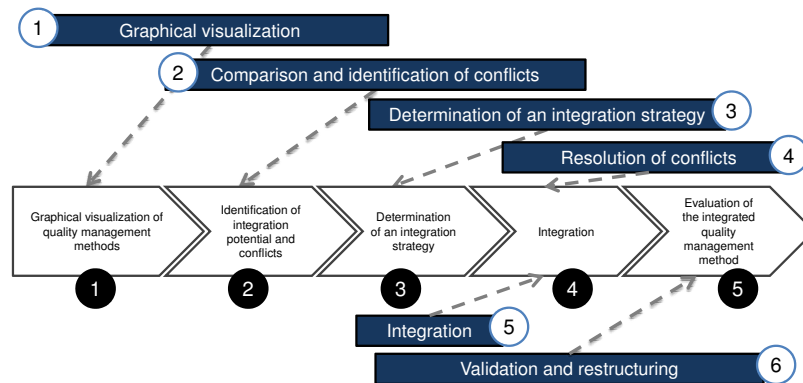


Fig. 2: Integration approach for quality management methods


After an appropriate integration strategy has been selected, the integration is performed. The integration varies depending on the integration strategy chosen. During integration also the conflicts are resolved. In a last step, the integrated method is evaluated, while the criteria as introduced by Greiffenberg [35] are modified and referred to. Each of the steps comprises several sub-steps. Step 4 (integration) is explained in more detail in the following.

5.3 Description of Step 4 – “Integration”

The integration itself (step 4) is the central step of the integration approach which merges two quality management methods A and B (e.g. Six Sigma and Work-Out). This step varies depending on the integration strategy chosen in the prior step 3. Nevertheless all integration strategies have a common pattern. At first the procedure models of the quality management methods are to be merged. The procedure model is the basis for embedding the quality techniques. The quality techniques are assigned to those activities supporting the creation of corresponding result documents (see [16]). Afterwards responsibilities respectively roles (see [16]) should be considered. Figure 3 shows the sub-steps that are to be performed for integrating methods A and B. In the example the following integration strategy has been chosen (see [18]): “*The quality management method A is declared as the base method which is enhanced by specific activities, quality techniques and roles of method B*”.

For integrating the procedure models five sub-steps have to be performed. At first (sub-step P1), the activities of the procedure model of method B should be specified regarding input-/output-relations (result documents). For example the activity “col-

lecting process data” has the input “data collection plan” and the output “process performance data”. Afterwards (sub-step P2) these activities are transferred to the procedure model of method A (P2: Identification of integration points). Consistency in the procedure model (see Table 3) has to be paid attention to at that point. No activities may use result documents as input which are produced in later activities. Naming conflicts (e.g. „performance metrics“ vs. „key measures“) have to be resolved (sub-step P3). Afterwards the new procedure model is constructed on a conceptual level (sub-step P4) and the user may consider (sub-step P5) how this new procedure model can be adapted for specific project situations in the sense of a “roadmap” (see [43]). That way the procedure model is already specified for certain project situations. Hereafter the quality techniques of both methods A and B are considered. Only those quality techniques of method B are transferred to the toolbox of quality techniques of method A that actually support the new procedure model (sub-step QT1). In addition, quality techniques may support each other, an aspect that Bruhn [32] calls complementary interdependencies. For example, the KANO-model enables a more precise prioritization of customer requirements in QFD (see e.g. [44]). If such interdependencies are given between quality techniques of method A and method B, further techniques from B may be transferred to the final toolbox of quality techniques (sub-step QT2). As a final step, the roles of the quality management methods have to be considered. From the role models of methods A and B a final role model for the integrated method has to be derived (sub-steps R1 and R2).



<i>Integration of the procedure models and the result documents</i>	<i>Integration of the quality techniques</i>	<i>Integration of roles</i>
P1: Specification of the activities to be transferred and their input/output-relations	QT1: Transfer of the quality techniques that support the creation of result documents in the new procedure model	R1: Merging of common roles
P2: Identification of integration points		
P3: Resolution of naming conflicts	QT2: Transfer of quality techniques that have complementary interdependencies	R2: Visualization of the new role model
P4: Visualization of the new procedure model		
P5: Enterprise-specific adaption		

Fig. 3: Integration of the quality management methods

As mentioned this procedure varies slightly for different integration strategies. The result of this procedure is the concept of the integrated quality management method.

6 Evaluation and Application of the Integration Approach

In the following the integration approach is reflected against the requirements defined in section 4. In addition, its application at an automotive bank is described. Table 5 shows those requirements focusing on the correct construction of the integration approach which can be confirmed at that point (see section 4). The practical applicability of the integration approach was tested at an automotive bank as a case study. At the automotive bank, Six Sigma (see [3]) had been introduced as the standard method

for quality management. Management intended to investigate a potential integration of Six Sigma and Work-Out (see [45]) to derive an integrated method. Work-Out was considered as a promising method for accelerating the implementation of improvement ideas within improvement efforts.

Table 5. Requirements regarding the correct construction of the integration approach

Requirements									
Input-/Output-completeness	Completeness regarding the method elements	Completeness regarding the procedure model	Consistency in the procedure model	Consistency of the result documents	Construction adequacy	Efficiency	Support of consistent method perception	Assurance of consistency of the terms used	Assurance of consistency of quality techniques
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

The task of developing an integrated method was assigned to the department “organization”. The project team consisted of three employees experienced in the Six Sigma concept. The Work-Out method had not been implemented at the automotive bank while the bank was unfamiliar with its functionality at that point. It was the intention to create one integrated method (from Six Sigma and Work-Out) since employees at the bank opposed the use of several methods in parallel. Thus a scenario was given the integration approach has been designed for.

The integration approach was performed stepwise by the project team leading to the integrated quality management method which was to be communicated to the employees afterwards. The graphical visualization of the quality management methods (step 1 – Figure 2) helped in creating a better understanding of the Work-Out method which was necessary to identify integration potential in the following. Although the creation of the corresponding models proved to be time-consuming the benefit of this step was appreciated for the identification of integration potential. Much effort was put into the identification of naming conflicts (step 2 – Figure 2) since both methods (Six Sigma and Work-Out) have similar concepts which however have different names (e.g. “stretch goals” vs. “project goals”). The comparison of both methods to find integration potential showed that Work-Out could provide activities, roles and quality techniques for enhancing the Improve-phase of the Six Sigma cycle (see [3]) by control-mechanisms for the implementation of improvement ideas. Because of that it was chosen (integration strategy) that Six Sigma should be the base method which is partially enhanced by quality techniques (e.g. Gallery of ideas [45]) and activities (e.g. “conduct Town-Meeting” [45]) of Work-Out (step 3 – Fig. 2). The integration (step 4 – Figure 2) was performed as shown in section 5.3. From that procedure an integrated method resulted which was mainly based on the company’s Six Sigma method and was extended by components of Work-Out to mitigate certain weaknesses in the Improve-phase of Six Sigma. Based on the use of the integration approach at the automotive bank its applicability as well as its ease of learning and flexibility (see section 4) could be judged. The integration approach proved suitable for integrating the methods Six Sigma and Work-Out in an adequate way, while the integration result was free from naming conflicts and competing interdependencies. Since the integration approach is characterized by a clear procedure

model and corresponding techniques, its ease of learning was confirmed after it had been applied for the first time. Nevertheless the identification of competing interdependencies was seen as a challenging task since profound knowledge regarding the quality techniques is necessary. Specific steps of the integration approach may be skipped emphasizing its flexibility. For example the quality management methods do not have to be visualized in case the user is familiar with their functionality.

7 Summary, Limitations and Outlook

Regarding the multitude of existing quality management methods, quality managers strive for ways to use the strengths of various approaches. Nevertheless the use of several methods in parallel is challenging, since a proper coordination is necessary.

Quality management is a discipline that lacks a theoretical foundation [14], hence no commonly accepted theory on integration exists. While some integration efforts are described in literature (see Table 1), integration is mostly performed in an ad-hoc manner in quality initiatives. Contrary to other disciplines such as data modeling (see [30], [36]), guidelines or well-known approaches for integration are missing in quality management. In section 5.1, it has been shown, that a transfer of established integration approaches from neighboring disciplines (e.g. [30], [38]) cannot be done due to specific challenges (see section 2.3). The paper at hand addresses this gap and introduces an integration approach supporting a quality manager to derive an integrated method for the purpose of quality management. The approach helps a company to combine existing methods (e.g. Lean Management) with newly arising quality management methods (e.g. Six Sigma). An ad-hoc introduction of quality management methods, which may cause problems in coordinating these methods, can be avoided. It became obvious that it was not possible to develop an integration approach equally suitable for all integration efforts. Because of that, a specific integration scenario is focused (see section 3). A limitation is that the integration approach is specially designed for that particular scenario. The integration approach is evaluated both against defined requirements and in a cooperation project. However a limitation is that the integration approach has so far only been applied in one cooperation project. Different interpretations of quality management methods require that a uniform way for describing the methods is found. In that context the method elements (see [16]) are used in the approach at hand. Strategic aspects (e.g. organizational concepts, etc.) are, however neglected; much more an operational interpretation of a quality management method is given. A limitation is that guidelines for implementing the integrated method in an enterprise are not part of the integration approach. Building ontologies for quality management in future research may support the user in comparing methods. These may be used for enhancing the integration approach as introduced.

References

1. McDonald, M.P., Aron, D.: Executive Summary: Reimagining IT: The 2011 CIO Agenda. Gartner Research (2011)
2. Stracke, C.: Process-oriented quality management. In: Ehlers, U.-D., Pawlowski, J.M. (eds.): Handbook on Quality and Standardisation in E-Learning. Springer, Berlin Heidelberg (2006)
3. Pande, P., Neumann, R., Cavanagh, R.: The Six Sigma Way - How GE, Motorola and other top companies are honing their performance. McGraw-Hill, New York et al. (2000)
4. Andersson, R., Eriksson, H., Torstensson, H.: Similarities and differences between TQM, six sigma and lean. The TQM Magazine 18, 282-296 (2006)
5. Imai, M.: Kaizen: The Key to Japan's Competitive Success. McGraw-Hill (1986)
6. Womack, J.P., Jones, D.T.: Lean Thinking. Simon & Schuster, New York (1996)
7. Pfeifer, T., Reissiger, W., Canales, C.: Integrating six sigma with quality management systems. The TQM Magazine 16, 241-249 (2004)
8. Töpfer, A.: Six Sigma, Balanced Score Card und EFQM-Modell im Wirkungsverbund. In: Töpfer, A. (ed.): Six Sigma. Springer, Berlin et al. (2007)
9. Snee, R., Hoerl, R.: Leading Six Sigma. Prentice Hall, New York et al. (2003)
10. Gijo, E.V., Rao, T., S.: Six Sigma Implementation – Hurdles and More Hurdles. Total Quality Management & Business Excellence 16, 721-725 (2005)
11. Samson, D., Challis, D.: Patterns of business excellence. Measuring Business Excellence 6, 15-21 (2002)
12. Bendell, T.: A review and comparison of six sigma and the lean organisations. The TQM Magazine 18, 255-262 (2006)
13. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75-105 (2004)
14. Klefsjö, B., Bergquist, B., Garvare, R.: Quality management and business excellence, customers and stakeholders: Do we agree on what we are talking about, and does it matter?. The TQM Journal 20, 120-129 (2008)
15. de Mast, J.: A methodological comparison of three strategies for quality improvement. International Journal of Quality & Reliability Management 21, 198-213 (2004)
16. Gutzwiller, T.A.: Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen. Physica, Heidelberg (1994)
17. Cooper, C.M., Hedges, L.V.: Research Synthesis As a Scientific Enterprise. In: Cooper, C.M., Hedges, L.V. (eds.): The Handbook of Research Synthesis. Russell Sage Foundation, New York (1994)
18. Johannsen, F.: State of the art concerning the integration of methods and techniques in quality management – literature review and an agenda for research. In: Proceedings ECIS 2011, Paper 41 (2011)
19. Bucher, P.: Integration von Six Sigma und ISO 9000 am Beispiel von GECTIS Austria. In: Töpfer, A. (ed.): Six Sigma. Springer, Berlin et al. (2007)
20. Low, S.P.: Towards TQM – integrating Japanese 5-S principles with ISO 9001:2000 requirements. The TQM Magazine 13, 334-341 (2001)
21. Ruiz, U., Simon, J.: Quality management in health care: a 20-year journey. International Journal of Health Care Quality Assurance 17, 323-333 (2004)
22. Töpfer, A., Günther, S.: Six Sigma im Wirkungsverbund mit ISO 9000:2000. In: Töpfer, A. (ed.): Six Sigma. Springer, Berlin et al. (2007)

23. Schmutte, A.M.: Six Sigma im Business Excellence Prozess - Wertorientierte Unternehmensführung mit Balanced Scorecard, EFQM und Six Sigma bei Siemens. In: Töpfer, A. (ed.): Six Sigma. Springer, Berlin et al. (2007)
24. Cua, K.O., McKone, K.E., Schroeder, R.G.: Relationships between implementation of TQM, JIT, and TPM and manufacturing performance. *Journal of Operations Management* 19, 675-694 (2001)
25. Revere, L., Black, K., Huq, A.: Integrating Six Sigma and CQI for improving patient care. *The TQM Magazine* 16, 105-113 (2004)
26. Ehie, I., Sheu, C.: Integrating six sigma and theory of constraints for continuous improvement: a case study. *Manufacturing Technology Management* 16, 542-553 (2005)
27. Bevan, H., Westwood, N., Crowe, R., O'Connor, M.: Lean Six Sigma: some basic concepts. *NHS* (2005)
28. Magnusson, K., Kroslid, D., Bergman, B.: Six Sigma umsetzen. Hanser, München (2004)
29. McQuater, R.E., Scurr, C.H., Dale, B.G., Hillmann, P.G.: Using quality tools and techniques successfully. *The TQM Magazine* 7, 37-42 (1995)
30. Batini, C., Lenzerini, M., Navathe, S.B.: A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys* 18, 323-364 (1986)
31. Strecker, S., Kargl, H.: Integration Deficits of IT Controlling. *Business & Information Systems Engineering* 1, 204-212 (2009)
32. Bruhn, M.: Qualitätsmanagement für Dienstleistungen. Springer, Berlin et al. (2011)
33. Weidenhaupt, K., Pohl, K., Jarke, M., Haumer, P.: Scenarios in System Development: Current Practice. *IEEE Software* 15 (2), 34-45 (1998)
34. Rosemann, M.: Gegenstand und Aufgaben des Integrationsmanagements. In: Scheer, A.-W., Rosemann, M., Schütte, R. (eds.): Integrationsmanagement – Arbeitsbericht des Instituts für Wirtschaftsinformatik Nr.65. University Münster (1999)
35. Greiffenberg, S.: Methodenentwicklung in Wirtschaft und Verwaltung. Dr. Kovac, Hamburg (2003)
36. Hars, A.: Referenzdatenmodelle – Grundlagen effizienter Datenmodellierung. Gabler, Wiesbaden (1994)
37. Rosemann, M.: Komplexitätsmanagement in Prozeßmodellen. Gabler, Wiesbaden (1996)
38. Van Hillegersberg, J., Kumar, K.: Using metamodeling to integrate object-oriented analysis, design and programming concepts. *Information Systems* 24, 113-129 (1999)
39. Kronlöf, K., Ryan, K.: Conclusions and Recommendations. In: Kronlöf, K. (ed.): Method Integration. John Wiley & Sons, Chichester et al. (1993)
40. Alter, S., Goeken, M.: Konzeptionelle Metamodelle von IT-Governance-Referenzmodellen als Basis der Kombination und Integration in einer Multi-Modell Umgebung. In: 9. Internationale Tagung Wirtschaftsinformatik (Band 1), pp. 705-714 (2009)
41. Spaccapietra, S., Parent, C.: View Integration: A Step Forward in Solving Structural Conflicts. *IEEE Transactions on Knowledge and Data Engineering* 6, 258-274 (1994)
42. Fritsch, V., Stössl, M.: Implementierung von Six Sigma in Abstimmung mit KVP bei VA TECH ELIN Transformatoren GmbH & Co, einem Unternehmen der Siemens Gruppe. In: Töpfer, A. (ed.): Six Sigma. Springer, Berlin et al. (2007)
43. Mirbel, I., Ralyté, J.: Situational method engineering: combining assembly-based and roadmap-driven approaches. *Requirements Engineering* 11, 58-78 (2006)
44. Tontini, G.: Integrating the Kano Model and QFD for Designing New Products. *Total Quality Management & Business Excellence* 18, 599-612 (2007)
45. Ashkenas, R.: Why Work-Out Works: Lessons from GE's Transformation Process. *Handbook of Business Strategy* 4, 15-21 (2003)

Leveraging Service Incident Analytics to Determine Cost-Optimal Service Offers

Axel Kieninger¹, Detlef Straeten², Steven Kimbrough³, Björn Schmitz¹, and Gerhard Satzger¹

¹ Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
{axel.kieninger,bjoern.schmitz,gerhard.satzger}@kit.edu

² IBM Deutschland GmbH, Walldorf, Germany
detlef.straeten@de.ibm.com

³ University of Pennsylvania, Philadelphia, USA
kimbrough@wharton.upenn.edu

Abstract. In this work we address the challenge for an IT service customer to select the cost-optimal service among different offers by external providers. We describe the customer's optimization problem by considering the negative monetary impact of potential service incidents on its business. First, we demonstrate that the information currently used in service level agreements may lead to suboptimal customer decisions. Second, we discuss how providers' private information about the behavior of service delivery environments could be leveraged by the customer when selecting service offers. Third, we propose a procurement auction as a mechanism to optimize total cost for the customer – choosing from different service offers by risk-neutral providers. In introducing this approach, we suggest that customers and providers collaborate to define service performance measures, which allow providers to better tailor service offers to customers' business requirements.

Keywords: Service Level Management, Service Level Agreement, Adverse Business Impact, Service Incident, Procurement Auction

1 Introduction

Outsourcing of IT services is popular and growing. The worldwide revenue generated with outsourcing totaled \$246.6 bn in 2011 – an increase of 7.8 percent compared to 2010 [1]. But in spite of the increasing significance of outsourced IT services, there are still opportunities for improvement (see, e.g., [2-3]).

In this work we address the business and IT alignment challenge (see, e.g., [4]) for an IT service customer to select the cost-optimal service among different offers by external providers. We describe the customer's optimization problem by considering the negative monetary impact of potential service incidents on its business.

Typically, the provision of outsourced services is governed by a *service level agreement* (SLA) – also including performance requirements for the vendor, such as

throughput, response time, and service outages. We focus on service outages of end-to-end services, which directly affect the customer's business performance; however, our findings can in many cases be generalized to also cover other performance dimensions.

In practice, performance requirements for system outages are typically expressed by two specifications: the minimum availability of the service over a given period of time and the maximum duration of any single service outage incident (see, e.g., [5]). For example, the minimum availability might be expressed as 99% uptime every month and the maximum incident duration might be specified as 30 minutes. Let us call this requirements specification method the *(min, max)* regime, where the target values for *min* and *max* are called *service level objectives* (SLOs).

We address two primary research questions in this paper: First, we analyze why the *(min, max)* method of performance specification of service outages leaves purchasers under-informed and why, as a consequence (under reasonable assumptions), decisions based on this method may be significantly suboptimal. We discuss how providers' private information about the behavior of service delivery environments could be leveraged by the *customer* when selecting service offers. Second, we examine how the customer could optimize its total cost in choosing from different service offers by risk-neutral providers. We propose a procurement auction as a mechanism to support customers in determining the cost-optimal solution to their optimization problem considering the negative monetary impact of potential service incidents on their business.

The remainder of the paper proceeds as follows: Section 2 briefly reviews related literature. We introduce the base scenario, which we address later on, in Section 3. In Section 4, we analyze the monetarily quantified negative impact of selected incident duration distributions on the performance of a customer business process. We discuss how a provider could support the customer in solving its decision problem in Section 5. In Section 6, we propose a service procurement auction approach that enables the customer to select its cost-optimal service solution. Finally, Section 7 summarizes and discusses our approach, and outlines the next steps towards application of these ideas in industry.

2 Related Work

We conducted a literature review following the methodology proposed by Webster and Watson [6]. In an extensive forward and backward search without temporal restriction, Google Scholar and CiteSeer served as our main sources of search. Keywords used were 'service level objective', 'service level agreement', 'service level management', 'optimal choice', 'business impact', 'incident management', 'decision theory' and 'customer' as well as combinations thereof.

Focusing on the research field of IT (outsourcing) we found six papers – by Jin et al. [7], Sauvé et al. [8], Cheng et al. [9], Taylor and Tofts [10], Barroero et al. [11] and Franke [12] – which are related to our approach that aims to support the consideration of adverse business impact in service offers. None of these works, however,

closely matches our scenario and approach. Our own prior work contains discussions of these papers and is more directly relevant:

In [13] we discuss the customer's optimization problem to select the cost-optimal SLA having concluded a long-term outsourcing contract with a single external provider. We propose a methodology to solve the optimization problem in a scenario without penalties. In [14] we extend that approach and discuss the customer's optimization problem in a setting where a provider has to compensate the customer for every single service incident to occur. We propose a single-attribute procurement auction approach to identify and select the cost-optimal service offer out of different providers' bids. In contrast to this work, our prior work does not consider the cases of providers bidding tuples of target values for service level indicators (i.e., performance measures) and a price (multi-attribute procurement auction) or of receiving incentive bonus payments for 'better-than-promised' performance. In [15] we propose a simulation-based procedure to monetarily quantify the negative impact of single service incidents on the customer business.

Furthermore, there are several areas of research complementary to our work: Taking a more technical perspective, Wittgreffe et al. [16] provide concepts aiming at the design of end-to-end SLAs that are directly targeted at business applications of the service customer. Breitgand et al. [17] present an SLA design approach which enables service providers to derive optimal response time SLOs for a service by analyzing historical performance data of the applied IT infrastructure. Blau et al. [18] use a multidimensional procurement auction to determine the socially efficient service composition in a service value network aiming at the maximization of welfare.

Summing up, none of the related works closely matches our scenario and approach, namely: (i) We monetarily measure the negative business impact of service offers by (ii) quantifying the incurred business cost resulting from the occurrence of combinations of service incidents and (iii) explicitly considering business cost functions which do not develop linearly with service performance attributes. We do this in a setting in which providers have private information they do not want to disclose (iv).

3 Base Scenario

In our base scenario, which we address later on, a customer company wants to purchase an end-to-end IT service from an external service provider to support a particular business process. This non-mission-critical, yet non-commodity service directly affects the customer's business performance, i.e., it contributes measurably to the business value created. Competing providers are able to offer the particular service with different SLAs, i.e., at different service levels and different prices.

We assume that all functional and non-functional properties (including base sizing) of the service, except for performance and price, have already been documented in a 'Request for Proposal' (RfP). For this purpose, requirements of the business process to be supported have been considered. The RfP also states a fixed delivery duration, which may be split into a number of reference periods.

To define service performance the customer has specified a set of service level indicators (performance measures), and associated service level objectives (target values). An SLA is represented by a tuple of service level objectives and a service price, which the customer will have to pay to the selected provider on a periodic basis for service provisioning. In the RfP, the customer has also defined penalty functions for each service level indicator. These specify the amount of money a provider will have to pay in case of service level breaches. Further, the RfP grants the customer the right of constant access to the monitoring information that the provider collects.

To compare service providers' offers, we assume the customer will estimate and monetarily quantify the negative impact of imperfect service (at specific service levels) on its business process. Also, the customer is assumed to consider its expectations of providers to achieve, to under- or even outperform the stated SLOs (cf. [13]). Using these 'business cost' estimates, the service price, as well as the expected penalty payments, the customer can assess the financial implications resulting from the selection of a specific SLA, i.e., it can determine total customer cost. We suppose that the rational customer in our setting aims to identify the cost-optimal service offer.

Since customer and provider have agreed on a fixed delivery duration, capital budgeting approaches, such as the net present value method, can be applied to discount cash flows (i.e., total customer cost) of several future reference periods. This allows the comparison of SLAs which state service prices and business cost changing over time, i.e., varying in different reference periods. For reasons of clarity, however, we will focus on a single reference period.

4 The Adverse Business Impact of Service Incidents

The (*min*, *max*) performance measures normally found in SLAs for specifying requirements for service outages may be called *aggregating* and *limiting* service level indicators. An example: *A logistics company C runs a high-bay warehouse with a management system that is operated and managed by the provider P as a service. This end-to-end service is defined to be available 99% of the time of a reference period [this is the SLO of the aggregating service level indicator min], whereas no single service incident may last longer than 30 minutes [this is the SLO of the limiting service level indicator max].*

The information lost in the aggregation may, however, have undesirable consequences for making decisions – as will be illustrated in the following. Using these established service level indicators the customer has no information about the specific probabilities of a service to achieve, to under- or even outperform the stipulated SLOs. Consequently, the customer is not able to determine with any precision the business cost resulting from the selection of an SLA. Even if the service level objective combinations of aggregating and limiting service level indicators are exactly met, customers may not be able to determine the adverse business impact. Let us see why.

Adverse business impact, resulting in business cost, is caused by each service incident (SI), i.e., by each 'unplanned interruption' to a service or 'reduction in the quality' of a service, which is observable from a business point of view (adapted from the

definition of ‘incident’ in [19]). We use business cost functions to describe the business cost induced by a single service incident of a certain type with regard to its incident attribute values. These business cost functions may develop non-proportionally to service incidents’ attribute values, i.e., they do not need to increase or decrease linearly.

Example (cont’d): *The warehouse management system is used to locate goods in the high-bay warehouse that have to be taken out of stock and shipped by truck. Every five minutes, one truck arrives at the warehouse in order to be loaded. Outages of the warehouse management system (i.e., service incidents of the type ‘outage incidents’, which are described by the single attribute ‘outage duration’) interrupt this workflow. For each minute a truck has to wait to be loaded, C incurs business cost of € 1. Figure 1 shows how C’s aggregated business cost increases with longer outage durations.*

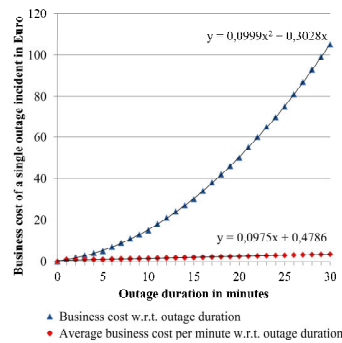


Fig. 1. Business cost of a single outage incident w.r.t. outage duration

The blue triangles in Figure 1 represent the (discrete) business cost values of single outage incidents with respect to their duration.¹ The red diamonds in Figure 1 describe the average business cost per minute an outage incident causes with respect to its duration.² For instance, an outage incident that lasts 25 minutes induces business cost of € 75, i.e., average business cost of € 3 per minute of outage.

When an aggregating service level indicator is used, the achieved service quality is calculated considering the values of a specific attribute of all service incidents of the same type (service incident class) that occur within a certain reference period. In order to quantify business cost in case of moderately non-linear business cost functions³, however, it is essential for the customer to understand which combination of incident attribute values actually realizes a specific *achieved* value (of an aggregating

¹ The discrete business cost values (blue triangles) can be approximated by the continuous business cost function $y=0.0999x^2+0.3028x$.

² The average business cost per minute values (red diamonds) can be approximated by the continuous function $y=0.0975x+0.4786$.

³ We obtained our results below assuming a moderately non-linear business cost function. Due to space limitations we cannot report a full sensitivity analysis. The business cost function we employ, however, is realistic and reasonably robust.

service level indicator). The distribution of single incidents' attribute values, e.g., of their outage durations, has to be considered.

To illustrate the potential consequences of different incident attribute value distributions on business cost, we continue our example:

We assume the provider to exactly achieve the target availability of 99% (i.e., 432 minutes of total outage time within a month) and to avoid single service incidents lasting longer than 30 minutes.

Let O_T be the total outage time permitted and x_{max} be the maximum length of a single outage incident. Furthermore, let x_i be the length of outage i during the reference period T and let $b(x_i) = 0.0999x_i^2 + 0.3028x_i$ be the non-linear business cost function. Now assume $x_i \leq x_{max}$, \forall_i , and, $x_i = x_j$, $\forall_{i,j}$.

If all outages occurring in T have the same duration x_{fix} this leads to significantly different total business cost for different values of x_{fix} (see Table 1):

Table 1. Different outage durations lead to significantly different business cost

x_{fix}	$b(x_{fix})$	number of incidents within T (rounded)	total business cost caused within T
10	€ 13.02	43	€ 559.77
20	€ 46.02	21	€ 996.34
30	€ 98.99	14	€ 1,385.92

Having shown the impact of different outage durations using this simple calculation we can now discuss a more complex setting: Figure 2 shows the relative frequencies of different outage incident durations for selected beta-distributions⁴.

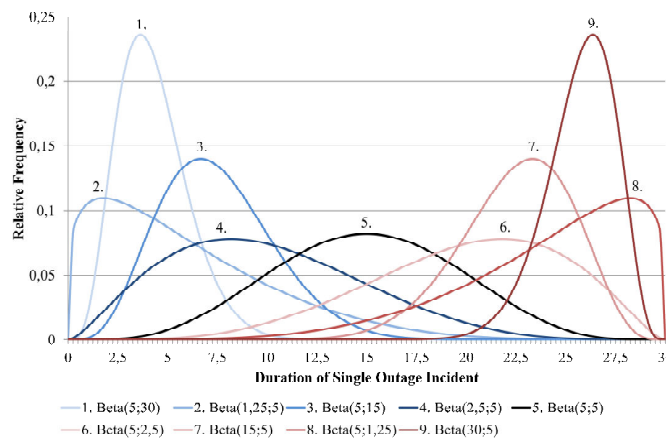


Fig. 2. Relative frequencies of beta-distributed outage incident durations on the interval (0;30]

⁴ We chose the beta-distribution because it can be specified within definite limits, it is very flexible and is analytically tractable. The beta-distribution allows to model a large variety of incident behaviors of service delivery environments, e.g., that of an environment which is characterized by many short outages, some outages of medium length and few long outages.

Since no single service incident may last longer than 30 minutes (SLO of the limiting service level indicator) the distributions are limited to the interval (0;30]. By assuming the provider exactly meets the SLOs promised, we are able to calculate the business cost a customer has to expect if these targets are realized through different beta-distributions of single outage durations.

We compute the following integral (see equation 1) for the given beta-distributions, where $f_{Beta}(x, \alpha, \beta, 0, 30)$ is the probability density function of a beta-distribution which is limited to the interval (0;30] and $g_{AvgBC/Min}(x)$ describes the average business cost per minute (depicted as red diamonds in Figure 1):

$$\int_0^{30} f_{Beta}(x, \alpha, \beta, 0, 30) * 432 * g_{AvgBC/Min}(x) dx \quad (1)$$

The probability density function is multiplied with the total outage duration, i.e., 432 minutes, in order to obtain the time which is spent on a certain incident attribute value. Figure 3 shows the resulting business cost values.

In case of non-constant marginal business cost functions, the customer has to consider the distribution of incident attribute values (a target value of an aggregating service level indicator is realized through), since different distributions will usually lead to different business cost.

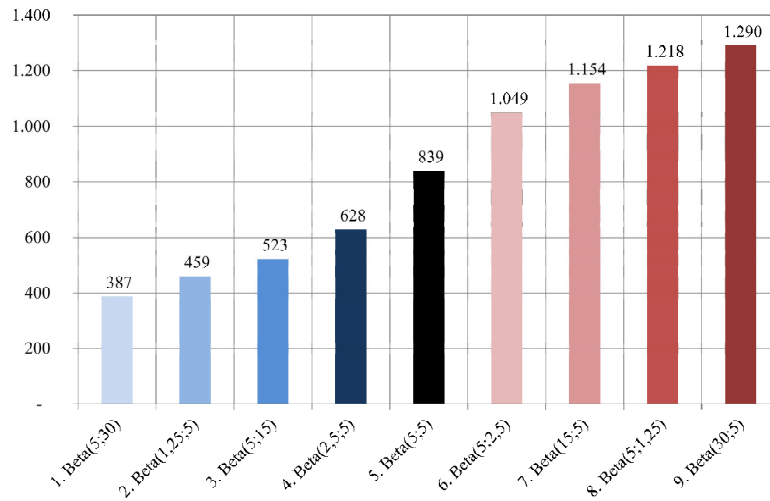


Fig. 3. Expected business cost given different beta-distributions for single outage durations

Even if the customer in our example knew that incident durations are beta-distributed and assumed a specific beta-distribution in order to estimate its expected business cost, this could lead to a significant misjudgment in case the customer selects inaccurate beta-distribution parameter values.

Table 2 shows the errors with regard to expected business cost (in percent) resulting from the assumption of a specific beta-distribution, while the actual outage duration distribution follows a different beta-distribution.

Table 2. Error with regard to expected business cost in percent if a ‘wrong’ beta-distribution is assumed

		Assumed beta-distribution				
		B(5;30)	B(5;15)	B(5;5)	B(15;5)	B(30;5)
Actual beta-distribution	B(5;30)	0.0 %	35.0 %	116.5 %	198.1 %	233.1 %
	B(5;15)	-25.9 %	0.0 %	60.4 %	120.9 %	146.7 %
	B(5;5)	-53.8 %	-37.7 %	0.0 %	37.7 %	53.8 %
	B(15;5)	-66.5 %	-54.7 %	-27.4 %	0.0 %	11.7 %
	B(30;5)	-70.0 %	-59.5 %	-35.0 %	-10.5 %	0.0 %

Consequently, the definition of SLOs for aggregating and limiting service level indicators is not sufficient to compare different service offers regarding the total business cost (and, thus, total customer cost) induced in the case of non-linear business cost functions.⁵ Performance measures have to consider the monetarily quantified negative business impact of service incidents instead of incident attribute values only.

Furthermore, the example shows that it would be negligent from a customer point of view not to reflect distributions of incident attribute values in service level agreements in case of moderately non-linear business cost functions. This idea, however, hinges on the providers’ ability to ascertain incident attribute value distributions at reasonable (additional) cost and, thus, price increases.

5 Service Incident Patterns

Providers usually possess private information about the service delivery environments (SDEs), i.e., combinations of organizational and technical service components they use to provide a service. Since outsourcing providers have many customers who purchase similar services, they are able to collect data about the frequency of service incidents that occur when using a particular type of SDE, i.e., to establish service incident histories for each SDE. By analyzing these data, providers can predict the frequency distributions of service incidents of a specific type having certain incident attribute values, i.e., determine the SDEs’ characteristic ‘service incident patterns’. If the customer had the deep insight a provider has, it could use this information to better address its optimization problem. In the following we will denote a tuple of attribute values of a single service incident as its service incident level (SIL), with the incident type determining the tuple structure.

Considering one service incident type⁶ only, Table 3 illustrates a provider’s private knowledge of an exemplary service delivery environment e . It lists in schematic form

⁵ ... even if we assume the service level objective combinations of aggregating and limiting service level indicators to be met. In practice, the stipulated SLOs might be even outperformed to certain degrees in different reference periods.

⁶ Service incidents sharing the same set of (one or more) attributes are regarded to be of the same service incident type (e.g. outage incidents, reduced throughput incidents, etc.). Incidents of the same service incident type affect the performance of a customer business process in a similar manner, but may differ in their attribute values (e.g., in the duration of outage (outage incident) or the degree of throughput reduction (reduced throughput incident)).

the data records about the combinations of sequential service incidents (CoSIs) the provider has collected about this type of service delivery environment when using it (i) during past reference periods (ii) for one or several other customers. The function $f_t(\text{SIL}_{a, b, \dots, n})$ describes the absolute frequency with which service incidents at the levels $\text{SIL}_{a, b, \dots, n}$ occurred during a specific reference period, whereas the index t labels a data record. We denote this table as CoSI/SIL frequency history.

To continue with our example: A row CoSI_2 in the CoSI/SIL frequency history could, for instance, document that there were five outage incidents with a duration in the interval $(0;5)$ minutes ($f_2((0;5))=5$), three incidents with a duration in the interval $[5;10)$ minutes ($f_2([5;10))=3$), ... in the data record '2'.⁷

Based on the information contained in the CoSI/SIL frequency histories, the provider determines the SDEs' characteristic incident patterns (one pattern for each service incident type), which state the expected (absolute) frequencies of service incidents at the different service incident levels to occur within a reference period.⁸

Table 3. A provider's private knowledge of the incident characteristics of a specific service delivery environment e (CoSI/SIL frequency history)

		Absolute frequencies of SIs at a certain SIL (historical data)				
		$\text{SIL}_{1,1, \dots, 1}$...	$\text{SIL}_{1,1, \dots, n}$...	$\text{SIL}_{a,b, \dots, n}$
Service Delivery Environment e	CoSI_1	$f_1(\text{SIL}_{1,1, \dots, 1})$...	$f_1(\text{SIL}_{1,1, \dots, n})$...	$f_1(\text{SIL}_{a,b, \dots, n})$
	CoSI_2	$f_2(\text{SIL}_{1,1, \dots, 1})$...	$f_2(\text{SIL}_{1,1, \dots, n})$...	$f_2(\text{SIL}_{a,b, \dots, n})$

	CoSI_t	$f_t(\text{SIL}_{1,1, \dots, 1})$...	$f_t(\text{SIL}_{1,1, \dots, n})$...	$f_t(\text{SIL}_{a,b, \dots, n})$

Providers have only little understanding of the business cost the customer incurs when a certain CoSI is realized. In our work we assume that the customer is able to describe business cost functions, which define the business cost caused by single service incidents with respect to their specific service incident levels (SILs). That is, the customer can state (discrete) BusinessCost/SIL vectors for each service incident type.

Example (cont'd): A customer's BusinessCost/SIL vector for the service incident type 'outage' states that outages with a duration in the interval $(0;5)$ minutes induce business cost of € 1.81, outages with a duration in the interval $[5;10)$ minutes induce business cost of € 8.82, ...

Furthermore, in the following, we assume the customer discloses this information to selected providers, treating them as 'trusted suppliers'. Knowing the customer's BusinessCost/SIL vectors (one vector for each service incident type), providers can combine the information contained therein with their private knowledge of service delivery environments' service incident patterns. In doing so, they are able to calcu-

⁷ The classification of attribute values into attribute intervals reduces the number of service incident levels to be considered and, thus, improves the applicability of our approach.

⁸ In future works, we will focus on the determination of service incident patterns from CoSI/SIL frequency histories.

late (provider-internal) indicators describing their service delivery environments' impact on the customer business process in terms of business cost.

In order to determine the monetary adverse business impact – for a single service incident type and a certain service delivery environment – a provider multiplies the characteristic service incident pattern p with the BusinessCost/SIL vector b . Equation 2 formally describes this operation.

$$p * b = r, \text{ where } p \in \mathbb{R}^s, b \in \mathbb{R}^s, r \in \mathbb{R} \quad (2)$$

The resulting value r represents the business cost value the customer would incur in a specific reference period for a specific type of incident. Assuming historical data to allow the prediction of service incident patterns to occur in future reference periods and service incident patterns to be stable – i.e., to have low variances with regard to frequencies of service incidents at certain SILs to occur – providers can arrive at accurate business cost estimates. These measures reflect the adverse business impact a customer incurs in case of constant *and* non-constant marginal business cost functions. Using this private information, providers can better understand the impact of their services on customers' business and, thus, define their service offers accordingly.

6 A Multi-Attribute Procurement Auction to Select Cost-optimal Service Offers

Based on the constructs introduced in the previous sections, we develop a procedure enabling a customer to solve its optimization problem – given different service offers by risk-neutral providers.

Today, a typical contract negotiation in IT outsourcing follows a structured approach [20]. First, by sending potential service providers a request for information, the customer identifies suppliers that offer the service in question. Afterwards, the customer defines a request for proposal specifying additional information about technical, economic and legal characteristics (sizing, mission-criticality, service partnership models, etc.) and further reduces the number of potential partners before the actual negotiation process begins.

Our proposal is to use a service procurement auction to conduct the negotiation. Auctions constitute a way to elicit bidders' private information [21]. In our scenario we use an auction to allow the customer to draw conclusions regarding the providers' reservation prices for delivering their services at characteristic service incident patterns. Thus, we enable the customer to better compare different service offers. In procurement auctions providers 'compete for the right to sell' their services [21].

For reasons of clarity, the method presented focuses on a single reference period only. As discussed in Section 3, capital budgeting approaches could be applied to compare multi-reference period service contracts, which state service prices and business cost changing over time.

6.1 Definition of the Procurement Auction

The objective of our approach is to compare service offers in order to identify and select the one that is minimizing the total customer cost. Therefore, we define an ‘optimal mechanism’ [21] for the procurement of the service.

In the following, we assume the customer reveals its business cost functions to the selected providers participating in the procurement auction. Providers taking part in the procurement auction have to bid a tuple of service incident patterns (one pattern for each service incident type with regard to service incident levels predefined by the customer) and service price.

A provider in our example, for instance, would bid a tuple consisting of an ‘outage incident pattern’ p (stating the absolute frequencies for outage incidents at the different service incident levels to occur) and the service price it demands with regard to this pattern.

We define the following two rules in order to make sure that providers have no incentive to state service incident patterns of SDEs untruthfully in their bids:

If the amount of service incidents at a specific service incident level exceeds the number stated in the service incident pattern (i.e., a provider’s bid) we define that the future contractual partner will have to compensate the customer for every additional service incident to occur in the amount of business cost defined in the corresponding customer business cost function. Due to this ‘1:1’-penalty rate, providers would directly add the additional business cost to be expected (for all types of service incidents) to their service prices.

On the other hand, we define that the future contractual partner (provider) will be rewarded for every single incident at a specific SIL it avoids (with regard to the number specified in the service incident pattern) in the amount of business cost defined in the corresponding customer business cost function. Due to this ‘1:1’-bonus rate, providers in a competitive situation would give the customer a discount for business cost they expect to avoid considering future rewards.

Since the customer has defined all types of service incidents in a way that these are independent of one another, providers will add up all expected penalties (positive) and rewards (negative) and the net service price to a single monetary value (the gross service price). As mentioned above, providers are supposed to bid a tuple of service incident patterns and gross service price.

Therefore, we can use a multi-attribute auction to determine the cost-optimal service offer for the customer. In such a ‘multi-attribute context’ where the customer reveals the negative impact of service incidents on its business and the providers have different ‘market expectations’, first-price auctions seem to achieve the best results [22].

Based on these findings, we define a multi-attribute, first-price, sealed-bid auction for our setting (see e.g., [21]). In our first-price auction, each provider is asked to submit one sealed bid, which states a tuple of service incident patterns and service price it offers. The provider who submits the bid leading to lowest total customer cost wins the auction and has to be prepared to deliver the service as described in the RfP, with the promised service incident patterns and at the offered gross service price.

6.2 Conducting the Procurement Auction

We now briefly describe the procurement auction.

- 1. Definition of service incident types (customer):** For the service in question, the customer formally describes the different business-relevant types of service incidents including their corresponding sets of attributes. Further, the specific value range of each attribute is defined.
- 2. Definition of service incident levels (customer):** The customer divides the value range of each attribute (characterizing a single service incident type) into a set of disjoint intervals. Therefore, the endpoints of each interval are defined in a way that attribute values leading to significantly different adverse business impacts are elements of different intervals. Then, the customer calculates the Cartesian product of the different sets of attribute intervals for each service incident type, i.e., all service incident levels to be considered are defined.

Example: The service incident type ‘reduced throughput incident’ is characterized by the two attributes ‘duration of throughput reduction’ (measured in minutes) and ‘degree of throughput reduction’ (measured in percent) with the value ranges $(0;c]$ and $(0;z]$. Considering the negative business impact of different attribute value combinations the customer now divides the range of the duration interval into three parts – $(0;a)$, $[a;b)$ and $[b;c)$ – and the range of the degree interval into four parts – $(0;w)$, $[w;x)$, $[x;y)$ and $[y;z]$. Calculating the Cartesian product of these two sets of attribute intervals, twelve service incident levels (3×4) are defined.

This step is repeated for all types of service incidents of the service in question.

- 3. Business cost assessment (customer):** The customer considers the negative business impacts, which service incidents at the different service incident levels (SIL) might cause, and assesses these monetarily. Consequently, for each service incident type the customer specifies a BusinessCost/SIL vector stating the business cost for each SIL defined in step 2. That is, the (discrete) business cost function regarding each service incident type is defined.
- 4. Invitation of providers (customer):** The customer invites a number of selected providers to take part in the procurement auction and sends the RfP including the business cost functions to these trusted suppliers.
- 5. Definition of bids (providers):** According to the procedure introduced in the previous section, each provider determines the characteristic incident patterns (one pattern for each service incident type) for the SDE it wants to offer. Furthermore, it calculates the expected penalty and benefit payments it assumes to have to pay / receive. Due to the competitive auction setting the providers will add / subtract these expected values to / from the net service price. That is, the resulting gross service price reflects all expected cash flows.
- 6. Bidding (providers):** The providers submit their sealed bids, i.e., tuples of service incident patterns and service price, to the customer.
- 7. Winner determination (customer):** Among all service offers the customer identifies the bid resulting in the lowest total customer cost. Therefore, the customer multiplies the absolute frequencies for service incidents at certain SILs to occur as stated in the service incident patterns (first part of a provider bid) with its BusinessCost/SIL vectors (one vector for each service incident type). It adds the re-

sulting business cost value to the gross service price (second part of a provider bid) and, thus, determines the total customer cost.

The provider bid winning the procurement auction determines the total customer cost (fixed value), since the winning provider has to compensate the customer for negative deviations from the service incident pattern offered but is rewarded for positive deviations as well. If the provider is rewarded, on the customer's side business cost are avoided in the same amount. Thus, the winning provider is rewarded if the service in question is delivered at a higher quality than stated in the service incident patterns, and is penalized if realized service quality is lower.

7 Conclusion and Outlook

To summarize, we have addressed a service customer's optimization problem of identifying and selecting a cost-optimal service offer meeting the required level of service criticality and assuming typically non-linear business impact behavior.

First, we analyzed the information that SLA decisions are typically based upon and illustrated deficiencies in the service level indicators usually used today. With regard to our first research question, we showed that the traditional (*min, max*) indicator regime for measuring service performance is not sufficient to support proper decision making by the customer in case of quantified negative business impact to develop non-linearly with service quality. Furthermore, we discussed how providers' private information about the behavior of service delivery environments could be leveraged by the customer when selecting service offers.

In order to address our second research question, we developed a procedure enabling customers to solve their optimization problem – given different service offers – using a procurement auction. Our proposed seven-step auction approach enables providers to calculate business cost measures and, thus, to predict their services' quantified negative business impact using the customer's business cost estimates. Thus, providers can tailor service offers according to customers' business requirements and serve their clients' needs more profoundly. Using the information contained in the customer's business cost functions as the basis for penalty and bonus rules, providers are free to balance resources they use to deliver a service. Moreover, the application of our procedure enables the customer to compare service offers regarding their expected adverse business impact and, thus, to select the total customer cost-optimal solution of its optimization problem – in case of linear and non-linear business cost functions.

Having put forward our approach, we are well aware of a number of limitations and challenges. First, we assumed providers are able to determine the discrete incident attribute value distributions (e.g., regarding outage lengths) when using specific service delivery environments. In order to demonstrate this capability, we are currently working on a data mining approach using monitoring data. Second, we assumed the customer is able to determine its business cost functions for all types of service incidents. We realize that a precise determination of business cost functions may be difficult to achieve. Nevertheless, we argue that an economically well-founded investment

decision about the service offer to purchase has to be made based on information about business cost. Even a rough estimate of business cost functions might significantly improve a decision with regard to total customer cost. Third, we assumed providers to be risk-neutral. This allows us the simplification of adding expected penalties to and subtracting expected bonus payments from the net service price when defining service offers, i.e., without adding risk premiums. This should be a realistic assumption at least for large providers supporting numerous, heterogeneous customers as they can spread and diversify their risk. In addition, we assumed that for non-mission-critical, yet non-commodity services, providers will accept a '1:1 penalty rate' as well as a '1:1 bonus rate'. These rates ensure that business cost information contained in providers' offers is not distorted when service prices are defined and that providers have no incentive to state service incident patterns untruthfully. The penalty rate will be accepted by providers, since they can add all expected penalties to the service price. In future work we plan to analyze the influence of different provider risk preferences on the solution of the customer optimization problem. Finally, we assumed the customer to reveal its business cost functions, which indicate the criticality of the service for its business. The competition among providers, which participate in the procurement auction, will prevent them from taking advantage of this information and claiming the complete consumer surplus, which results from different service quality and service price combinations, leading to specific total customer cost.

We are convinced that our approach, enabling customers to compare different service offers and to select the cost-optimal service solution, provides valuable insights for both, service providers and their customers. It will help both parties to address the challenge of business and IT alignment in a different way – exploiting analytical insights gained for service delivery environments. At the same time, the limitations mentioned above leave a broad field for further research.

References

1. Gartner: Gartner Says Worldwide IT Outsourcing Market Grew 7.8 Percent in 2011, <http://www.gartner.com/it/section.jsp>
2. Unterharnscheidt, P., Kieninger, A.: Service Level Management: Challenges and their Relevance from the Customer's Point of View. In: Proceedings of the 16th Americas Conference on Information Systems, Paper 540. Association for Information Systems, Lima, Peru (2010)
3. Lacity, M.C., Kahn, S.A., Willcocks, L.P.: A review of the IT outsourcing literature: Insights for practice. *The Journal of Strategic Information Systems* 18 (3), 130–146 (2009)
4. Luftman, J., Ben-Zvi, T.: Key Issues for IT Executives 2010: Judicious IT Investments Continue Post-Recession. *MIS Quarterly Executive* 9 (4), 263–273 (2010)
5. Sturm, R., Morris, W., Jander, M.: Foundations of Service Level Management. SAMS, Indianapolis, Indiana (2000)
6. Webster, J., Watson, T.: Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* 26 (2), xiii–xxiii (2002)
7. Jin, L., Machiraju, V., Sahai, A.: Analysis on Service Level Agreement of Web Services. Technical report HPL-2002-180, Hewlett-Packard Laboratories, Palo Alto, California (2002)

8. Sauv e, J., Marques, F., Moura, A., Sampaio, M., Jornada, J., Radziuk, E.: SLA design from a business perspective. In: Sch onw alder, J., Serrat, J. (eds.): Proceedings of the 16th International Workshop on Distributed Systems, Operations and Management. LNCS, Vol. 3775, pp. 72–83. Springer, Barcelona (2005)
9. Cheng, F., Gamarnik, D., Jengte, N., Min, W., Ramachandran, B.: Modeling Operational Risks in Business Processes. IBM Research Report RC23672, Thomas J. Watson Research Center, Yorktown Heights, New York (2005)
10. Taylor, R., Tofts, C.: Death by a thousand SLAs: a short study of commercial suicide pacts. Technical report, Hewlett-Packard Laboratories, Bristol (2005)
11. Barroero, T., Motta, G., Pignatelli, G., Bochicchio, M., Longo, A.: Aligning IT service levels and business performance: A case study. In: Proceedings of the International Conference on Services Computing, pp. 570–577. IEEE Computer Society, Miami, Florida (2010)
12. Franke, U.: Optimal IT Service Availability: Shorter Outages, or Fewer?. *IEEE Transactions on Network and Service Management* 9 (1), 22–33 (2012)
13. Kieninger, A., Satzger, G., Straeten, D., Schmitz, B., Baltadzhiev, D.: Business Cost Budgets: A Methodology to Incorporate Business Impact into Service Level Agreements. *International Journal of Service Science, Management, Engineering, and Technology* 3 (3), 49–64 (2012)
14. Kieninger, A., Schmitz, B., Straeten, D., Satzger, G.: Incorporating Business Impact into Service Offers: A Procedure to Select Cost-Optimal Service Contracts. In: Proceedings of the 18th Americas Conference on Information Systems, Paper 19. Association for Information Systems, Seattle, Washington (2012)
15. Kieninger, A., Berghoff, F., Fromm, H., Satzger, G.: Simulation-based Quantification of Business Impacts Caused by Service Incidents. In: Cunha, J.F., Snene, M., N voa, H. (eds.): IESS 2013 Porto. LNBIP, Vol. 143, pp. 170–185. Springer, Berlin et al. (2013)
16. Wittgreffe, J., Dames, M., Clark, J., McDonald, J.: End-to-end service level agreements for complex ICT solutions. *BT technology journal* 24 (4), 31–46 (2006)
17. Breitgand, D., Henis, E., Shehory, O., Lake, J.M.: Derivation of Response Time Service Level Objectives for Business Services. In: 2nd International Workshop on Business-Driven IT Management, pp. 29–38. IEEE Press, Munich (2007)
18. Blau, B., Conte, T., van Dinther, C.: A Multidimensional Procurement Auction for Trading Composite Services. *Electronic Commerce Research and Applications* 9 (5), 460–472 (2009)
19. OGC, Office of Government Commerce: IT Infrastructure Library (ITIL): Service Operation. The Stationary Office (TSO), London, UK (2007)
20. Br utigam, P.: IT-Outsourcing: Eine Darstellung aus rechtlicher, technischer, wirtschaftlicher und vertraglicher Sicht. 2nd edition, Schmidt, Berlin (2009)
21. Krishna, V.: Auction Theory. 2nd edition, Academic Press, Amsterdam (2010)
22. Bichler, M.: An experimental analysis of multi-attribute auctions. *Decision Support Systems* 29 (3), 249–268 (2000)

Sustainability and Security

Hans-Knud Arndt¹ und Kai Rannenberg²

¹ Otto-von-Guericke-Universität, Magdeburg

² Goethe-Universität, Frankfurt am Main

Nachhaltigkeit und Sicherheit sind gesellschaftlich breit diskutierte Themen. Auch im Bereich der Informations- und Kommunikationstechnik (IKT) haben sie durch Herausforderungen wie Green IT oder den Schutz kritischer Infrastrukturen an Bedeutung gewonnen. Dass Nachhaltigkeit und Sicherheit auch in der Wirtschaftsinformatik Einzug gehalten haben, zeigt sich z. B. in den Schlagworten „Nachhaltiges Informationsmanagement“ oder „Informationssicherheit“.

Nachhaltigkeit und mehrseitige Sicherheit sind eine spannende Kombination. Einerseits müssen funktionierende Sicherheitssysteme ohnehin für einen nachhaltigen Ausgleich der Interessen der Beteiligten und den langfristigen Schutz von Daten, etwa in Archiven, Grundbüchern oder Zertifikaten sorgen, weil sie sonst nicht stabil sind. Andererseits gehen Green IT oder effizientes Energie- und Ressourcenmanagement meist mit der Verarbeitung sensibler, sensitiver und oft auch personenbezogener Daten einher, was erhebliche Herausforderungen an Datenschutz und Sicherheit stellt.

Die für diesen Track (aus 27 Einreichungen mit einer Akzeptanzquote von 29,63 %) ausgewählten acht Beiträge berühren diese Themen und in ersten Ansätzen auch ihre Kombination.

Die Autoren Opitz, Ere, Thies, Kolbe und Zarnekow analysieren in ihrem Beitrag „Kennzahlenbasierte Erfolgsmessung von Green-IT-Maßnahmen – Eine empirische Analyse zum aktuellen Stand in Forschung und Praxis“ den Umsetzungsstand von Kennzahlen für den Bereich der Green-IT von Organisationen unterschiedlicher Größe und Branche und kommen auf der Grundlage einer Befragung zu dem Schluss, dass der Einsatz von Green-IT-Kennzahlen einerseits grundsätzlich möglich ist, der konkrete Einsatz andererseits aber von der jeweiligen Unternehmensgröße, der Nachhaltigkeitsorientierung und der Green-IT-Expertise abhängig ist.

In dem Beitrag „Experimental Evaluation of a Process Benchmarking Tool in a Green Business Process Management Context“ von den Autoren Gräuler und Teuteberg wird ein neuer Ansatz für ein Benchmarking von Geschäftsprozessen/ Semantisches Prozessmanagement („Semantic Business Process Benchmarking“) für ein umweltbezogenes Geschäftsprozessmanagement („Green Business Process Management“) vorgeschlagen, indem Metamodelle, Ontologien, Umweltleistungskennzahlen und Metriken miteinander verknüpft werden. Da ein solches Benchmarking üblicherweise ein zeit- und kostenintensives Vorgehen darstellt, wird in diesem Beitrag auf eine (teilweise) Automatisierung eingegangen, indem in dem vorgeschlagenen Ansatz eine größere Flexibilität bezogen auf die zu verwendende Terminologie, den zu verwendenden Grad der Abstraktion als auch die zu verwendenden Modellierungsnotationen ermöglicht wird.

Der Beitrag „Modellgestützte Analyse und Optimierung der Energieeffizienz betrieblicher Informations- und Kommunikationstechnik“ von den Autoren Vaupel und Leukel setzt bei den Unternehmensarchitekturmodellen (EAM) an und stellt zunächst fest, dass die derzeitig vorliegenden Architekturmodelle nur unzureichend die automatisierte Bestimmung der Energieeffizienz und die Bewertung von Maßnahmen zur Erhöhung der Energieeffizienz unterstützen. Deshalb wird in diesem Beitrag eine Methode zur modellgestützten Energieeffizienzanalyse und –optimierung im Rahmen von Unternehmensarchitekturmodellen vorgeschlagen, prototypisch implementiert und durch ein Simulationsexperiment evaluiert.

Die Autoren Widmer, Premm und Karaenke gehen in ihrem Beitrag „Energy-aware Service Allocation for Cloud Computing“ der Frage der Energieeffizienz von Cloud-Computing-Lösungen nach. Auf Grundlage eines spieltheoretischen Forschungsrahmens werden die Handlungen von Anbietern als auch Kunden von IT-Dienstleistungen innerhalb eines Cloud-Computing analysiert. Darauf aufbauend wird ein neuer Allokationsmechanismus für eine energieeffiziente IT-Dienstleistungsallokation vorgeschlagen und evaluiert.

Der Beitrag „Nachhaltigkeit in IT-Organisationen: Ein Forschungsrahmen für das Nachhaltige Informationsmanagement“ von den Autoren Ereik, Löser und Zarnekow stellt zunächst fest, dass in den letzten Jahren Nachhaltigkeitsthemen stärker in das Blickfeld der Wirtschaftsinformatik gerückt sind und auf einen Paradigmenwechsel innerhalb der Wirtschaftsinformatik hindeuten. Um theoretische und konzeptionelle Fragestellungen der Wirtschaftsinformatik vor dem Hintergrund eines nachhaltigen Informationsmanagement besser bearbeiten zu können, schlägt dieser Beitrag einen Strukturierungsrahmen zur ganzheitlichen Adaption von Nachhaltigkeit in Forschung und Praxis der Wirtschaftsinformatik vor.

Die Autoren Loske, Widjaja und Buxmann zeigen in dem Beitrag „Unrealistischer Optimismus der Cloud Computing Anbieter bezüglich IT Sicherheitsrisiken - Eine Bedrohung für die Nutzer?“, dass Anbieter von Cloud Computing die IT Sicherheitsrisiken ihrer eigenen Angebote in vielen Fällen systematisch unterschätzen. Dieses Verständnis der systematischen Verzerrungen bei der Risikowahrnehmung ermöglicht, das tatsächliche Risiko besser zu bewerten und so gezielter Strategien zur Verbesserung der IT Sicherheit im CC zu entwickeln.

Die Autoren Meier, Fuchs und Pernul beschreiben in dem Beitrag „Managing the Access Grid – A Process View to Minimize Insider Misuse Risks“, wie Rollenmodellierung durch ein „Access Grid“ erleichtert werden kann.

Die Autoren Milicevic und Goeken entwickeln in dem Beitrag „Systematic Review and Meta-Analysis of IS Security Policy Compliance Research. First Steps towards Evidence-Based Structuring of the IS Security Domain“ ein Darstellungsformat, auf dem Entwurfsentscheidungen, etwa zu Information Security Policies basieren können.

Die Einreichungen zeigen einerseits beeindruckende Fortschritte, andererseits aber auch die bei Nachhaltigkeit und Sicherheit noch weiter bestehenden Herausforderungen auf, so dass wir Vortragenden und Publikum fruchtbare Diskussionen wünschen.

Dank gebührt den Gutachtern und den Mitgliedern des Programmkomitees, dabei besonders an Dr. Andreas Albers, der immer wieder einsprang und für einen reibungslosen Ablauf sorgte, sowie – last but not least – der Leipziger Organisationsleitung der WI2013.

Programmkomitee

Rafael Accorsi, Albert-Ludwigs-Universität Freiburg
Andreas Albers, Goethe-Universität Frankfurt am Main
Hannes Ferrath, Universität Hamburg
Ulrich Greveler, Hochschule Rhein-Waal
Rüdiger Grimm, Universität Koblenz
Ralf Isenmann, Fraunhofer Institut für System- und Innovationsforschung Karlsruhe
Dogan Kesdogan, Universität Siegen
Ioannis Krontiris, Goethe-Universität Frankfurt am Main
Jorge Marx Gómez, Carl von Ossietzky Universität Oldenburg
Andreas Möller, Leuphana Universität Lüneburg
Stefan Naumann, Umwelt-Campus Birkenfeld
Günter Pernul, Universität Regensburg
Hartmut Pohl, Hochschule Rhein-Sieg
Jens Strüker, Albert-Ludwigs-Universität Freiburg
Frank Teuteberg, Universität Osnabrück
Volker Wohlgenuth, Hochschule für Technik und Wirtschaft Berlin

Unrealistischer Optimismus der Cloud Computing Anbieter bezüglich IT Sicherheitsrisiken – Eine Bedrohung für die Nutzer?

André Loske, Thomas Widjaja, und Peter Buxmann

CASED - Center for Advanced Security Research Darmstadt, Darmstadt, Germany
{andre.loske,thomas.widjaja,peter.buxmann}@cased.de

Abstract. Anbieter von Cloud Computing (CC) Lösungen versprechen zahlreiche technische und ökonomische Vorteile gegenüber klassischen IT Outsourcing Konzepten. Der Paradigmenwechsel hin zu CC induziert jedoch auch neuartige IT Sicherheitsrisiken, die zu erheblichen Bedenken seitens potenzieller Anwender führen. Anbieter von CC Lösungen betonen hingegen stetig die hohen Standards der IT Sicherheit ihrer eigenen Angebote. Auf Grundlage der kognitionspsychologischen Theorie des „unrealistischen Optimismus“ und einer empirischen Untersuchung unter deutschen CC Anbietern zeigen wir, dass Anbieter die IT Sicherheitsrisiken ihrer eigenen CC Angebote in vielen Fällen systematisch unterschätzen. Das Verständnis der systematischen Verzerrungen bei der Risikowahrnehmung ermöglicht Wissenschaftlern, Anbietern und Nutzern das tatsächliche Risiko besser zu bewerten und so gezielter Strategien zur Verbesserung der IT Sicherheit im CC zu entwickeln.

Keywords: Cloud Computing, Unrealistischer Optimismus, IT Sicherheit, psychologische Faktoren, Risikomanagement

1 Einleitung

Outsourcing der Informationstechnologie (IT) an externe Anbieter ist heute Bestandteil der IT Strategie vieler Unternehmen. Cloud Computing (CC) stellt eine Weiterentwicklung klassischer IT Outsourcing (ITO) Konzept unter Verwendung moderner Kommunikationstechnologien dar, das auf zunehmendes Interesse stößt, aber derzeit auch Gegenstand kontroverser öffentlicher und wissenschaftlicher Diskussionen ist. CC ist ein Modell „*for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*“ [1]. Obwohl CC gegenüber klassischen ITO Konzepten eine Vielzahl technischer und ökonomischer Vorteile verspricht, bleibt die Akzeptanz bei den Nutzern weit hinter den Erwartungen zurück [2]. Insbesondere das wiederholte Auftreten aufsehenerregender Zwischenfälle hat potentielle Nutzer in den letzten Jahren hinsichtlich der IT Sicherheitsrisiken (ITSR) von

CC stark sensibilisiert und in vielen Fällen langfristig von der Nutzung abgeschreckt [3]. Im April 2011 haben bspw. hunderte Kunden einen Großteil ihrer gespeicherten Daten durch einen fatalen Systemabsturz der Amazon EC2 Dienste verloren. Wenige Monate später verursachte ein Blitzschlag einen über 48 stündigen Ausfall von Microsofts CC Dienst "Business Productivity Online Suite". Während des gesamten Zeitraums hatten die Nutzer weder Zugriff auf Ihre E-Mails, Kalender und Kontakte noch auf Ihre Dokumente im Managementsystem [4]. Sicherheitsexperten betonen allerdings das Vorhandensein geeigneter Konzepte und Maßnahmen, die einen umfassenden Schutz der CC Angebote hinsichtlich der ITSR gewährleisten können [5]. Angesichts der Häufung sicherheitsrelevanter Zwischenfälle in der Cloud stellt sich die Frage, ob die Anbieter die ITSR erheblich unterschätzen und in Folge erforderliche Sicherheitsmaßnahmen und -konzepte nicht hinreichend umsetzen?

Eine mögliche systematische Unterschätzung der ITSR durch die CC Anbieter beruht im Wesentlichen auf dem Umstand, dass nicht das tatsächliche Risiko, sondern das seitens einer Person wahrgenommene Risiko, bspw. durch den Sicherheitsbeauftragten oder CIO, die Risikobewertung für ein Unternehmen bestimmt. Die Höhe des wahrgenommenen Risikos weicht allerdings oftmals signifikant vom tatsächlichen Risiko ab, da z. B. erforderliche Informationen fehlen und kognitive Abläufe die Einschätzung unbewusst in bestimmte Richtungen verzerren können [6]. Insbesondere im Bereich der IT existieren i. d. R. keine vergangenheitsbezogenen Daten bzgl. des Schadens und der Auftrittswahrscheinlichkeit bestimmter ITSR, die eine objektive Quantifizierung des tatsächlichen Risikos ermöglichen würden. Das Fehlen quantitativer Daten ist im Wesentlichen darauf zurückzuführen, dass die IT im Allgemeinen einem schnellen technologischen Wandel mit kurzen Produktlebenszyklen unterliegt und außerdem Sicherheitsvorfälle in vielen Fällen nicht entdeckt oder nicht systematisch dokumentiert werden [7]. Das wahrgenommene Risiko eines Entscheidungsträgers basiert dann häufig auf sozialen Vergleichen mit anderen Personen bzw. Unternehmen [8-9]. Menschen tendieren allerdings im Allgemeinen dazu, ihr persönliches Risiko erheblich geringer einzuschätzen als das Risiko der Vergleichsperson. Diese systematische Unterschätzung des Risikos basiert hauptsächlich auf einem abstrakten Gefühl der persönlichen Unverwundbarkeit und wird in der Literatur als „unrealistischer Optimismus“ (UO) bezeichnet [10]. UO in der Risikowahrnehmung konnte bereits von Wissenschaftlern in unterschiedlichen Disziplinen, z. B. bzgl. gesundheitlicher Probleme, Autounfällen oder Kriminalität, nachgewiesen werden [11-13].

Systematische Unterschätzungen eines Risikos sind dabei insbesondere hinsichtlich der reduzierten Motivation zur Umsetzung erforderlicher Präventions- oder Schutzmaßnahmen als überaus kritisch zu betrachten [14-15]. Raucher kennen z. B. im Allgemeinen die, mit dem Rauchen verbunden, Gesundheitsrisiken, wie Lungenkrebs oder Herzinfarkt. Jedoch erst mit dem Bewusstsein, um die persönliche Verwundbarkeit durch, bspw. in Folge des Auftretens erster Symptome, leiten Personen entsprechende Maßnahmen ein [16]. Gleichmaßen benötigen die Verantwortlichen seitens der CC Anbieter zunächst ein Bewusstsein für die Verwundbarkeit ihrer Dienste durch bestimmte ITSR, bevor Entscheidungen für die Umsetzung notwendige und ggf. kostenintensive Sicherheitsmaßnahmen getroffen werden. Das Wissen um

die Existenz eines IT Sicherheitsrisikos allein ist demnach im Allgemeinen nicht ausreichend. Es ergeben sich die folgenden Forschungsfragen für den UO im CC:

1. *Sind die CC Anbieter unrealistisch optimistisch bzgl. der ITSR?*
2. *Hat UO negative Folgen für die Awareness der CC Anbieter bzgl. der ITSR?*
3. *Variiert der Grad des UO der CC Anbieter in Abhängigkeit der Charakteristika einzelner ITSR?*

In der vorliegenden Studie zeigen wir anhand der spezifischen ITSR von CC [4], dass der UO sowohl relativ zu anderen Wettbewerbern als auch insgesamt [10] eine systematische Unterschätzung der relevanten ITSR seitens der Anbieter bedingt. Grundlage der Untersuchung des UO ist eine quantitative empirische Studie unter den CC Anbietern im deutschen Markt, wobei die Entscheidungsträger gebeten wurden die ITSR jeweils für das eigene Unternehmen sowie den durchschnittlichen Wettbewerber im Markt zu bewerten. Der Vergleich der Durchschnittswerte aus der Einschätzung des ITSR des eigenen Unternehmens und der des durchschnittlichen Wettbewerbers erlaubt dabei eine differenzierte Analyse systematischer Verzerrungen bzw. des UO [11]. Darüber hinaus weisen wir nach, dass der UO bei den einzelnen Anbietern zu einem signifikant reduzierten Bewusstsein (Awareness) bzw. Auseinandersetzung der Verantwortlichen mit den ITSR und –maßnahmen [17] im Bereich des CC führt.

Mit der Schaffung eines besseren Verständnisses der kognitiven Abläufe und insbesondere der s. g. „systematischen kognitiven Fehlleistungen“ bei der Wahrnehmung von Risiken, können die Ergebnisse den Anbietern in erster Linie eine Verbesserung ihrer Strategien zur Risikobewertung, auch bei zukünftigen Technologieinnovationen, ermöglichen. Eine korrekte Bewertung der ITSR seitens der Anbieter stellt eine Grundvoraussetzung dar, um mögliche Bedrohungen rechtzeitig zu identifizieren und Sicherheitsmaßnahmen in erforderlichem Umfang zu implementieren. Durch die Implementierung der objektiv erforderlichen Sicherheitsmaßnahmen kann sowohl der Schutz der CC Angebote gesteigert als auch der finanzielle Aufwand für die Sicherheit optimiert werden. Mit einer Steigerung der IT Sicherheit kann außerdem eine langfristige Verbesserung der Akzeptanz von CC erwartet werden.

Auch wenn insbesondere die Marketingabteilungen der CC Anbieter angesichts der zögerlichen Nachfrage unablässig den umfassenden Schutz der Dienste betonen, sollten die Nutzer den Stand der IT Sicherheit anhand unserer Ergebnisse kritisch hinterfragen und für Risiken mit typischerweise hohen Verzerrungen ggf. eigene Sicherheitsüberprüfungen vor Vertragsabschluss durchführen.

2 Grundlagen: ITSR im CC und Unrealistischer Optimismus

2.1 Wahrgenommene ITSR von CC

Auf Basis von Cunningham (1967) wird wahrgenommenes Risiko in der Literatur oftmals als *“the felt uncertainty regarding the possible negative consequences of adopting a product or service”* verstanden [18]. Die Analyse von (insb. wahrgenommenen) Risiken hat im Kontext von ITO eine lange Tradition – mit der Entwicklung

des ITO-Konzepts hat sich jedoch der Fokus der betrachteten Risikodimensionen verschoben: Während bei der Forschung zu traditionellem ITO eher strategische und finanzielle Risiken im Vordergrund standen [19-20], änderte sich der Betrachtungsgegenstand mit dem Aufkommen von Application Service Providing (ASP) und CC in Richtung eher technisch bedingter Risiken [21-22]. Aktuelle Studien zeigen, dass gerade die IT sicherheitsbezogenen Risiken den größten Einfluss auf die Adoptionsentscheidung im CC Kontext haben [21].

Wir definieren das wahrgenommene ITSR im CC Kontext als *das vom Entscheider wahrgenommene Risiko für die Sicherheit der IT des Unternehmens wenn CC als Bezugsmodell verwendet wird*. Damit bauen wir auf Forschungsergebnissen zur Konzeptualisierung der wahrgenommenen ITSR im Kontext von CC aus einer vorangegangenen Studie auf [4]. Hierbei wurde auf Basis einer umfangreichen Literaturrecherche, eines anschließenden Q-Sorts (6 IS Experten), strukturierten Interviews mit 24 IT Sicherheitsexperten und einer empirischen Untersuchung unter 356 Unternehmen eine umfassende Konzeptualisierung von wahrgenommenen ITSR für den CC Kontext entwickelt und validiert (in der Studie als PITSR - Perceived IT Security Risk) bezeichnet. Dabei konnten wir „Verfügbarkeit“, „Vertraulichkeit“, „Integrität“, „Leistung“, „Anpassbarkeit und Wartung“ und „Zurechenbarkeit“ als die sechs Risikodimensionen wahrgenommener ITSR identifizieren. Die Risikodimension Verfügbarkeit bezieht sich dabei darauf, dass der Zugriff auf das Angebot und die Daten zu jedem vom Kunden gewünschten Zeitpunkt möglich ist. Unter Vertraulichkeit verstehen wir, dass Daten ausschließlich von autorisierten Benutzern gelesen werden. Die Dimension Integrität umfasst Risiken bezüglich der Veränderung von Daten durch Unbefugte. Leistungsrisiken betreffen Vorfälle, durch die eine Nutzung des CC Angebots und der Daten nicht in der Geschwindigkeit erfolgen kann, die den Anforderungen der Kunden entspricht. Unter Anpassbarkeit und Wartung verstehen wir, dass die Anpassungen des Angebots an eigene bzw. geänderte Anforderungen möglich und Wartung sowie Support gewährleistet sind. Zurechenbarkeitsrisiken treten auf, wenn Authentifizierungsmechanismen umgangen werden können und Aktionen im Rahmen der Nutzung des Angebots nicht eindeutig identifizierbaren Benutzern zugeordnet werden können. Zu den sechs Dimensionen wurden zudem 31 IT-Sicherheitsrisiken von CC identifiziert (vgl. Tabelle 3 in Kap. 3.3).

2.2 Unrealistischer Optimismus

Menschen tendieren in vielen Fällen dazu, sich anderen überlegen zu fühlen. Obwohl wissenschaftliche Studien [23] zeigen, dass dieses Phänomen nicht zwangsläufig mit negativen Konsequenzen verbunden sein muss und für das Selbstwertgefühl bzw. das psychische Wohlbefinden einer Person sogar unabdingbar sein kann [24], verdeutlichen resultierende Verzerrungen in der Bewertung von Risiken gleichzeitig eine mögliche Gefahr. Individuen schreiben sich selbst vielfach wünschenswerte Attribute zu und interpretieren vorhandene Informationen oder unbekannte Sachverhalte in für sich positiver Weise [16]. Aufgrund dieser kognitiven Prozesse unterschätzen Personen typischerweise besonders die eigene Wahrscheinlichkeit für das Auftreten negativer Ereignisse im Vergleich zu anderen erheblich und lassen auch im Hinblick auf

bekannte Risiken nicht die notwendige Vorsicht walten [11]. Diese s. g. „systematische kognitive Fehlleistung“ wird in der Literatur als „unrealistischer Optimismus“, „optimistische Verzerrung“ oder „optimistischer Fehlschluss“ bezeichnet [10].

Frühere Studien über die Verzerrung der Risikowahrnehmung konnten nachweisen, dass Personen im Allgemeinen besonders UO bei der Einschätzung ihrer Verwundbarkeit durch verschiedene negative Ereignisse zeigen. So schätzt bspw. Menschen das eigene Risiko für das Auftreten gesundheitlicher Probleme, wie z. B. Herzinfarkt, chronische Krankheiten oder auch AIDS, signifikant geringer ein als das Risiko einer anderen Person mit gleichem Geschlechts, Alter und Bildungsstand [11]. Neben den gesundheitlichen Risiken konnte UO von Wissenschaftlern in einer Vielzahl anderer Bereichen, z. B. hinsichtlich Risiken des Autofahrens, Rauchens oder Kriminalität, gezeigt werden [12-13], [16]. Auch in unserer Disziplin konnte in Bezug auf Risiken der Internetnutzung sowie der IT Sicherheit im Allgemeinen bereits UO bei den Entscheidungsträger in Unternehmen nachgewiesen werden [25-26].

Das Vorhandensein von UO bei einem Individuum kann grundsätzlich bewertet werden, wenn die Risikoeinschätzung der Person mit dem tatsächlichen Risiko verglichen werden kann. Allerdings sind in vielen Bereichen, wie bspw. der IT, keine geeigneten quantitativen Daten verfügbar, um das tatsächliche Risiko bestimmen zu können. Die Bewertung des UO ist dann für eine einzelne Person im Allgemeinen nicht möglich, da die Einschätzung der befragten Person, dass sie einem geringeren Risiko ausgesetzt sei, durchaus korrekt sein könnte [27]. Die UO Forschung bedient sich in diesem Fall einem Vergleich der Risikobewertungen einer bestimmten Personengruppe, wobei die Teilnehmer aufgefordert werden ihr Risiko im Vergleich zu einer Referenz (direkte Methode) oder ihr eigenes Risiko und das Risiko der Referenz separat (indirekte Methode) zu bewerten [15]. Die Referenz des Vergleichs hängt dabei vom Gegenstand der Untersuchung ab und kann sowohl eine andere Person mit bestimmten Merkmalen als auch ein Wettbewerber bzw. ein anderes Unternehmen sein. Anhand eines Vergleichs der Mittelwerte der Selbsteinschätzungen und des Durchschnitts der Bewertungen der anderen Personen kann untersucht werden, ob eine Person bzw. ein Personenkreis ein Risiko systematisch verzerrt wahrnimmt [11]. Ist bspw. die Risikowahrnehmung einer repräsentativen Stichprobe von Rauchern nicht systematisch durch UO verzerrt, sollte die Differenz zwischen dem Durchschnitt der Einschätzungen des eigenen Risikos und der Bewertung des durchschnittlichen Rauchers gleich Null sein. Andernfalls würden sich alle Raucher im Durchschnitt einem anderen Risiko ausgesetzt sehen als der durchschnittliche Raucher und eine Verzerrung der Risikobewertung dieser Personengruppe wäre offenkundig.

Grundlage dieser Vorgehensweise bildet die Theorie des sozialen Vergleichs nach Festinger (1954) [9]. Personen streben zwar grundsätzlich zunächst eine Bewertung ihres Risikos mit objektiven Maßnahmen an, bedienen sich allerdings beim Fehlen entsprechender Mittel einem Vergleich mit anderen Personen, denen sie ähnliche Eigenschaften zurechnen [28]. Auch wenn der kognitive Mechanismus in erster Linie einer objektiven Einschätzung eines Risikos dienen soll, zeigen Studien in diesem Bereich, dass unterschiedliche Einflussfaktoren, wie bspw. Selbstüberschätzung oder Wünsche, den Vergleich beeinflussen können [29]. So neigen Personen bspw. unbewusst dazu, sich mit anderen zu vergleichen, von denen sie wissen, dass sie hinsicht-

lich bestimmter Eigenschaften benachteiligt sind. Auf diese Weise werden Bedrohungen negativer Ereignisse relativiert und das persönliche Wohlbefinden gesteigert [30]. Analog würde eine solche optimistische Verzerrung bspw. die Zufriedenheit des Entscheidungsträgers im CC hinsichtlich seines beruflichen Erfolges steigern und gleichzeitig die Wahrnehmung der Verwundbarkeit der Dienste durch die ITSR reduzieren [26].

Als wesentliche Einflussfaktoren auf den UO wurden von den Wissenschaftlern in früheren Studien die wahrgenommene Kontrolle und die s. g. „soziale Distanz“ zur Vergleichsperson identifiziert [31]. Die wahrgenommene Kontrolle beschreibt dabei die Erwartung einer Person, inwieweit sie den Ausgang eines bestimmten Ereignisses in ihrem Sinne beeinflussen kann. Außerdem bestimmt die Charakteristik des Vergleichsziels als soziale und psychologische Distanz den Grad des UO einer Person. Die Befragten zeigen im Allgemeinen einen höheren UO, wenn sie sich mit einer durchschnittlichen anderen Person statt einer bestimmten Person, wie bspw. einem Freund, vergleichen sollen, da ersteren eher schlechtere Eigenschaften zugerechnet werden [13].

3 Empirische Studie – Unrealistischer Optimismus im CC

3.1 Datenerhebung und Analyse

Zur Identifikation und Bewertung des unrealistischen Optimismus in der Risikowahrnehmung der CC Anbieter wurde ein Fragebogen entwickelt und in einem mehrstufigen Prozess anhand kognitiver Interviews [32] mit 4 Wissenschaftlern und 7 Experten aus der Praxis getestet, wodurch einige Formulierungen geschärft werden konnten. Die finale Version des Fragebogens enthält Indikatoren zur empirischen Bewertung der (Gesamt-) ITSR, der (Einzel-) ITSR sowie Elemente zur Messung des Bewusstseins (*Awareness*) bzgl. ITSR von CC. Der Fragebogen wurde an 247 CC Anbieter im deutschen Markt (Stand: Dezember 2011) verteilt, die im Wesentlichen anhand entsprechender Publikationen [33] sowie Datenbanken, unterstützt von einer systematischen Recherche in einem sozialen Netzwerk im Internet (Xing), ermittelt werden konnten. Sofern entsprechende Informationen verfügbar waren, haben wir grundsätzlich zunächst den CIO oder IT Sicherheitsbeauftragten des Unternehmens kontaktiert. Bei vielen kleineren CC Anbietern konnte ausschließlich der Geschäftsführer bzw. CEO ermittelt werden. Die Fragebögen wurden anschließend oftmals innerhalb des Unternehmens an den Verantwortlichen weitergeleitet, der uns im Folgenden als Ansprechpartner für die Studie diente.

Die Datenerhebung fand im Zeitraum vom 10. Juni bis 20. Juli 2012 statt. Die Ansprechpartner wurden durch das Angebot eines detaillierten Ergebnisberichts mit einer Übersicht über die Risikobewertung der Wettbewerber sowie einer Erinnerung per E-Mail zur Teilnahme motiviert. Darüber hinaus wurden nach Ablauf des halben Datenerhebungszeitraums alle bekannten Ansprechpartner telefonisch kontaktiert und an die Studie erinnert. Nach Ablauf des Zeitraums haben wir insgesamt 58 ausgefüllte Fragebögen (23,5%) erhalten, wobei 11 aufgrund schlechter Datenqualität oder feh-

lender Angaben bei einzelnen Indikatoren aussortiert werden mussten. Insbesondere in Anbetracht der Schwierigkeiten bei der Erhebung von Daten bei IT Führungskräften stellt dies ein gutes Ergebnis dar, das auch auf Interesse der CC Anbieter an der Thematik schließen lässt [34].

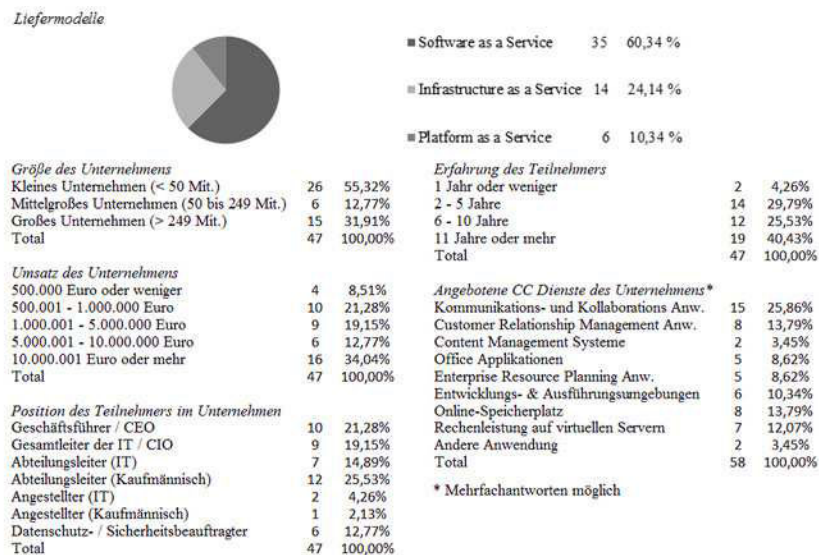


Abb. 1. Deskriptive Teilnehmerdaten der empirischen Untersuchung

Eine Analyse der Charakteristik der deskriptiven Teilnehmerdaten (s. Abbildung 1) lässt auf eine gute Repräsentativität der Stichprobe schließen [35]. Sowohl im Hinblick auf die angebotenen Dienste (60,3 % Software-as-a-Service; 25,9 % Infrastructure-a-Service; 10,3 % Platform-as-a-Service) als auch auf die Liefermodelle sowie die Unternehmensgröße bilden die Teilnehmer ungefähr den Durchschnitt des CC Markts ab [33]. In Anbetracht der Kritikalität systematischer Antwortausfälle bei der Bewertung des unrealistischen Optimismus, bspw. könnten sich die teilnehmenden Anbieter intensiver mit der IT Sicherheit im CC befassen und in Folge objektiv einem geringeren Risiko ausgesetzt sein, haben wir weitere Untersuchungen der Stichprobe durchgeführt. Nach Armstrong und Overton (1977) haben wir die frühesten 25% mit den 25% letzten (alle nach der schriftlichen und telefonischen Erinnerung) der eingegangenen Antworten verglichen [36]. Die Vorgehensweise prüft die Auswirkungen der Höhe des Interesses der Befragten auf das Antwortverhalten unter der Annahme, dass Teilnehmer mit hohem Interesse sich frühzeitig an der Studie beteiligen. Wir konnten in der Stichprobe mit t-Tests keine signifikanten Unterschiede zwischen den Antworten in den betrachteten Variablen identifizieren. Im Rahmen der Telefonanrufe bzw. der Kontaktaufnahme in sozialen Netzwerken haben wir die Ansprechpartner außerdem nach dem Grund für die ggf. fehlende Bereitschaft zur Teilnahme befragt. In den meisten dieser Fälle haben unternehmensinterne Richtlinien grundsätzlich eine

Beteiligung an Umfragen jeglicher Art verboten oder fehlende Zeit haben es den Ansprechpartner nicht erlaubt sich an der Studie zu beteiligen.

3.2 Verwendete Skalen

Zur Sicherstellung der Konstrukt- und Inhaltsvalidität des Messmodells haben wir Skalen und Elemente aus vorhergehenden wissenschaftlichen Studien mit geringfügigen Anpassungen der ursprünglichen Formulierungen adaptiert (siehe Tabelle 1).

Tabelle 1. Verwendete Skalen der empirischen Untersuchung

Konstrukt	Indikator	Quelle
Wahrgenommen (Einzel-) ITSR für die CC Dienste	Wie bewertet Ihr Unternehmen das Risiko für (potenzielle) Nutzer des Cloud Computing-Angebots (<i>Ihres Unternehmens / Ihrer Wettbewerber</i>), dass (<i>IT Sicherheitsrisiko, siehe Tabelle 3 in Kap. 3.3</i>) <ul style="list-style-type: none"> überhaupt nicht riskant - überaus riskant 	Indikatoren basieren auf Ackermann et al. (2012)
Wahrgenommenes (Gesamt-) ITSR für die CC Dienste	Zusammenfassend und unter Berücksichtigung aller Faktoren, die die Sicherheit der IT (potentieller) Nutzer betreffen, wäre es für die allgemeine IT Sicherheit der Nutzer ... das Cloud Computing Angebot (<i>unseres Unternehmens / unserer Wettbewerber</i>) zu nutzen. <ul style="list-style-type: none"> überhaupt nicht riskant – überaus riskant 	Indikatoren basieren auf Featherman und Pavlou (2003)
Bewusstsein (<i>Awareness</i>) bzgl. ITSR von CC	<ul style="list-style-type: none"> Unser Unternehmen hat einen guten Überblick über die Bedrohungen der IT Sicherheit von Cloud Computing (<i>Wissen</i>) Unser Unternehmen hat sich sehr intensiv mit dem Thema Sicherheit im Cloud Computing auseinandergesetzt (<i>Verhalten</i>) Cloud Computing setzt die Nutzer aus unserer Sicht keinen besonderen ITSR aus (<i>Einstellung</i>) 	Indikatoren basieren auf Kruger und Kearney (2006); Featherman und Pavlou (2003)

Zur empirischen Bewertung der Wahrnehmung der einzelnen ITSR und des aggregierten Gesamtrisikos für das CC Angebot wurden die Indikatoren von Ackermann et al. (2012) [4] bzw. Featherman und Pavlou (2003) [37] verwendet. Jeder dieser Indikatoren wurden jeweils für die CC Angebote des eigenen Unternehmens und des durchschnittlichen Wettbewerbs getrennt abgefragt. Zur Messung des Bewusstseins (*Awareness*) für die IT Sicherheit der CC Anbieter wurden die Indikatoren von Kruger und Kearney (2006) auf den CC Kontext angepasst [17]. Die Auswahl der Referenz bzw. des Bezugsobjekts der Risikobewertung ist insbesondere hinsichtlich der sozialen Distanz, welche die Höhe des UO beeinflusst [8], sowie der Vergleichbarkeit der Ergebnisse relevant. Als Vergleichsobjekt zum eigenen Unternehmen wurde der durchschnittliche Wettbewerber ausgewählt, definiert als Anbieter mit Diensten ähnlicher Spezifikation im selben Marktsegment. Auf diese Weise kann die soziale Distanz fixiert und die Risikobewertungen auf äquivalente Vergleichsobjekte bezogen

werden, so dass auch verschiedenartigen Angeboten in absoluter Hinsicht verglichen werden können [38]. Die Teilnehmer sollten die ITSR anhand der Auswahl Ihrer Risikowahrnehmung auf einer 7er Skala zwischen zwei vorgegebenen Adjektiven bewerten. Die Auseinandersetzung mit der IT-Sicherheit wurde mit einer 7er Likert-Skala gemessen, wobei 1 die geringste und 7 die höchste Zustimmung des Befragten zu einer Behauptung repräsentieren.

3.3 Ergebnisse der empirischen Untersuchung

Forschungsfrage 1 (UO bei CC Anbietern). Wenn die Risikowahrnehmung der CC Anbieter im deutschen Markt nicht systematisch aufgrund UO verzerrt ist, müsste mathematisch für die Differenz des Durchschnitts der Risikowahrnehmungen für das eigenen Unternehmen und der Bewertungen des durchschnittlichen Wettbewerbers $\bar{D} = 0$ gelten, wobei R_{ijk} die Bewertung des ITSR Indikators i ($i = 0$: Gesamt-ITSR) in Bezug auf die Referenz j ($j = 1$: eigenes Unternehmen; $j = 2$: durchschnittlicher Wettbewerber) durch den CC Anbieter $k \in K$ ist.

$$\bar{M}_{ij} = \sum_k \frac{R_{ijk}}{|K|} \quad (1)$$

$$\bar{D}_i = \bar{M}_{i1} \quad (2)$$

Die Anwendung eines zweiseitigen t-Tests zeigt, dass sich die Mittelwerte der Risikobewertungen zwischen den eigenen Unternehmen und den durchschnittlichen Wettbewerbern hochsignifikant ($t(46) = -8,84$; $ps < 0,001$) mit $\bar{D}_0 = -1,66$ unterscheiden. Die CC Anbieter schätzen folglich im Durchschnitt das ITSR der eigenen Dienste signifikant geringer ein als das des durchschnittlichen Anbieters, so dass auf eine systematische Verzerrung der Risikowahrnehmung der CC Anbieter im deutschen Markt aufgrund von UO geschlossen werden [11]. Die Ergebnisse zeigen darüber hinaus, dass die CC Anbieter insgesamt das Risiko für die CC Dienste des eigenen Angebots als sehr gering ($\bar{M}_{01} = 1,55$) sowie des durchschnittlichen Angebots als eher gering ($\bar{M}_{01} = 3,21$) einschätzen. Ein einzelner Anbieter k kann genau dann als unrealistisch optimistisch betrachtet werden, wenn seine Wahrnehmung des ITSRs für das eigene Unternehmen signifikant vom Durchschnitt (über alle befragten Anbieter) der Bewertungen des ITSRs des durchschnittlichen Anbieters abweicht [11].

$$UO_{ik} = R_{i1k} - \bar{M}_{i2} \quad (3)$$

35 CC Anbieter (74,5%) aus unserer Stichprobe weisen eine negative Differenz zwischen der Wahrnehmung des ITSR der eigenen Dienste und der durchschnittlichen Risikobewertung aller Anbieter auf ($UO_{0k} < 0$) und können folglich als unrealistisch Optimistisch betrachtet werden.

Forschungsfrage 2 (UO und Awareness der CC Anbieter für ITSR). Alle Anbieter im Markt haben angegeben, einen guten Überblick über die IT Sicherheitsbedrohungen von CC ($M = 6,49$) zu haben und sich sehr intensiv mit dem Thema Sicherheit im CC auseinanderzusetzen ($M = 6,27$). Eine besondere Bedrohung der CC Dienste durch ITSR sehen die Anbieter im Allgemeinen eher nicht ($M = 5,21$).

Ein Vergleich der Mittelwerte zwischen Anbietern die *UO* ($UO_{0k} < 0$) und Anbietern die keinen *UO* ($UO_{0k} \geq 0$) in der Auseinandersetzung mit ITSR gezeigt haben, verdeutlicht hochsignifikante ($D = 0,65$; $t(33) = 5,12$; $ps < 0,001$) bzw. signifikante ($D = 0,82$; $t(19) = 3,67$; $ps < 0,05$) Abweichungen (zweiseitiger t-Test), wobei positive Differenzen eine höhere Zustimmung der Anbieter ohne *UO* repräsentieren (s. Tabelle 2). Anbieter deren Risikowahrnehmung nicht durch *UO* verzerrt ist, haben sich folglich intensiver mit der IT Sicherheit im CC auseinandergesetzt und haben einen besseren Überblick. Demgegenüber sind Anbieter, bei denen kein *UO* nachgewiesen werden konnte, erheblich kritischer gegenüber der IT Sicherheit von CC ($D = -0,86$; $t(14) = -2,07$; $ps < 0,01$) eingestellt. Eine negative Differenz zeigt in diesem Fall eine geringere Zustimmung dieser Anbieter zur Aussage, dass CC im Allgemeinen mit keinen besonderen ITSR verbunden ist. Insgesamt zeigen Anbieter, denen keine Verzerrung der Risikowahrnehmung aufgrund *UO* festgestellt werden konnte, tendenziell ein höheres Bewusstsein gegenüber den ITSR im CC [17]. Die Resultate bestätigen die Befunde in anderen Forschungsgebieten im Bereich der ITSR von CC, wonach der *UO* signifikant negative Auswirkungen auf das Bewusstsein für Risiken und damit die umgesetzte, Sicherheitsmaßnahmen hat [14-15].

Tabelle 2. Mittelwerte der *Awareness* bzgl. ITSR von CC (N=47)

<i>Awareness</i> bzgl. ITSR im CC	UO ^a	M ^b	SD ^c	D ^d	SE ^e
Wissen	kein UO (≥ 0)	6,92	0,15	0,65***	0,31
	UO (< 0)	6,21	0,80		
Verhalten	kein UO (≥ 0)	6,71	0,49	0,82*	0,22
	UO (< 0)	5,90	0,78		
Einstellung	kein UO (≥ 0)	4,85	1,90	-0,86**	0,75
	UO (< 0)	5,71	1,25		

^a $UO \geq 0$: kein *UO* empirisch nachgewiesen; $UO < 0$: *UO* empirisch nachgewiesen.

^b Mittelwert (M) der Indikatoren auf 7er Likert-Skala gemessen mit 1 (stimme überhaupt nicht zu) und 7 (stimme voll und ganz zu).

^c Standardabweichung (SD)

^d Differenz (D) der Mittelwerte, wobei ein positiver Wert eine höhere Zustimmung und eine negative Differenz eine geringere Zustimmung der Anbieter repräsentiert, denen kein *UO* empirisch nachgewiesen werden konnte. Signifikanzen mit zweiseitigem t-Test: *** $ps < 0,001$; ** $ps < 0,01$; * $ps < 0,05$.

^e Standardfehler (SE)

Das Bewusstsein (*Awareness*) für ITSR hat im Allgemeinen starke Auswirkungen auf die konsequente Umsetzung von IT Sicherheitsmaßnahmen in Unternehmen [39]. Empirische Studien in diesem Bereich zeigen, dass die Unternehmen, deren Mitarbeiter über ein hohes Bewusstsein für die Relevanz von ITSR verfügen, objektiv erheb-

lich besser geschützt sind. Neben höheren Investitionen im Bereich der IT Sicherheit wurden die Maßnahmen in diesen Unternehmen eher organisationsweit umgesetzt bzw. von den Mitarbeitern stetig mitgetragen [17]. Das reduzierte Bewusstsein für ITSR hat somit auch negative Folgen für die IT Sicherheit der CC Angebote [26].

Forschungsfrage 3 (UO und Charakteristika der ITSR von CC). Die Analyse der empirischen Bewertungen der untersuchten 31 Risikoelemente von CC durch die Anbieter für das eigene Angebot und das der durchschnittlichen Wettbewerber, zeigt eine signifikante, systematische Verzerrung der Risikowahrnehmung [11]. Aufgrund der überschneidungsfreien und vollständigen Abdeckung der Gesamtheit der ITSR von CC durch die 31 Risikoelemente liefern die empirischen Befunde eine starke Bestätigung für das Vorhandensein von UO im CC (Forschungsfrage 1) [4]. Die Untersuchung erfolgt mit einem zweiseitigen t-Test anhand der Annahme, dass die Risikoeinschätzung der Anbieter in den Einzelrisiken nicht durch UO verzerrt sind und folglich die Differenzen der Mittelwerte zwischen der Bewertung des eigenen Unternehmens und des durchschnittlichen Wettbewerbers gleich Null ist ($\sum_{i=1}^{31} \bar{D}_i = 0$; vgl. Forschungsfrage 1: Formel 2) [11]. Die Annahme kann in der Stichprobe allerdings mit $\sum_{i=1}^{31} \bar{D}_i = -0,90$ verworfen werden, wobei sich alle Einzelrisiken signifikante und negative Differenzen aufweisen (vgl. Tabelle 3). Die Anpassbarkeits- und Wartungsrisiken sind dabei im Mittel durch die höchste Verzerrung ($\sum_{i=1}^7 \bar{D}_i = -1,02$) gekennzeichnet. Die übrigen Risikofacetten weisen im Durchschnitt eine ähnliche negative Differenz der Mittelwerte auf (Integritätsrisiken: $\sum_{i=8}^{12} \bar{D}_i = -0,89$; Leistungsrisiken: $\sum_{i=13}^{16} \bar{D}_i = -0,87$; Verfügbarkeitsrisiken: $\sum_{i=17}^{22} \bar{D}_i = -0,86$; Vertraulichkeitsrisiken: $\sum_{i=23}^{26} \bar{D}_i = -0,85$; Zurechenbarkeitsrisiken: $\sum_{i=27}^{31} \bar{D}_i = -0,81$).

Eine differenzierte Betrachtung der Risikobewertungen der einzelnen ITSR verdeutlicht teilweise erhebliche Unterschiede zwischen den Abweichungen der jeweiligen Mittelwerte bzw. der relativen Höhe der Unterschätzung, die bereits erste Rückschlüsse auf die Ursachen des UO ermöglichen. So unterscheidet sich bspw. das IT Sicherheitsrisiko mit der höchsten Risikoverzerrung („Unzureichende Wartung“ (#1): $\bar{D}_1 = -1,42$; $t(46) = -5,46$; $ps < 0,001$) zu dem mit der geringsten Differenz („Identitätsdiebstahl“ (#31): $\bar{D}_{31} = -0,46$; $t(46) = -3,11$; $ps < 0,01$) durch die Kontrollmöglichkeiten seitens des Anbieters. Der Anbieter kann im Allgemeinen leichter die Wartung seiner CC Dienste kontrollieren als z. B. den Diebstahl der Anmeldedaten der Nutzer, da dieses Risiko zum großen Teil auch vom Verhalten der Kunden abhängig ist. Analog hat ein Anbieter bspw. direkten Einfluss auf die Bereitstellung der CC Angebote („Einstellung des Angebots“ (#2): $\bar{D}_{17} = -1,33$; $t(46) = -5,13$; $ps < 0,01$), aber weniger Kontrolle bzgl. einem möglichen Versagen interner IT Systeme („Nichtverfügbarkeit interner Systeme“ (#30): $\bar{D}_{22} = -0,51$; $t(46) = -2,26$; $ps < 0,01$). Die Ergebnisse weisen auf eine Bestätigung der Theorie hin, wonach die wahrgenommene Kontrolle [12] über ein negatives Ereignis bzw. ITSR signifikanten Einfluss auf die Höhe des UO und somit dem Grad der Unterschätzung eines Risikos haben [10], [26]. Forschungsfrage 3 kann als empirisch bestätigt betrachtet werden, da die Einzelrisiken nicht den gleichen Grad an UO aufweisen und sich hinsichtlich ihrer Charakteristika unterscheiden [4].

Tabelle 3. UO der Anbieter in den (Einzel-) PITSR von CC (N=47)

#	IT Sicherheitsrisiko ^a	D ^b	SD ^c
1	[1] Unzureichende Wartung (Wart.)	-1,42***	1,79
2	[17] Einstellung des Angebots (Verf.)	-1,33***	1,20
3	[2] Zeitlich ungünstige Updates (Wart.)	-1,25***	1,22
4	[13] Minderleistung nach Vertragsabschluss (Leist.)	-1,17***	1,24
5	[27] Aktionen unzureichend protokolliert (Zurech.)	-1,13***	1,26
6	[8] Datenmanipulation während Übertragung (Int.)	-1,08**	1,53
7	[3] Fehlender technologischer Fortschritt (Wart.)	-1,08***	1,35
8	[18] Verlust von Zugriff auf Daten (Verf.)	-1,04***	1,12
9	[14] Unzureichende Skalierbarkeit (Leist.)	-1,04**	1,43
10	[9] Datenmanipulation beim Anbieter (Int.)	-1,00***	1,14
11	[10] Ändern von Daten auf internen Systemen (Int.)	-1,00***	1,35
12	[23] Abhören der Übertragung (Vert.)	-1,00**	1,38
13	[24] Datenweitergabe durch Anbieter (Vert.)	-1,00**	1,50
14	[19] Datenverluste beim Anbieter (Verf.)	-0,96**	1,40
15	[28] Unzureichende Trennung von Kunden (Zurech.)	-0,96**	1,33
16	[4] Inkompatible Geschäftsprozesse oder Software (Wart.)	-0,92**	1,50
17	[5] Unzureichender Datenimport (Wart.)	-0,92***	1,21
18	[25] Einsehen von Daten beim Anbieter (Vert.)	-0,88**	1,33
19	[6] Unzureichende Anpassbarkeit (Wart.)	-0,83*	1,58
20	[29] Zugriff ohne Autorisation (Zurech.)	-0,83***	1,09
21	[7] Proprietäre Technologien (Wart.)	-0,75*	1,48
22	[11] Datenveränderung während Übertragung (Int.)	-0,75**	1,03
23	[20] Ungewollte Ausfälle (Verf.)	-0,71**	1,04
24	[21] Angriffe auf Verfügbarkeit (Verf.)	-0,71**	1,04
25	[15] Geschwindigkeitsprobleme (Leist.)	-0,71**	1,20
26	[30] Nichtzurechenbarkeit bei internen Aktionen (Zurech.)	-0,67*	1,31
27	[12] Datenveränderung beim Anbieter (Int.)	-0,63**	0,97
28	[16] Geschwindigkeitsprobleme interner Systeme (Leist.)	-0,58*	1,14
29	[26] Einsehen von Daten auf internen Systemen (Vert.)	-0,54*	1,18
30	[22] Nichtverfügbarkeit interner Systeme (Verf.)	-0,51**	0,78
31	[31] Identitätsdiebstahl (Zurech.)	-0,46**	0,72

^a ITSR: [Elementnr.] Name (Risikofacette)

^b Differenz (D) der Mittelwerte für die einzelnen ITSR gemessen auf einer 7er Skala mit 1 (überhaupt nicht riskant) und 7 (überaus riskant), wobei eine signifikant negative Differenz eine systematische Unterschätzung des jeweiligen IT Sicherheitsrisikos verdeutlicht. Signifikanz mit zweiseitigem t-Test: *** $ps < 0,001$; ** $ps < 0,01$; * $ps < 0,05$.

^c Standardabweichung (SD)

4 Zusammenfassung der Ergebnisse und Ausblick

Auf Grundlage der Theorie des UO konnten wir empirisch eine systematische Verzerrung in Form einer Unterschätzung der ITSR seitens der CC Anbieter im deutschen Markt nachweisen. Darüber hinaus konnten wir zeigen, dass die resultierende Unterschätzung der ITSR signifikant negative Auswirkungen auf die Auseinandersetzung der Entscheidungsträger mit IT Sicherheit bzw. die „*Awareness*“ für diese Risiken hat. Die Betrachtung der 31 Einzelrisiken des CC ermöglichte uns in diesem Zusammenhang sowohl die Risikowahrnehmung der Entscheider vollständig hinsichtlich des Vorhandenseins von UO zu analysieren als auch erste Rückschlüsse auf relevante Einflussfaktoren in unserem Forschungsgebiet zu ziehen. In allen Facetten des ITSRs konnte dabei eine systematische Unterschätzung der Risiken durch die Verantwortlichen gezeigt werden, wobei in Abhängigkeit der Charakteristika eines ITSR, bspw. der Kontrollierbarkeit durch den Anbieter, der Grad des UO variiert.

Insbesondere im Hinblick auf die Einflussfaktoren des UO im CC ist weitere Forschung erforderlich, um die Auswirkungen der bekannten Einflussgrößen im Bereich der IT zu analysieren und ggf. spezifische Faktoren zu identifizieren [29]. Darüber hinaus sollte die Untersuchung auf internationaler Ebene fortgeführt werden, um den Einfluss kultureller Faktoren auf die subjektive Risikowahrnehmung zu untersuchen.

Während die systematische Verzerrung der Risikowahrnehmung seitens der Anbieter im deutschen CC Markt im Durchschnitt eindeutig nachgewiesen werden konnte, unterliegt die Analyse einzelner Anbieter methodischen Einschränkungen. Obgleich die Methode nach Weinstein (1980) [11] eine gute Approximation des individuellen UO liefert, ist die Quantifizierung des tatsächlichen Risikos, dem eine Person ausgesetzt ist, notwendig, um Messfehler und falsche Klassifizierungen vollständig ausschließen zu können.

Nach unserem Kenntnisstand haben wir mit der vorliegenden Studie erstmals die Theorie des UO auf ein ITO Konzept übertragen, wobei wir dieses vollständig hinsichtlich Verzerrungen in der Risikobewertung untersuchen konnten. Vor dem Hintergrund der Bedeutung der Risikowahrnehmung bzw. der „*Awareness*“ der Entscheidungsträger für die Umsetzung möglicher IT Sicherheitsmaßnahmen können die Ergebnisse wichtige Implikationen für zukünftige IT Sicherheitsforschungen und das Risikomanagement liefern. Das Verständnis der kognitiven Prozesse ermöglicht außerdem auch Anbietern und Nutzern gezielt Strategien zu entwickeln, um Verzerrungen in der Risikowahrnehmung zu adressieren und auf diese Weise langfristig sowohl die IT Sicherheit als auch den finanziellen Aufwand zu optimieren. Der Anwendungsfall des CC verdeutlicht dabei die zunehmende Bedeutung der IT Sicherheit für ITO Beziehungen [4], so dass in Zukunft vermehrt ganzheitliche Ansätze des IT Sicherheits- und Risikomanagements unter Berücksichtigung subjektiver Risikowahrnehmung sowie unbewussten „kognitiven Fehlleistungen“ aller Beteiligten erforderlich werden.

Literatur

1. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. National Institute of Standards and Technology (2011)
2. Vaquero, L.M., Rodero-Merino, L., Morán, D.: Locking the sky: a survey on IaaS cloud security. *Computing* 91 (1), 93-118 (2011)
3. Pring, B.: Cloud Computing: The Next Generation of Outsourcing. Gartner Group (2010)
4. Ackermann, T., Widjaja, T., Benlian, A., Buxmann, P.: Perceived IT Security Risks of Cloud Computing: Conceptualization and Scale Development. In: Proceedings of the 33rd International Conference on Information Systems, Orlando (2012)
5. Hange, M.: Security Recommendations for Cloud Computing Providers. Federal Office for Information Security (2011)
6. Gigerenzer, G.: Dread Risk, September 11, and Fatal Traffic Accidents. *Psychological Science* 15 (4), 286-287 (2004)
7. Kankanhalli, A., Teo, H.-H., Tan, B.C.Y., Wei, K.-K.: An integrative study of information systems security effectiveness. *Int. Journal of Information Management* 23, 139-154 (2003)
8. Klein, W.M.P.: Self-Prescriptive, Perceived, and Actual Attention to Comparative Risk Information. *Psychology & Health* 18, 625-643 (2003)
9. Festinger, L.: A Theory of Social Comparison Processes. *Human Relations* 7, 117-140 (1954)
10. Weinstein, N.D., Klein, W.M.: Unrealistic optimism: Present and future. *Journal of Social and Clinical Psychology* 15, 1-8 (1996)
11. Weinstein, N.D.: Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39, 806-820 (1980)
12. McKenna, F.P.: It won't happen to me: Unrealistic optimism or illusion of control?. *British Journal of Psychology* 84, 39-50 (1993)
13. Perloff, L.S., Fetzer, B.K.: Self-Other Judgments and Perceived Vulnerability to Victimization. *Journal of Personality and Social Psychology* 50, 502-510 (1986)
14. Adams, J.: Cars, Cholera, and Cows: The Management of Risk and Uncertainty. *Policy Analysis* 335 (1999)
15. Helweg-Larsen, M., Shepperd, J.A.: Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature. *Personality and Social Psychology Review* 5, 74-95 (2001)
16. McKenna, F.P., Warburton, D.M., Winwood, M.: Exploring the limits of optimism: The case of smokers' decision making. *British Journal of Psychology* 84, 389-394 (1993)
17. Kruger, H.A., Kearney, W.D.: A prototype for assessing information security awareness. *Computers & Security* 25 (4), 289-296 (2006)
18. Cunningham, S.M.: The major dimensions of perceived risk. In: Cox, D.F.: Risk taking and information handling in consumer behavior. Harvard University Press (1967)
19. Quinn, J., Hilmer, F.: Strategic outsourcing. *Sloan Management Review* 35, 43-55 (1994)
20. Earl, M.J.: The risks of outsourcing IT. *Sloan Management Review* 37, 26-32 (1996)
21. Benlian, A., Hess, T.: Opportunities and risks of software-as-a-service: Findings from a survey of IT executives. *Decision Support Systems* 52 (1), 232-246 (2011)
22. Ackermann, T., Miede, A., Buxmann, P., Steinmetz, R.: Taxonomy of Technological IT Outsourcing Risks: Support for Risk Identification and Quantification. In: Proceedings of the 19th European Conference on Information Systems, Paper 240. AIS (2011)
23. Brown, J.D., Collins, R.L., Schmidt, G.W.: Self-Esteem and Direct versus Indirect Forms of Self-Enhancement. *Journal of Personality and Social Psychology* 55, 445-453 (1988)

24. Taylor, S.E., Brown, J.D.: Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin* 103, 193-210 (1988)
25. Campbell, J., Greenauer, N., Macaluso, K., End, C.: Unrealistic optimism in internet events. *Computers in Human Behavior* 23, 1273-1284 (2007)
26. Rhee, H.-S., Ryu, Y.U., Kim, C.-T.: Unrealistic optimism on information security management. *Computers & Security* 31, 221-232 (2012)
27. Rothman, A.J., Klein, W.M., Weinstein, N.D.: Absolute and Relative Biases in Estimations of Personal Risk. *Journal of Applied Social Psychology* 26, 1213-1236 (1996)
28. Wood, J.V.: Theory and Research Concerning Social Comparisons of Personal Attributes. *Psychological Bulletin* 106, 231-248 (1989)
29. Shepperd, J.A., Carroll, P., Grace, J., Terry, M.: Exploring the Causes of Comparative Optimism. *Psychologica Belgica* 42, 65-98 (2002)
30. Wills, T.A.: Downward comparison principles in social psychology. *Psychological Bulletin* 90, 245-271 (1981)
31. Klein, W.M., Weinstein, N.D.: Social comparison and unrealistic optimism about personal risk. In: Buunk, B.P., Gibbons, F.X. (eds.): *Health, coping, and well-being: Perspectives from social comparison theory*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ (1997)
32. Bolton, R.N.: Pretesting Questionnaires: Content Analyses of Respondents' Concurrent Verbal Protocols. *Marketing Science* 12, 280-303 (1993)
33. Velten, C., Janata, S.: *Cloud Vendor Benchmark 2011*. Experton Group (2011)
34. Poppo, L., Zenger, T.: Do formal contracts and relational governance function as substitutes or complements?. *Strategic Management Journal* 23, 707-725 (2002)
35. Heberlein, T., Baumgartner, R.: Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review* 43, 447-462 (1978)
36. Armstrong, J.S., Overton, T.S.: Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research* 14, 396-402 (1977)
37. Featherman, M.S., Pavlou, P.A.: Predicting e-services adoption: a perceived risk facets perspective. *Int. J. Hum.-Comput. Stud.* 59, 451-474 (2003)
38. Karakayali, N.: Social Distance and Affective Orientations. *Sociological Forum* 24, 538-562 (2009)
39. Siponen, M.T.: A conceptual foundation for organizational information security awareness. *Information Management & Computer Security* 8, 31-41 (2000)

Managing the Access Grid - A Process View to Minimize Insider Misuse Risks

Stefan Meier¹, Ludwig Fuchs², and Günther Pernul¹

¹University of Regensburg, Department of Information Systems, Regensburg, Germany
{stefan.meier, guenther.pernul}@ur.de

²Nexis GmbH, Regensburg, Germany
Ludwig.Fuchs@nexis-secure.de

Abstract. It is generally agreed upon the fact that the quality of Identity- and Access Management (IAM) data such as user accounts, access privileges or consistent user representation among different security domains is low. Growing user populations in medium- and large-sized organizations lead to a so called “identity chaos” in which over-privileged employees increase the risk of insider misuse. Recent governance and compliance mandates have amplified the importance of minimizing these risks. In order to fulfill these requirements, organizations focus on implementing role-based user management. To set up a role-based access control system, they face the challenge of modeling suitable roles for their employees. In this paper we show how the role modeling process can be improved by utilizing the so called access grid, a visualization technique to incorporate human interaction into the process of role creation.

Keywords: Insider Misuse, Visual Role Mining, Role Mining, Role Development, Identity Management

1 Introduction

Effectively administrating employees’ access to sensitive applications and data is one of the biggest security challenges for today’s organizations. A typical medium-sized or large enterprise manages millions of user access privileges¹ that are spread across thousands of IT resources. Job and position changes of employees further complicate the task of correctly managing users and their access rights. Major security problems arise because of manual management of user accounts spread across various applications. As a result, employees accumulate excessive entitlements over time. This accumulation violates the principle of the Least Privilege and increases the risk of insider threats. A study investigating industry espionage across several organizations recently underlined the importance of insider threats, revealing that confidential information is mostly stolen by internal employees and not by external attackers [1]. Addi-

¹ In the remainder, the terms permissions, entitlements and access privileges are used interchangeably

tionally, compliance considerations emphasized the need for centralized, automated and secure management of employees and their access privileges. Cleven and Winter [2] state that regulations like the Sarbanes-Oxley Act of 2002 (SOX) [3] or Basel III [4] lead to an increased importance of secure IAM. Organizations need to be able to verify and prove evidence that user's access privileges conform to regulatory and corporate guidelines. In order to regain the control of the managed digital identities and their entitlements, companies aim at migrating from identity-based towards role-based user management in which identities are no longer directly related to access privileges but via roles acting as intermediary between identities and entitlements. The so called Role-based Access Control (RBAC) paradigm simplifies user management, reduces administrative costs and increases the overall security level [5-6].

In order to gain the benefits of RBAC, companies are forced to initially model a set of valid business roles for their employees. Role modeling is a task commonly executed by IT experts in cooperation with business representatives like departmental managers. The human interpretation of potential role candidates therefore represents a critical process step during role modeling. Thus, supporting the human interactor during his tasks is mandatory for successful role definition. Recently, the Role-Mining Process Model (RMPM) has been proposed, integrating the necessary steps for role creation into a process-oriented framework [7]. To increase the applicability of the RMPM and to overcome drawbacks of existing role modeling approaches, we propose the usage of the so called access grid. We show how the access grid can be facilitated in order to reduce potential insider misuse risks and define high-quality business roles in a process-oriented fashion.

This paper is structured as follows: Section 2 presents related work while Section 3 investigates visualization techniques in order to choose the grid-based visualization as the basis for the access grid. In Section 4 the integration of the access grid into the RMPM is shown. Our approach is then evaluated in Section 5 using a real-world case study. Finally, Section 6 concludes our findings and reveals future research directions.

2 Related Work

The different aspects of insider threats have been discussed by various authors, e.g. Probst et al. [8], analyzing the detection and mitigation of insider threats. Closely related, aspects of compliance in the information systems discipline have been discussed by Cleven and Winter in [2]. The authors give an overview of compliance needs and regulations and underline how the growing number of regulations and laws affects businesses. Furthermore, the need for additional research, especially on the development of new concepts and solutions to achieve compliance, is pointed out.

Implementing role-based user management conforming to the RBAC standard is commonly seen as a main element of compliant IAM and thus for minimizing insider threats. To complete the initial task of role definition, three main approaches have been proposed: Role engineering, role mining and a hybrid combination of both techniques [9]. Role engineering can be considered as the mainly manual approach, modeling roles top-down on the basis of employees' job descriptions, business processes

and organizational structuring. Neumann and Strembeck, e.g., focus on scenario-driven role engineering [10]. Role mining, on the contrary, represents the tool-based approach to identify clusters of similar users bottom-up on the basis of the existing access privilege assignments using data mining techniques. Lately, researchers have agreed upon the fact that only a hybrid combination of those two approaches leads to an applicable role catalogue in the context of enterprise-wide IAM [9].

The RMPM offers an incremental and iterative approach for hybrid role definition using role engineering as well as role mining techniques (see Figure 1). It splits the role mining tasks pre-processing (data cleansing and data selection), role detection and post-processing (integration of business semantics and hierarchy optimization) activities [7]. Role engineering results like e.g. existing work profiles or job descriptions of employees can act as input data for role mining while at the same time role mining output feeds into a new role engineering iteration.

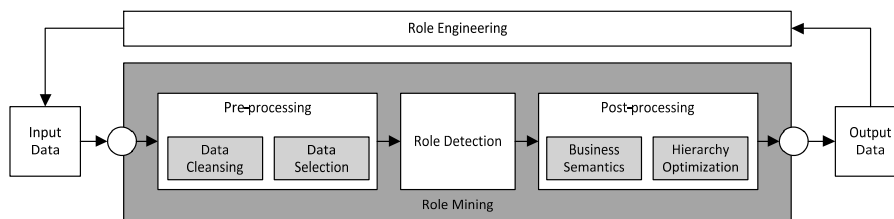


Fig. 1. The Role-Mining Process Model [7]

Several researchers already underlined the importance of data visualization during role mining. In [11], e.g., a two-dimensional matrix is facilitated to picture potential roles generated by a role mining algorithm. In other work, Zhang et al. suggest the use of a two-dimensional matrix for displaying similarities between users based on their permissions and indicating the quality of a current RBAC state [12]. Further visualizations were proposed in [9] and [12].

However, looking at the different role mining approaches underlines, that none of the current solutions provides a consistent visual component supporting the human interactor, commonly an IT-related role developer, during the process of role modeling. Existing solutions mainly concentrate on one single task or parts of the role mining activities respectively. The RMPM as the first comprehensive role modeling process integrates the single activities into a structured approach, but does not focus on the presentation and interpretation of partial results by a human. In this paper we thus focus on the role developer's tasks during the various RMPM activities. The overall goal is supporting his decision making process.

3 Visualizing Identity and Access Data

The previous section revealed the importance of information visualization in the context of role modeling and insider threat detection. In order to support a human interactor during his tasks, visualizations have to fulfill certain requirements. Their

main goal is the appropriate picturing of the aggregated or non-aggregated information depending on the current activity. In the following, we present different visualization techniques and evaluate them shortly using a set of core requirements defined below. Note that the given list is not exhaustive but rather represents the most important requirements depending on literature research and practical experience.

- Scalability: Visualizations used during role modeling have to be capable of representing large datasets containing of potentially thousands of users, permissions and permission assignments.
- Perceptibility: Visualized identity-related data needs to be interpreted intuitively by a role developer. He needs, in particular, to be able to identify patterns like role candidates as well as outliers or potential errors within the given data.
- Aggregation Level: Identity-related data needs to be visualized on various aggregation levels. Aggregated information gives an overview of the underlying data structures while at the same time the role developer might need to drill down in order to gather detailed information about the displayed employee- and permission data.
- Consistency: A suitable visualization technique should be applicable during all RMPM activities. This leads to an increased user experience and simplifies the iterative comparison and communication of results.

3.1 Visualization Techniques

In the following we focus on the evaluation of common visualization techniques in order to rate their applicability during the RMPM. A comprehensive overview of basic and advanced visualization techniques is given in [13]. Due to space constraints we excluded generally unsuited visualization techniques like glyphs and icons [14], pixel techniques or bar charts and histograms in the following [13]. They lack scalability and are not capable of intuitively displaying relationships between entities like employees and their permissions. Due to the resulting low perceptibility and scalability they are not applicable consistently within the RMPM.

Table Visualizations. Table visualizations (Figure 2a) are used to display two entities and the relations between them in table cells [13]. They are capable of representing large amounts of data, increasing their scalability, mainly through the flat representation of the data. However, table visualizations are not able to provide an easily understandable and aggregated overview of the data. As they display the relationship between employees and permissions using binary values (either a 0 or 1), they struggle with a low perceptibility. A table visualization of a large department containing more than one hundred employees and several hundred different permissions represented by binary values, for instance, can hardly be interpreted by a human.

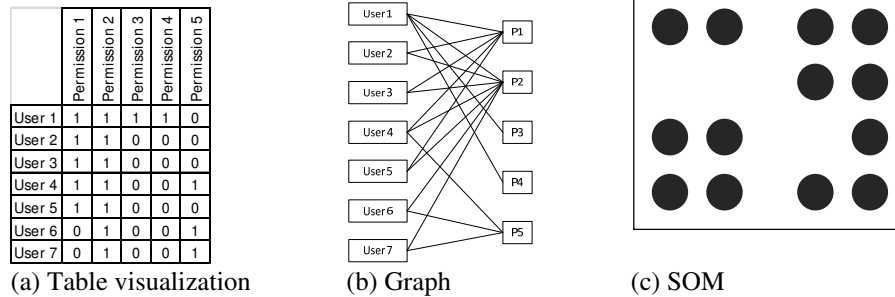


Fig. 2. Different visualization types

Graph-based Visualizations. The usage of graph-based visualization techniques (Figure 2b) consisting of nodes and edges for presenting role structures has already been proposed by several authors (e.g. in [15]). Graphs can be used to browse data on different aggregation levels and provide cluster visualizations in order to highlight outliers or potential roles, resulting in a high perceptibility. Nevertheless, graph-based visualizations have major drawbacks in scalability. Kreuzeler et al. state that they only scale until approximately 100 nodes [16], while access data of large organizations commonly exceeds this level significantly. Thus, even though graph-based visualizations are capable of displaying aggregated role information, they struggle with displaying large amounts of non-aggregated access data.

Neural Networks. SOMs (Figure 2c) have already been applied for outlier detection during role mining and data quality management [17]. Due to their high scalability and perceptibility, they are able to aggregate large amounts of data on the basis of a two-dimensional visualization. However, SOMs only allow for an aggregated data representation and can only be applied for displaying large amounts of data. If the amount of data is too low, the cluster quality significantly is reduced as SOMs position elements (like employees and their entitlements) in relation to other, potentially un-similar elements. If the basic population of displayed elements is too low or their similarity is too high, SOMs, for instance, display very similar elements within different parts of a network even though they should be located right next to each other. Practical experience shows that SOMs struggle when displaying access data consisting of less than approximately 100 employees.

Grid-based Visualizations. Similar to basic table visualizations, grid-based visualizations (Figure 3) are building on a two-dimensional matrix. Commonly, one axis is used for displaying employees while the other highlights their permissions. Figure 3 shows an example of an unsorted grid. Each grey colored cell represents a permission assignment while each empty cell shows that there is no such assignment. Previous work like [11-12] builds on a two-dimensional matrix in order to support the task of creating single roles. It, however, does not consider nor evaluate the usage of the grid consistently during pre- and post-processing steps of role modeling projects.

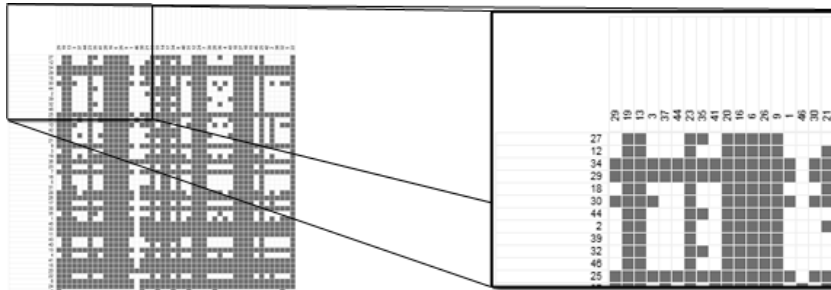


Fig. 3. Example of an access grid

In comparison to table visualizations, grid-based visualizations provide an easily interpretable overview of aggregated as well as non-aggregated data. By using colored symbols instead of characters, a role developer gets an improved overview of the underlying data structures, leading to a high perceptibility. Due to the two-dimensional representation of employees and permissions, grid-based visualizations offer a high level of consistency. In contrast to all aforementioned techniques, they can be applied during the pre-processing, role detection as well as post-processing activities of the RMPM. This high level of consistency gives the human interactor the option to work with the same visualization technique during all of his tasks. Due to these reasons we propose the usage of an enhanced grid-based visualization during the RMPM.

3.2 Characteristics of the Access Grid

In the following we shortly present the core characteristics of our proposed grid-based visualization (from hereinafter called access grid) before we embed it into the single RMPM activities in Section 4. In general, the access grid enhances the basic grid visualization technique by integrating intelligent **data sorting mechanisms** as well as a **high level of interactivity** and the inclusion of **business semantics**. It is thus no longer used as a trivial representation of two-dimensional data but rather acts as core element for information interpretation and human decision making during the RMPM.

Business Semantics. The semantic enrichment of data is one critical success factor during the application of the RMPM [7]. In contrast to two-dimensional matrix representations of access data, the access grid thus provides element coloring in order to display business-related information like employees' job title or department. The human interactor can, for instance, highlight departmental assignments of employees or visualize the homogeneity of employees' clustered job positions. This information supports him during the identification of over-authorized employees or the selection of one or more business roles from a large list of role candidates generated by role mining.

Intelligent Grid Sorting. As mentioned in [11] and [18], sorting of rows and columns is also critical when applying grid-based visualization techniques. Using a suitable sorting mechanism, interesting employee clusters and outliers might be visualized appropriately while the same patterns cannot be identified using another sorting mechanism. Therefore, the sorting of rows and columns is an optimization problem which cannot be solved in the context of identity management [11]. In general, there is a need for improved and sophisticated sorting methods which are able to focus on different aspects of used input data and reveal patterns and outliers depending on the current RMPM activity. Existing sorting mechanisms only facilitate simple similarity-based or graph-based algorithms based on pre-calculated roles.

Interactivity. Another characteristic and major improvement of the access grid is its interactivity. By not just statically visualizing data but rather allowing for human interaction and dynamic adaption, it increases applicability and simplifies the RMPM activities. The functionalities include common and practical features like drill-down and drill-up, panning and selecting. With zooming, it is possible to focus on details of the access grid or get an overview of the whole data. This is required as access grids commonly might spread over 100+ employees and 100+ permissions, leading to 10000+ cell matrices.

One main improvement of the access grid in comparison to existing visualization techniques is the deep integration into the RMPM by offering dynamic context menus. This, for instance, allows the human interactor to select certain areas of the grid and define underlying permission assignments e.g. as over-authorizations which need to be reviewed by a responsible authority like the head of the respective department. The grid offers the functionality required for delegating and forwarding the approval request to the appropriate business expert. The same holds for the role modeling task. Role candidates identified by a clustering algorithm can easily be highlighted, altered and saved for further approval.

4 Embedding the Access Grid into the Role Mining Process

In this section of the paper we present the integration of the access grid into the RMPM following the concept of visual data mining [19]. The first goal is the improvement of the interpretability of automatically generated results by a human role developer. The second goal is the process-oriented manual generation of new knowledge by the role developer (e.g. the manual detection of outliers and roles) on the basis of an intuitive visualization of large amounts of data.

4.1 Pre-processing with the Access Grid

The first activity according to the RMPM is the pre-processing of input data. Companies starting a role modeling project commonly need to initially investigate their access privilege structures for errors like over-privileged employees or unsuited input data [7]. The access grid supports both, data cleansing as well as data selection.

By visualizing grouped access privileges, it allows for manually detecting outliers in the given data structures. The human interactor can drill down into departments of a company and analyze the respective access grids for conspicuous privilege assignments. Furthermore, automatically identified potential data errors can be displayed and highlighted. This way, the outcome of automated outlier detection mechanisms as well as role engineering input (like the enforcement of security policies defining rules for privilege assignments) can be validated. In addition, the exclusion of unsuited input data from further activities is supported by the access grid. In contrast to basic data filtering, e.g. using database queries, the displaying of business semantics (e.g. department assignments, permission criticality levels, etc.) allows for a simplified data selection. The role developer can, e.g., exclude unsuited data like critical entitlements that shall not be modeled into roles. The simplified example in Figure 4 underlines the suitability of the access grid for data cleansing and data selection activities.

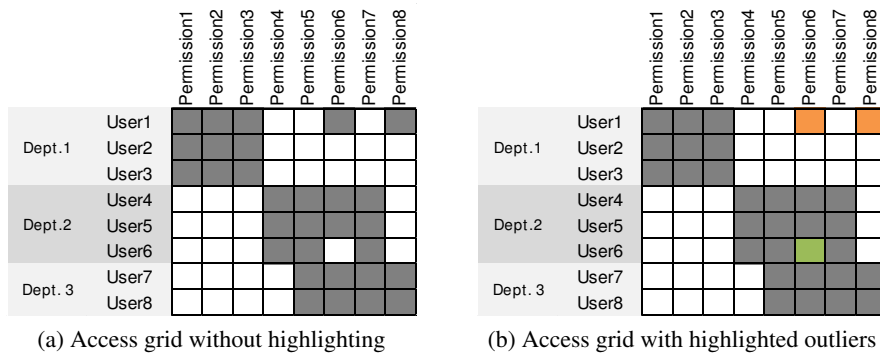


Fig. 4. Using the access grid during pre-processing

In both sub-figures similar users suitable for role modeling can be detected (e.g. User 7 and User 8 sharing identical permissions) without additional business-related knowledge. It is also possible to identify users with untypical access privileges. Inspecting Figure 4a, the role developer could highlight the potentially excessive Permissions 6 and 8 of User1 and the missing Permission 6 of User 6. In the example in Figure 4b, automatic outlier detection mechanisms already identified and highlighted those privilege assignments as suspicious. The role developer then is able to decide whether to revoke or respectively grant the access privileges or mark them as valid exceptions.

By displaying further attributes like the employees' departmental assignment or job title, cleansing and data selecting might benefit even when there is no feedback from business representatives available. This is a common challenge at the beginning of role modeling projects when companies start with a small team of IT-related role developers. The access grid allows for an initial interpretation of the access privilege quality. In case the structures reveal large areas of similar users, role modeling complexity is lower and the data quality is higher than in heterogeneous environments.

4.2 Role Detection Using the Access Grid

According to the RMPM, after pre-processing the actual role detection takes place. The access grid supports this task by allowing for interactive manual role detection on the one hand, as well as the validation of automatically generated or pre-existing roles on the other hand. This might even include pre-defined employee groups provided by business experts during role engineering activities. They can be displayed, essentially supporting the hybrid role development process.

The access grid furthermore supports various data sorting mechanisms. In case no roles have been pre-generated, it facilitates a similarity-based user sorting algorithm starting with the detection of user groups depending on the assigned access privileges. Subsequently, similar user groups and group members are clustered². Permissions assigned to similar users thereby form a role candidate which can be easily detected by concentrating on columns in the access grid where all cells are filled. Another option is to sort users by their job description and permissions by their occurrence. This might reveal entitlements related to a single job or tasks in the organization and therefore lead to the identification of according roles.

While the above description uses the humans' ability to detect patterns in the access grid, another option is to iteratively run automated role mining algorithms and display the results after each run as shown in Figure 5. This might be useful in case a company wants to build their efforts on several different role detection algorithms in parallel, each with different parameters. A company might for instance use two algorithms, one detecting only flat roles (Figure 5a) and another one to detect role hierarchies (Figure 5b). In Figure 5b, Role 4 is a parent role of Role 2 and Role 3 using a top-down inheritance relationship. Displaying the respective results using the access grid, the identification and selection of the best fitting role set depending on the companies' goals can be simplified by a visual comparison. Moreover, the access grid might highlight roles covered by all algorithms as potentially good roles.

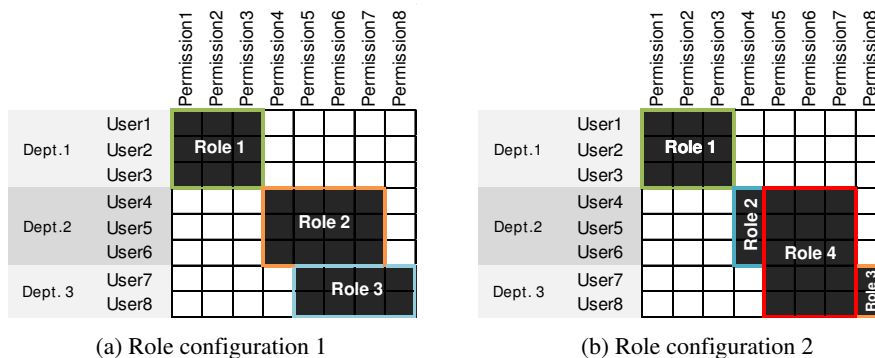


Fig. 5. Using the access grid during role detection

² Note that it is not the goal of this paper to present sorting mechanisms for the access grid. We have developed two sorting mechanisms; however, due to space constraints it is not possible to present them in detail. We plan to publish them in future work.

4.3 Post-processing with the Access Grid

During post-processing, the output data of role mining activities needs to be refined. This includes the optimization of role hierarchies, the splitting or merging of roles, as well as the further integration of business semantics to improve roles.

Hierarchy optimization can be supported by displaying overlapping roles in the access grid. The human interactor can select the overlapping area and model a role hierarchy. This refinement process can also be partly automated using a hierarchy modeling algorithm for detecting roles to be refined. This reduces the efforts of manual review by allowing for a visual inspection during role optimization. Without using the access grid, it is hardly possible to model and review role hierarchies within a large project setting consisting of thousands of employees and permissions.

The identification of roles spreading over various departments within a company is one challenge existing role modeling approaches struggle with. The access grid supports the integration of business semantics in order to support this task of role improvement. Even though the role developer analyzes one specific department, the access grid can, for instance, inform him about similar roles existing within other departments. The role developer can then refine those roles and their scope. He might aim at reducing administrative efforts and thus merge similar roles from various departments into one global role.

One example for role improvement is shown in Figure 6a where a role developer requests the merging of two existing roles (Role 2 and Role 3) into a new, larger role Role 4. In case the newly required access privileges (previously white areas in the access grid now covered by Role 4) are approved by a responsible authority, the merging can be executed. Additionally, business-related feedback resulting from a RMPM iteration can be used to improve role quality. In case a departmental manager alters a previously created role during his approval activity, he might request a role merging and a reduction of the permissions included in the role. In Figure 6b, for example, the role developer receives such a request, investigates the reasonableness of the decision and executes the merging of Role 2 and Role 3 into Role 4, essentially leading to an exclusion of certain permissions from the role definition (Permission 4 and Permission 8)

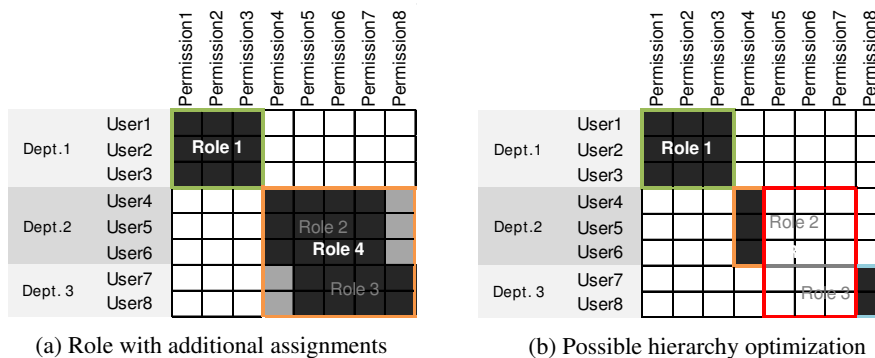


Fig. 6. Using the access grid during post-processing

5 Managing Insider Misuse Risks: A Case Study

After the application of the access grid during the RMPM has been presented, the paper continues with an evaluation of the proposed approach in a real-world application scenario, using a large, complex, and potentially erroneous dataset. We implemented the proposed access grid functionality as a web-based java application which is part of a larger tool³ for analyzing identity data and business roles. The used dataset for this evaluation has been anonymized and originates from the IAM repository (Microsoft Active Directory) of a medium-sized organization in the retail sector. The company operates in large parts of Europe and has more than 3000 employees and nearly 1000 managed access privileges (Active Directory security groups). For readability reasons we present the results of the “Service” department consisting of 15 employees, 51 different access privileges and three sub-departments. Figure 7 displays the department’s initial sorted access configuration showing the employees colored according to their departmental assignment on the left side. The anonymized permissions are represented by the columns. Grey Coloring of a grid element represents an existing access privilege of a certain employee.

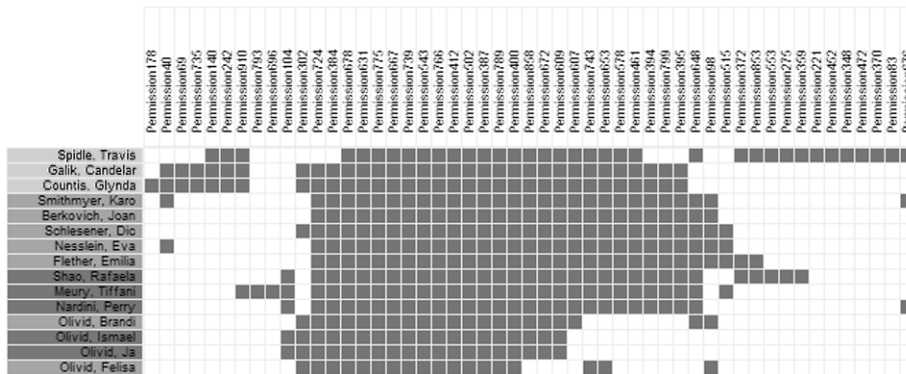


Fig. 7. Initial access grid of the “Service” department

Pre-processing. At first we applied the access grid during the pre-processing of the initial input data. During these activities, the access grid supports the role developer with its automated access privilege sorting and the inclusion of semantic information. We executed several automated data analysis algorithms to detect potentially erroneous privilege assignments and to subsequently highlight those outliers. Additionally, several over-authorizations in the predominantly white areas or on the edges of the bigger tiles in the access grid have been manually marked as outliers (Figure 8a, dark grey coloring). Subsequently, the responsible manager of the “Service” department has been asked to review the outliers in a hybrid fashion, resulting in an improved (and re-sorted) access grid (Figure 8b). Note that, depending on the company’s preferences, the role developer himself might also be accredited to approve the outliers.

³ <http://www.nexis-secure.de>

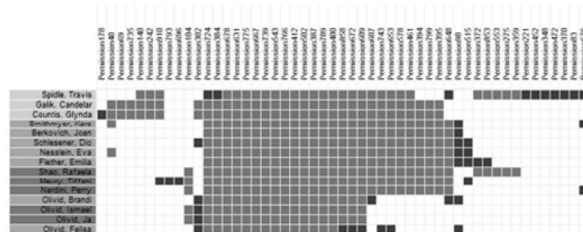


Fig. 8 (a). Data cleansing using the access grid - highlighted outliers

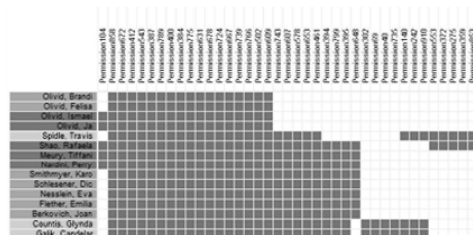


Fig. 8 (b). Data cleansing using the access grid - cleansed access data

Role Detection. After the input data has been cleansed, we used automated role mining algorithms to detect potential roles. Figure 9a displays the resulting three roles (grey, light grey and dark grey rectangles) representing the entitlements all employees share in the respective sub-departments. The role developer at this point is able to alter the role candidates by adding or removing access privileges or employees, essentially increasing or reducing the size of the rectangle representing the roles. He furthermore is capable of discarding identified roles or manually adding new roles. Figure 9b, for instance, shows the re-sorted access grid including one additional manually created role (dark grey rectangle). After the four defined roles have been approved by the responsible authority, the role developer might conclude role modeling for the “Service” department or add additional roles for the remaining uncovered access privileges (Figure 9b, grey coloring, at the right side). Practical experience revealed that using different sorting algorithms leads to altered role candidates. In future work we are thus going to investigate the effects of grid sorting on the manual role modeling in detail.

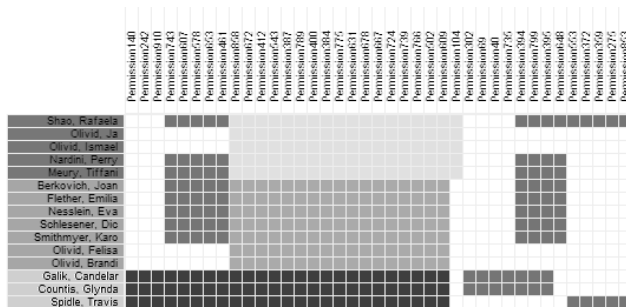


Fig. 9 (a). Role detection and creation using the access grid - automatically detected roles

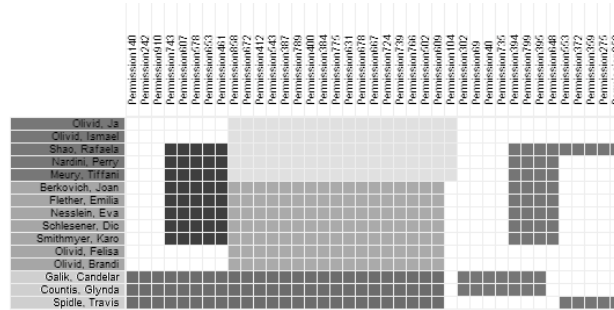


Fig. 9 (b). Role detection and creation using the access grid - manual role creations

Post-processing. Finally, the approved roles were post-processed and refined by setting up a role hierarchy. As mentioned, the three roles previously identified by the role mining algorithm represent the entitlements shared by all employees in the sub-departments of the “Service” department. The grid reveals a large overlap regarding their access privileges (see Figure 10a, light grey coloring). The role developer is thus capable of selecting the respective area and creating a new parent role for the three child roles, essentially setting up a role hierarchy.

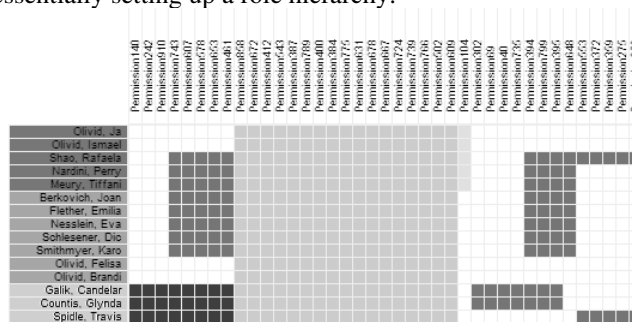


Fig. 10 (a). Role improvement using the access grid - hierarchy refinement

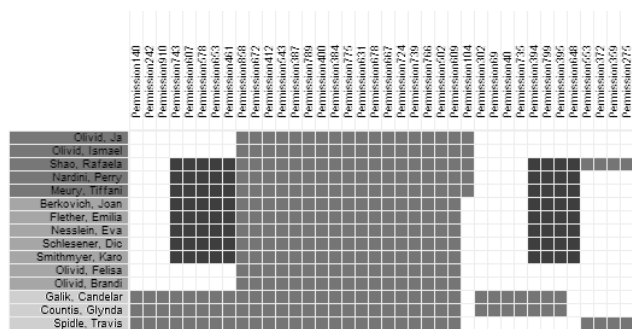


Fig. 10 (b). Role improvement using the access grid - single role refinement

Besides the role hierarchy modeling, other previously defined roles can further be improved. During post-processing, the access grid for instance highlights the potential extension of the manually defined fourth role (Figure 10b, dark grey coloring at the right side). In this case, the role developer previously missed to include those permissions within the role. During post-processing, he can facilitate the role extension suggested by the access grid and refine the role.

6 Conclusions

This paper proposed the usage of the so called access grid as an interactive visualization technique integrated into the Role-Mining Process Model for reducing insider misuse and supporting role modeling. The comparison of visualization techniques applicable during data cleansing and role mining activities in general in Section 3 revealed the access grid as the superior approach. Its high degree of usability in different project scenarios ranging from medium-sized to large-scale projects outpaces traditional limited visualization techniques.

During RMPM activities, the access grid provides intuitive decision support for responsible human interactors. Firstly, it has been shown that supporting the cleansing of excessive access privileges with the access grid leads to an improved user management within organizations. Potential insider misuse can be prevented by ensuring compliance with the principle of the Least Privilege. Secondly, the business semantic inclusion and interactivity of the access grid have been revealed as means to improve role detection and role refinement.

We have subsequently evaluated the benefits of the access grid throughout a real-world industry project. This evaluation, amongst others, revealed the impact and practical usability of the access grid and pointed at several potential future extensions. Above all, applying optimized grid sorting techniques as well as examining their effect on the access grid needs to be investigated during future research efforts as they have significant impact on data cleansing and role modeling. To further improve the human perceptibility, colorizing needs to be enhanced in order to allow for flexible and dynamic colorizing of different attributes. For applying the access grid in large environments with 10,000+ employees and permissions, approaches for data partitioning, e.g. based on departments, job descriptions or functional units in the organization additionally need to be investigated in detail, potentially including evaluation results from various case studies.

Acknowledgement

The research leading to these results was supported by “Regionale Wettbewerbsfähigkeit und Beschäftigung”, Bayern, 2007-2013 (EFRE) as part of the SECBIT project (<http://www.secbit.de>).

References

1. Corporate Trust - Business Risk & Crisis Management GmbH: Studie: Industriespionage 2012
2. Cleven, A., Winter, R.: Regulatory Compliance in Information Systems Research - Literature Analysis and Research Agenda. In: Enterprise, Business-Process and Information Systems Modeling. LNBIIP, Vol. 29, pp. 174–186. Springer, Berlin Heidelberg (2009)
3. Sarbanes-Oxley Act of 2002, <http://www.gpo.gov/fdsys/pkg/BILLS-107hr3763enr/pdf/BILLS-107hr3763enr.pdf>
4. Basel III: A global regulatory framework for more resilient banks and banking systems, <http://www.bis.org/publ/bcbs189.pdf>.
5. Sandhu, R., Ferraiolo, D., Kuhn, R.: The NIST model for role-based access control. In: Proceedings of the fifth ACM workshop on Role-based access control - RBAC'00, pp. 47–63. ACM Press, New York (2000)
6. Gallaher, M.P., O'Connor, A.C., Kropp, B.: The Economic Impact of Role-Based Access Control. RTI (2002)
7. Fuchs, L., Meier, S.: The Role Mining Process Model - Underlining the Need for a Comprehensive Research Perspective. In: ARES 2011, pp. 35–42. IEEE (2011)
8. Probst, C.W., Hunker, J., Gollmann, D., Bishop, M.: Aspects of Insider Threats. In: Insider Threats in Cyber Security. Springer (2010)
9. Fuchs, L., Pernul, G.: HyDRo – Hybrid Development of Roles. In: Information Systems Security. LNCS, Vol. 5352, pp. 287–302. Springer (2008)
10. Neumann, G., Strembeck, M.: A scenario-driven role engineering process for functional RBAC roles. In: Proceedings of the seventh ACM symposium on Access control models and technologies - SACMAT'02, pp. 33–42 (2002)
11. Colantonio, A., Di Pietro, R., Ocello, A.: Role Mining in Business: Taming Role-Based Access Control Administration. World Scientific Publishing Co. Pte. Ltd. (2012)
12. Zhang, D., Ramamohanarao, K., Versteeg, S., Zhang, R.: RoleVAT: Visual Assessment of Practical Need for Role Based Access Control. In: 2009 Annual Computer Security Applications Conference, pp. 13–22. IEEE (2009)
13. Hoffman, P.E., Grinstein, G.G.: A Survey of Visualizations for High-Dimensional Data Mining. Information Visualization in Data Mining and Knowledge Discovery. Morgan Kaufmann Publishers Inc. (2002)
14. Lee, M.D., Reilly, R.E., Butavicius, M.E.: An Empirical Evaluation of Chernoff Faces, Star Glyphs, and Spatial Visualizations for Binary Data. In: Proceedings of the Asia-Pacific symposium on Information visualisation, pp. 1–10 (2003)
15. Zhang, D., Ramamohanarao, K., Ebringer, T.: Role Engineering using Graph Optimisation. In: Proceedings of the 12. ACM symposium on Access control models and technologies, pp. 139–144. ACM (2007)
16. Kreuseler, M., Schumann, H.: A Flexible Approach for Visual Data Mining. IEEE Transactions on Visualization and Computer Graphics 8, 39–51 (2002)
17. Fuchs, L., Pernul, G.: Reducing the Risk of Insider Misuse by Revising Identity Management and User Account Data. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA) 1 (1), 14–28 (2010)
18. Ong, K.-H., Ong, K.-L., Ng, W.-K., Lim, E.-P.: CrystalClear: Active Visualization of Association Rules. In: ICDM'02 International Workshop on Active Mining (2002)
19. Ankerst, M.: Visual Data Mining. dissertation.de (2001)

Systematic Review and Meta-Analysis of IS Security Policy Compliance Research. First Steps towards Evidence-Based Structuring of the IS Security Domain

Danijel Milicevic and Matthias Goeken

Frankfurt School of Finance & Management, Frankfurt am Main, Germany
{d.milicevic,m.goeken}@fs.de

Abstract. Given the short supply of empiricism in ISS research, existing empirical evidence needs to be processed further than the scope of a single paper may allow. Other fields of science have long recognized the need for higher level analyses of research results in order to make them accessible to practitioners and develop a knowledge base. In our paper we perform an exhaustive literature research in the realm of empirical ISS research. As one of the recent research hotspots, we perform a systematic review of research results in Information Security Policy (ISP) compliance. We analyze and discuss the heterogeneity of research results and suggest a presentation format that may allow ISS practitioners to base their ISP design decisions on.

Keywords: Information Security, Policy, Compliance, Evidence-Based Research, Systematic Review

1 Introduction

Literature in the field of information systems security (ISS) is heavily dominated by conceptual and theoretical research, as has been shown by Siponen et al. [1]. They show that a majority of contributions is subjective-argumentative in nature and only a small percentage is based in empiricism. This is the cause for a set of problems for the research field. Given the limited amount of empirical research, the research community has been struggling to identify and agree upon a common set of theories as their kernel theories, as well as stable findings to construct an agreed-upon knowledge base. This makes the research field and the research results vulnerable to criticism regarding the scientific process and lack of rigor. However, this is not just a problem in the field of ISS research, but also other fields of information systems (IS) research. Without on-going efforts to provide and better discuss evidence for research findings and to consolidate existing research for IS (and potentially ISS) kernel theories and the construction of a knowledge base may remain fruitless for some time to come.

The lack of rigor, however, does not tip the scale in favor of an elevated relevance of ISS research to practice in its stead. Legislation and industry regulations put ISS practitioners increasingly under pressure to find sources of legitimacy for their deci-

sion-making. This opens up an opportunity for research to fill this void and find new avenues of interaction with ISS practice; an opportunity which, given the lack of empirical research and thus confidence in research results, has not yet been seized. Thus, ISS practitioners rarely become consumers of ISS research and as a result turn to so-called best practice frameworks or standards, which try to offer guidance and a point of reference when justifying their decision-making. ISS researchers have criticized this trend, as many of these standards have not been validated and are not developed based on a rigorous process. Thus the results are not necessarily reliable, due to the fact that they are based on little to no evidence [2]. However, ISS research has yet to offer a practical alternative and as such the use of such ISS frameworks and standards remains the primary foundation of managerial decision making in the ISS domain.

Interestingly enough, in other scientific disciplines we find more systematic treatment of research findings, most prominently in evidence-based medicine but also for example in psychology and educational research [3]. Scholars in these fields argued that there is much knowledge to be gained from summarizing existing research findings. Similarly it has been argued that the IS discipline would also benefit from better synthesizing findings and by adopting an evidence-based research approach [3-4].

To cumulate research findings, systematize the knowledge base, and communicate scientific results so called systematic reviews are gaining increasing acceptance and importance in other research disciplines. They are “attempts to collate all empirical evidence ... in order to answer a specific research question” [5] such as “what evidence do we have with respect to the effectiveness of countermeasures in ISS”. In this respect they differ from “traditional review” which tend to be summaries of broad topics answering questions such as “what is known about X” [6]. We argue that systematic reviews and meta-analysis are important means for both, to consolidate knowledge in the field of ISS research and to communicate the knowledge to other research communities and practitioners.

In this paper we lay the foundation by systemically analyzing existing empirical research findings. In a first step we assess the state of empirical research in the field of ISS in general by means of an exhaustive literature research. Building upon a set of empirical research results we focus then on factors that are relevant to the chosen research topic, information security policy compliance. By analyzing both collaborating and conflicting empirical research results, we derive a summary of findings table, a factor map, and perform meta-analysis on a selection of findings. These main contributions of our paper can be used by researchers as a reference point in their cumulative research, as well as be the basis for ISS practitioners in their decision-making.

The remaining paper is structured as follows: in a related work section we present existing literature research in the IS security policy compliance field and examine where they may fall short in terms of collecting empirically validated evidence. The third section describes our research methodology before we present the steps and results of our exhaustive literature research in the fourth section. Building on those preliminary results we extract empirical research which is relevant to our research topic and structure these findings, before performing a meta-analysis to test for significance of the pooled effects. In a discussion we identify seemingly stable research findings and make suggestions on the procedure of evaluating candidates for the body

of knowledge. We also discuss limitations of our work before concluding in the final section and suggesting interesting paths for future research.

2 Related Work

Reviews of existing research have a long-standing tradition in research, most commonly found in the form of literature reviews [7]. The goal of a literature review is two-fold. On the one hand, its purpose is to increase the rigor of research by basing it on the existing knowledge base. On the other hand, by describing existing research findings and their shortcomings in regard to the proposed research question, the researcher can demonstrate the contribution and thus relevance of his own research [8].

Given our objective, to identify empirical evidence in regard to information security policy compliance as well as assessing the state of this research stream, we are interested both in the type of available ISP compliance research as well as the existence of prior reviews and structuring efforts in this field.

One of the earliest works on social constructs in regard to information security policies is Kabay [9]. By transferring “well-established principles of social psychology” [9], Kabay contributed to an emerging research stream, with increasing numbers of publications and academic teaching starting around 1995 [10-11]. However, as Pahnla et al. [10] attest, up until their work in 2007, little to no empirical research had been published on this topic. In their overview three studies are brought up as examples for empirical research prior to 2007, which, however, must be dismissed under our scope. Straub [11] and Straub and Welke [12] both do not focus on policy compliance or information security policies in general in their work and as such do not collect data specifically for this research question. Woon et al. [13] look at the deployment of security features within home networks, which only in subsequent research would have implications on security policies and compliance with them.

A more detailed look into existing empirical literature will be provided in a later section. To the best of our knowledge no structured analysis of the current body of knowledge regarding ISP compliance has been published. In order to not only identify any kind of related work, but especially empirical evidence, we require a rigorous method for the literature search and analysis, as confidence in our results can only rely on our ability to transparently document the applied process [8].

3 Research Methodology

Levy and Ellis [14] suggest focusing on and finding high quality research by utilizing journal rankings. Given the often broad nature of such rankings and the very specific and narrow research stream in our scope, we decided to stray from the process in this aspect and cast a wider net by using research publishing portals as our starting points.

Another adjustment is due to the limited findings in the initial search for related work. While performing backward searches on articles found, a certain variety in terminology became apparent. Compliance was often subsumed under terms such as “user behavior”. Thus “compliance” became a too limited option as a search term.

The focus on empirical research leads to an additional sets of keywords. Since empirical evidence is derived from the application of empirical research methods, we chose a set of research methods as keywords that would cover the majority of empirical research, including: experiments, surveys, meta-analyses, field and case studies.

For the overall systematic review we apply the reference process proposed by Goeken [3]. It covers the steps from specifying the research question to the final presentation of the results. Table 1 shows the reference process with all activities.

Table 1. Phases of the Reference Process for a Systematic Review [3]

Phase	Output
1. Defining the research question	Research Question
2. Building the Infrastructure	Conceptual meta-model as a framework to represent the subject matter Classification system
3. Searching the literature	Preliminary inclusion of studies based on database research
4. Selecting the studies for inclusion	Set of final eligible studies
5. Assessing the quality of included studies and structuring of their results	Assessed and structured studies
6. Combining the results	Representation of the gained and integrated results
7. Create a structured report	Report on the findings and the evidence gained

In the **first phase** the research question must be defined. The research question will set the scope for the review and influences the direction subsequent activities take.

The broad research question at hand is how the two factors “controls” and “social factors” influence each other in IS security. To address this research question we look for answers in a specific case, in order to contribute to the original research question in a cumulative manner. As such we specify IS security policies as the IS security controls in question and analyze the relationship between this control and social factors that are influencing the effectiveness of said control. In order to answer the research question “What social factors influence the effectiveness of IS security controls?” we thus need to accumulate answers for specific instances of IS security controls. By structuring existing research and performing a meta-analysis, we aim to answer the research question: “What social factors have significant effects on a users’ intent to comply with IS security policies?”

The **second phase** is the building of an infrastructure. Having a conceptual model as a framework helps structure the research results early on, but must be balanced in accordance to whether or not a deductive approach seems beneficial. In our case we derive a simple tuple of concepts from our research question: social factors and ISP compliance. Goeken [3] suggests that choosing too broad concepts may result in a too high inclusion rate during the literature search, while specific conceptualizations may be too inflexible to accommodate most research. We – similarly to our decision to use portals instead of journals as starting points for our literature search – are choosing to make the search grid as big as possible to have a high degree of completeness possible by doing an exhaustive search. During the **third phase** the actual literature search is performed. This phase will be explained in detail in the following section. The **fourth phase**, dealing with the selection and inclusion in the review, will be discussed based

on a summary of findings table, which lists all relevant studies and factors. **Phase five** assesses the results and quality of included studies. For this the studies are classified based on an ordinal-scaled set of research methods and proposed hypotheses will be marked as either supported or not-supported. In the **sixth phase** we attempt to combine the results, which can be done either quantitatively using statistical methods to perform a meta-analysis, or by means of a qualitative, narrative discussion. Due to potential method plurality amongst the selected studies, we're applying both methods. We aggregate studies in a so-called summary of findings table, represent the variables that have been investigated in the literature by means of a "factor map", and, in addition, we will analyze prominent variables in greater depth by performing a meta-analysis. The **final phase** of writing a structured report is generally covered by the respective publication, which is made based on the performed research.

4 Collection and Processing of Empirical Evidence

As mentioned before, it is crucial for a literature search to be transparent, as it directly affects the credibility of the results. As such, verbosity and transparency are necessary and settings like queries, search fields and the following processing of the search results must be made accessible to the reader. Three elements had to be specified for the literature search process: 1) The source(s) the search is performed in. 2) The keywords used in the search queries. 3) The (optional) filtering process [16].

In our case we searched five large publishing portals, which cover – depending on the available subscription – more than 10,000 titles (ABI/INFORM, EBSCO, Emerald, ScienceDirect and SpringerLink). In every portal we chose whenever possible the combined Title-Abstract-Keywords field, otherwise linked the respective fields using Boolean operators and exhausting the possible combinations.

The keywords for the query comprised of synonyms (in the broader sense) of information security, such as IT security and system security. These keywords were linked using an 'AND' operator to a research method. The research methods queried for were: experiment, case study, survey, field study and meta-analysis. As a broad wildcard keyword we also used "empirical". The goal was not to determine the most accurate numbers during the literature search, which would have been challenging also due to some portals carry the same titles via licensing agreements. The goal was to achieve completeness; as such, redundancy was expected and planned for by using unique identifiers during data entry. The full combination of portals and keywords were used. The specific numbers per research method can be requested from the authors. The filtering process was stacked as follows: during the first stage all search results were captured and stored. In stage 2 the authors reviewed the title and abstract of each article and removed articles that were out of scope and belonged to a research area other than IS security, which results in 516 remaining studies. For stage 3 the authors retrieved the full articles and read the relevant sections of the articles (research methodology, data collection and findings) to exclude those that did not deliver actual empirical evidence, e.g. articles that used fictional data for conceptual work.

Overall, the literature search started with 2286 articles and ended with a final set of 354 articles with empirical results. This concluded phase three of the search process.

5 A Look at Empirically Validated Factors of ISP Compliance

With a database of articles in ISS research in place, the data had to be made accessible for phase four, during which the selection and inclusion of studies for the systematic review takes place. The articles got coded and information about the tested hypotheses got stored in “hypotheses-triplets”. These information triplets consisted of: the factor, the end point and the hypothesized effect. Article Id, along with relevant information about the research results, got linked to the edges of the resulting graph.

Whenever provided, the data-triplets got populated with information about the used statistical method and potential variant of it, along with information like use of bootstrapping and data source (e.g. whether IS students or IS professionals got surveyed). After preparing the data using this process, the selection process consisted of merely filtering the hypotheses-triples in the graph database for the keywords “policy” and “compliance” along with variations of these words. Out of the 354 papers with empirical methodology and results in all of ISS research, merely 9 papers dealt with similar constructs for the end point “compliance” in regard to the control “information security policies” and provided the required statistical key figures for further analysis.

The assessment of quality, which is phase five, can be performed using an ordinal scale based on the research methods, as it is common in evidence-based medicine and/or using the secondary data about the data collection (like population size, demographics, etc.). For example, a case study would provide a lower quality of evidence than a meta-analysis. A detailed description and discussion of evidence quality is out-of-scope for this paper, but can be found at Goeken and Patas [4]. Table 2 shows the summary of findings, a listing of the final selection of research papers and their variables, which are to be reviewed in the sixth phase of the research process.

The table is structured by the respective study which provides empirical evidence, information that helps determine the quality of the evidence as well as the conceptuple we determined in phase two, which are part of the infrastructure. In the last column the research results are listed in a condensed form, where a *** represents a significant correlation and thus a supported hypothesis. By default all effects are considered to be positive in the original hypothesis. Exceptions are marked as such; e.g. the bolded “Apathy” in the study by Foltz et al. [22]. Research results that conflict with a majority of other findings are bolded. As such is the case with “attitude” in the Herath and Rao [19] study, which is not supported in contrast to four studies in which this relationship was significant and thus the hypothesis behind it was supported.

Table 2. Summary of Findings

Study	Quality of Evidence: Method, Type of Evidence	Variables of the Research Question		Findings with respect to the variables
		Social Factors	ISP Compliance	
Hazari et al. 2008 [15]	- Survey - Empirical, quantitative evidence	- Attitude - Subjective Norm - Behavioral Control	Behavioral Intention	- Attitude *** - Perceived Behavioral Control *** - Subjective Norm
Bulgurcu et al. 2010 [16]	- Survey - Empirical, quantitative evidence	- Attitude - Normative Beliefs - Self-Efficacy	Intention to Comply	- Attitude *** - Normative Beliefs *** - Self-Efficacy ***
Johnston and Warkentin 2010 [17]	- Lab Experiment - Empirical, quantitative evidence	- Response Efficacy - Self-Efficacy - Social Influence	Behavioral Intent	- Response Efficacy *** - Self-Efficacy *** - Social Influence ***
Zhang et al. 2009 [18]	- Survey - Empirical, quantitative evidence	- Attitude - Subjective Norms - Perceived Behavioral Control	Intention	- Attitude *** - Subjective Norms - Perceived Behavioral Control ***
Herath and Rao 2009a [19]	- Survey - Empirical, quantitative evidence	- Attitude - Punishment Severity - Detection Certainty - Subjective Norm - Descriptive Norm	Security Policy Compliance Intention	- Attitude - Punishment Severity *** - Detection Certainty *** - Subjective Norm *** - Descriptive Norm ***
Myry et al. 2009 [20]	- Survey - Empirical, quantitative evidence	- Prevent. reasoning - Conventional reasoning - Postconventional reasoning - Openness to change - Conservation	(Hypothetical and Actual) Compliance with ISP	- Preventional reasoning *** - Conventional reasoning - Postconventional reasoning - Openness to change *** - Conservation *** - (negative effect)
Herath and Rao 2009b [21]	- Survey - Empirical, quantitative evidence	- Severity of Penalty - Certainty of Detection - Normative Beliefs - Peer Behavior - Perceived Effectiveness	Policy Compliance Intention	- Severity of Penalty *** - (negative effect) - Certainty of Detection *** - Normative Beliefs *** - Peer Behavior *** - Perceived Effectiveness ***
Foltz et al. 2008 [22]	- Survey - Empirical, quantitative evidence	- Attitude - Social Trust - Apathy (negative) - Subjective Norm - Perceived Behavioral Control	Behavioral Intention	- Attitude *** - Social Trust *** - Apathy *** - Subjective Norm - Perceived Behavioral Control
Pahnila et al. 2007 [10]	- Survey - Empirical, quantitative evidence	- Intention to comply - Information quality - Rewards	Actual compliance	- Intention to comply *** - Information quality *** - Rewards

Switching the perspective from studies to the studied constructs, we developed a Factor Map by summarizing the factors found in the studies' hypotheses. The studies of Myyry et al. [20] and Pahnla et al. [10] were disregarded, as their social and outcome constructs deviated from the more heterogenic research questions proposed by the remaining seven studies. The common thread between those was the Theory of Planned Behavior (TPB) by Ajzen [23], which puts the focus on the social constructs "subjective norm", "attitude" and "perceived behavioral control". During the operationalization and due to augmentations of the TPB model, further constructs have been introduced. The two factors "rewards" and "information quality" identified by [10] have been omitted, as neither had a direct influence on the intention to comply with a security policy. Figure 1 shows the Factor map of common constructs.

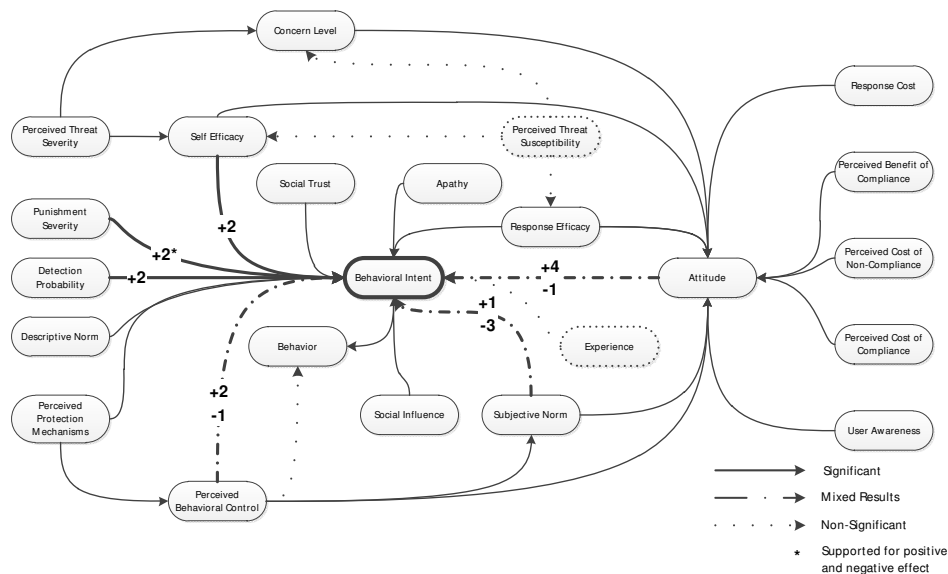


Fig. 1. Factor Map of Constructs in selected ISS Policy Compliance studies

Numbers on edges which have been investigated in multiple studies represent the amount of significant and non-significant findings. This helps identify findings which derive from common results for further analysis via meta-analysis in the next section.

6 Meta-Analysis of Selected Constructs

Meta-analytic procedures can be applied to further examine the findings from the summaries of findings table and factor map, and to focus the results with respect to the variables under investigation. The latter means that, in the following, we will refer to the constructs (variables) and try to deeper analyze the evidence we can derive by integrating the findings from the primary studies presented in table 2.

According to Higgins and Green [5] meta-analysis is “the use of statistical techniques in a systematic review to integrate the results of included studies” and is applied on the basis of “two or more studies”. In accordance with Matt and Cook [25] we assume that utilizing meta-analytic practice “will increase the precision with which an association is estimated” and that the generalization of an association that has been examined across the available primary studies might be strengthened. In what follows we are going to concentrate on three variables (“perceived behavioral control”, “attitude”, and “subjective norm”, as these have mixed findings, which can be seen in the factor map in figure 1) and discuss the findings with respect to the direction and magnitude of the effects across studies. Our analysis is based on the meta-analysis procedures described in Hunter and Schmidt [24]. All three studies that examine “perceived behavioral control” find a positive correlation between this variable and behavioral intention. It is important to mention that the study of Foltz et al. [22] does not find a significant relationship and that their hypothesis H1 (“Perceived control will positively affect behavioral intention”) is not supported by their data (see the forest plot in figure 2 where the line through the square representing the confidence interval (CI) crosses the line of no effect (0,00)). The studies of Hazari et al. [15] and Zhang et al. [18] show significant and positive effects since the squares representing the correlation are located to the right of the line of no effect and the CIs does not touch or cross the line of no effect. When determining the weight of single studies, the sample size of the studies is the main factor in a meta-analysis [5], [24].

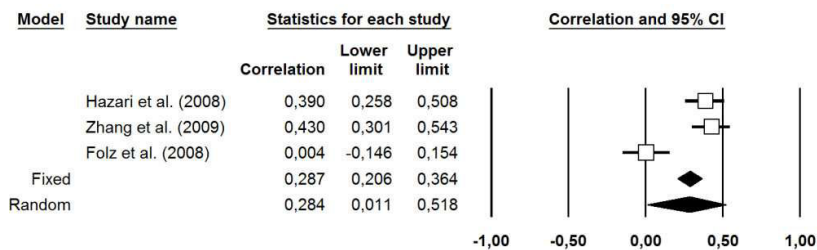


Fig. 2. Meta-analysis and Findings concerning the Variable “perceived behavioral control”

From a meta-analytical point of view, the results of the non-significant findings are important to compute an overall effect. The overall estimate is represented as a diamond in the forest plot. The center of the diamond represents the pooled point estimate for the effect and the width of the diamond is the certainty of the result, presented as a 95% CI. We utilized two models to calculate this overall effect (fixed effect model and the random effect-model). Even in the random effect-model – which is a more conservative estimator resulting in a greater width of the diamond – the overall effect is still significant, due to the fact that the estimate is based on a larger sample. In a meta-analytic context and if homogeneity assumptions hold, this is regarded as an increase of the precision. Hence, from the three studies and the meta-analysis performed, we can derive evidence that “perceived behavioral control” is likely to have a positive significant effect on behavioral intention.

Regrettably, we cannot explain the dissimilarities with respect to the significance, because the differences in either the demographics or the methodology does not allow for explaining the different results: E.g. all three studies base the variables on the theory of planned behavior and the studies of Foltz et al. [22] and Hazari et al. [15] are based on student samples, whereas Zhang et al. [18] sent the survey to an industry panel. Heterogeneity of the samples seems to have no influence on the results. In table 2 we found five studies that incorporate the variable "attitude". Comparable to the precedent analysis, the majority of studies discover a positive and significant association – only the findings of Herath and Rao [21] are not significant. The overall effect estimate is positive and significant in both models due to the larger sample size in the meta-analysis, resulting in a higher precision. Hence, if homogeneity of the studies is considered acceptable, the analysis indicates, that there is evidence for a positive association between “attitude” and “intention to comply”.

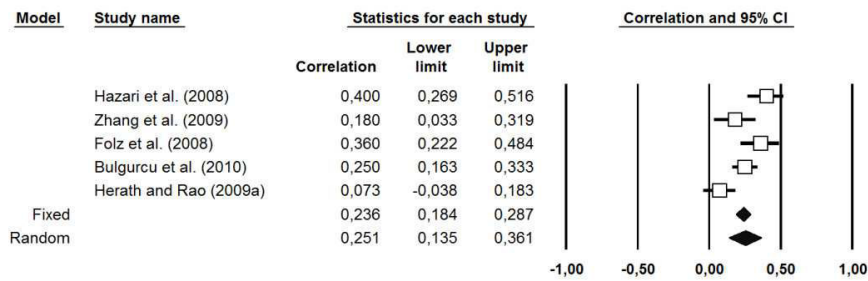


Fig. 3. Meta-analysis and Findings concerning the Variable “attitude”

In addition, the five studies allow for a more comprehensive generalization: The studies of Bulgurcu et al. [16] and Herath and Rao [19] are based on larger samples and the participants are working in different roles, companies of different sizes etc. (instead of using students as proxies). Therefore, it is reasonable to conclude, that the findings have a higher validity than the findings concerning “perceived behavioral control”. In this respect it would be interesting to perform subgroup-analysis on them.

One interesting finding is that in three of four studies the association between "subjective norms" and compliance intention is non-significant, because the CIs or the squares representing the effect are touching the line of no effect.

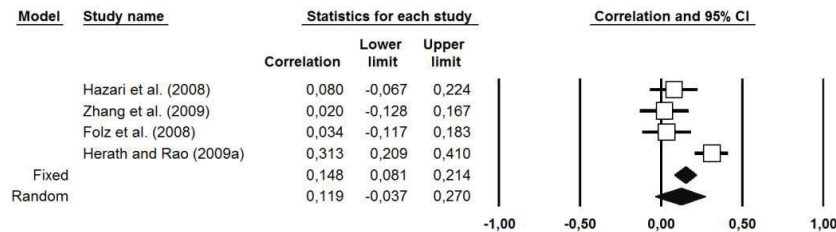


Fig. 4. Meta-analysis and Findings concerning the Variable “subjective norm”

In addition, the computation of the pooled effect does not give a clear picture, because the more conservative random effects-model shows that the effect is not significant. Hence, in this case we cannot derive clear and positive evidence from the meta-analysis.

7 Discussion

Our systematic review ultimately led to an investigation of the three social factors found in the Theory of Planned Behavior (TPB): “attitude”, “subjective norm” and “perceived behavioral control”. As the summary of findings table and factor map showed, these factors had mixed results in the set of studies we selected and thus were good candidates for further analysis. We investigated in our meta-analysis the pooled effects of said variables using both a fixed effects-model and a random effects-model. The latter, being a more conservative estimator, was used to determine whether or not the pooled larger sample had a significant effect or not. For the variable “perceived behavioral control” the estimator showed a significant positive effect, despite the not-significant results of the Foltz et al. [22] study. Potential reasons for the mixed results were discussed, but given a similar demographic (student proxies) to a study where significant effects were measured [18], we cannot determine the cause for those results, other than an error in the measurements.

We found similar results in the meta-analysis of the “attitude” variable, which appears to have the strongest positive effect on a users’ intention to comply with an IS security policy. The Herath and Rao [19] study is the exception for this variable and needs to be examined further, given that it is also the exception for the variable “subjective norm”. The variable appears to be a weak predictor for the intention of a user. This finding, based also on the pooled effect using the random effects-model, has also been suggested by Armitage and Conner [26] in their meta-analytic review of TPB use in literature. Overall our meta-analysis also could not show a significant effect between “subjective norm” and behavioral intention.

Another interesting finding, which can be found both in the summary of findings table and the factor map, are two significant effects of “punishment severity” on behavioral intention, as one study suggests a positive significant effect [19] and the other a negative significant effect [21]. Since the same dataset was used, we assume the reason for this difference may lie in the data analysis, potentially the performed bootstrapping. As such further research is required to determine the effect of (perceived) punishment severity on the intention to comply with an IS security policy.

The summary of findings table lists the variables which have been examined in the literature so far. Based on the amount of studies which cover a certain construct, the designer of an IS security policy may have a higher confidence in said results, if they are collaborated, e.g. by a meta-analysis. As such, our result types, summary of findings table, factor map, and the meta-analysis forest plot may serve as a foundation for design-decision-making. Multiple reasons can be the cause for heterogeneous results when compiling a systematic review. Analyzing and pinpointing potential causes is

one of the key elements in finding ways to consolidate empirical results to a shared knowledge base. In the example of the Herath and Rao [19] study and the atypical result of “attitude” having no significant effect, the authors hypothesize that it may be due to “reasons such as context, sample or other extraneous factors”. These factors can vary from organizational culture to small differences in method parameters or – despite best efforts – faults in the data. Researchers must not only apply methods and discuss the results, but also reflect on the applicability of said method. Since Mingers [27] call for method plurality there has been much talk about extending the “toolkit” and tackling research questions from different angles to triangulate new “truths”, but very few followed through and reflected on the current tools of the trade [28]. Another source for conflicts in study results is the theoretical foundation and structuring of items for tools like surveys. As D’Arcy and Herath [29] show, even with the same underlying theoretical foundation, a plethora of contradictory results may occur.

The benefits of structuring empirical research both for research and practice are clear. With stronger structuring of the existing body of knowledge future research is standing on better footing and is more “consumable” for non-researchers; a demand more than a decade old [31]. Practitioners on the other hand are able to base their decision-making on scientific results rather than industry studies and “expert opinion”. The neutral corner, which science occupies in the eyes of practitioners, should be a major reason for practitioners to look for an intellectual exchange with the scientific community, but the scientific community must meet practitioners in the middle and find simplicity whenever possible, another demand that has not yet been met [32].

Our findings have implications for the direction of future research regarding policy compliance, related research topics, and ISS practice. Out of the three major social factors stemming from TPB, only “attitude” showed strong support in our meta-analysis. While the random effect-model supported “perceived behavioral control” as a factor influencing the intention to comply with a policy, confidence in the correlation is less robust than it was the case for “attitude”. “Subjective norms” as a factor does not find support in our meta-analysis nor in the majority of studies we reviewed. For information security researchers this should indicate the need to reflect on the applicability of TPB as a theoretical foundation for their future compliance research.

Our results also influence related research topics in ISS, especially under the design-science paradigm. The design of security-related artifacts like methods, models and tools can benefit from a strong grounding in empirical evidence. An information security policy is such a design artifact and can have a higher degree of efficiency or efficacy when positively correlated factors are considered in its design. With the strong support we found for the factor “attitude”, design-science researchers need to adapt it as a critical success factor for security controls related to user compliance. This may suggest increased success of policy deployment if it is combined with security awareness training, which helps users develop a security mindset. In a field where approaches akin to engineering are common (e.g. requirements elicitation amongst stakeholders and consultation of “expert opinion”), insights from empirical research can improve design and be beneficial in the adaptation and configuration of controls for the specific needs of organizations that deploy them. Practitioners can base their decisions on empirical evidence that is collated by means of meta-analysis, which

elevates the level of confidence in research findings. As such the development of security guidelines and frameworks (as best practice or reference models) can benefit from systematic reviews as it allows better grounding of the selection and prioritization of controls in empiricism rather than then popular opinion.

Another goal of our contribution is to motivate other researchers to perform similar analyses on other ISS topics. While it is a de-facto standard to review and discuss related work by other researchers, important aspects like the quality of the empirical evidence are hardly ever considered. We urge fellow researchers and practitioners to consider the applicability of research methods and the generalizability of results when discussing research. Not only would this help the research field itself with syndicating research results and moving towards a theory kernel, but it also helps ISS practitioners get easier access to relevant information for their decision-making process.

8 Conclusion

In this paper we have presented a way of accessing, structuring and utilizing empirical evidence that can be found within ISS research. Specifically, we presented a selection of empirical studies that hold empirical evidence for the research stream of ISS policy compliance and may hold candidates for a consolidated knowledge base for this topic.

Using a combination of existing methodologies and applying minor subject-dependent adjustments, we showed a transparent documentation of a literature search process to find and filter relevant empirical findings for a structured review. A strong focus was put on verbosity in order to allow for reproducibility of the presented results, as it is the only path to increase confidence through rigor.

The main contributions of our research are the three main results, the summary of findings table, the factor map, and the meta-analysis. By means of the summary of findings table and the factor map we intend to represent empirical findings in the policy compliance research area at a glance. Both give an overview of relevant studies and represent the associations studied so far in the literature. Furthermore we utilized meta-analysis practices to calculate findings from different studies in a condensed manner. We were able to derive evidence for a selection of studies and the variables they examine, in particular for three social factors. In this respect, meta-analysis helped answer the research question. We confirmed significant effects for the two social constructs “attitude” and “perceived behavioral control” towards the intention to comply with an ISS policy using estimator models for the pooled effect. While the meta-analysis provides clear evidence for two factors and their influence on compliance intention, there are ambiguous results for the factor “subjective norms”. Hence, further analysis and primary research seems to be necessary.

From a methodological point of view, the analysis we performed here is to some degree simplistic. E. g. due to space limitations we did not perform a homogeneity analysis or a sensitivity analysis. In addition, in further analysis, formal procedures to correct effects of study artifacts and errors should be applied in a systematic and transparent way in order to improve the comparability of the results reported by single studies [24-25]. Our intention in this paper was to demonstrate the applicability and

usefulness of meta-analysis in the IS security domain and to put forward a case for the evidence-based structuring of findings. For this purpose we outlined concrete adaptations to ISS practice when dealing with information security policy development.

In combination with a strong cumulative research tradition, the review and structuring of existing research results can lead to progress for the field as a whole and strengthen then foundation for future research, as existing knowledge becomes more accessible to ISS practitioners and the existing theory/practice disconnect is reduced.

References

1. Siponen, M., Willison, R., Baskerville, R.: Power and Practice in Information Systems Security Research. In: ICIS 2008 Proceedings, Paper 26 (2008)
2. Siponen, M., Willison, R.: Information security management standards: Problems and solutions. *Information & Management* 46, 267-270 (2009)
3. Goeken, M.: Towards an Evidence-based Research Approach in Information Systems. In: ICIS 2011 Proceedings, Paper 10 (2011)
4. Goeken, M., Patas, J.: Evidence-Based Structuring and Evaluation of Empirical Research in Requirements Engineering – Fundamentals, Framework, Research Map. *Business & Information Systems Engineering* 2 (3), 175-185 (2010)
5. Higgins, J.P.T., Green, S.: *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 <http://www.cochrane-handbook.org/> (2011)
6. Gough, D.: Qualitative, Quantitative and Mixed Methods Systematic Reviews to Support Professional Decision Making in Education. In: Böttcher, W., Dicke, J.N., Ziegler, H. (eds): *Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis*. Waxmann (2009)
7. Garfield, E.: Proposal for a new profession: scientific reviewer. *Essays of an Information Scientist* 3, 84-87 (1977)
8. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: Proceedings of the 17th European Conference on Information Systems (ECIS), pp. 2206-2217 (2009)
9. Kabay, M.E.: Social Psychology and Infosec: Psycho-Social Factors in the Implementation of Information Security Policy. In: Proceedings of the 16th National Computer Security Conference, pp. 274-283 (1993)
10. Pahlila, S., Siponen, M., Mahmood, A.: Employees' Behavior towards IS Security Policy Compliance. In: Proceedings of the 40th Hawaii International Conference on System Sciences, pp. 156-166 (2007)
11. Straub, D.: Effective IS Security: An Empirical Study. *Information Systems Research* 1 (3), 255-276 (1990)
12. Straub, D., Welke, R.J.: Coping with Systems Risk: Security Planning Models for Management Decision-Making. *MIS Quarterly* 22 (4), 441-469 (1998)
13. Woon, I.M.Y., Tan, G.W., Low, R.T.: A Protection Motivation Theory Approach to Home Wireless Security. In: Proceedings of the 26th International Conference on Information Systems, pp. 367-380 (2005)
14. Levy, M., Ellis, T.J.: A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science Journal* 9, 181-212 (2006)

15. Hazari, S., Hargrave, W., Clenney, B.: An Empirical Investigation of Factors Influencing Information Security Behavior. *Journal of Information Privacy & Security* 4 (4), 3-20 (2008)
16. Bulgurcu, B., Cavusoglu, H., Benbasat, I.: Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly* 34 (3), 523-548 (A1-A7) (2010)
17. Johnston, A.C., Warkentin, M.: Fear Appeals and Information Security Behaviors: An Empirical Study. *MIS Quarterly*, 34 (3), 549-566 (A1-A4) (2010)
18. Zhang, J., Reithel, B.J., Li, H.: Impact of perceived technical protection on security behaviors. *Information Management & Computer Security* 17 (4), 330-340 (2009)
19. Herath, T., Rao, H.R.: Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems* 18, 106-125 (2009)
20. Myyry, L., Siponen, M., Pahnla, S., Vartiainen, T., Vance, A.: What levels of moral reasoning and values explain adherence to information security rules? An empirical study. *European Journal of Information Systems* 18, 126-139 (2009)
21. Herath, T., Rao, H.R.: Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems* 47, 154-165 (2009)
22. Foltz, C.B., Schwager, P.H., Anderson, J.E.: Why users (fail to) read computer usage policies. *Industrial Management & Data Systems* 108 (6), 701-712 (2008)
23. Ajzen, I.: The theory of planned behaviour. *Organizational Behavior and Human Decision Processes* 50 (2), 179-211 (1991)
24. Hunter, J.E., Schmidt, F.L.: *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage, Newbury Park, CA (2004)
25. Matt, G.E., Cook, T.D.: Threats to the validity of research syntheses. In: Cooper, H., Hedges, L.V., Valentine, J.C. (eds.): *The handbook of research synthesis and meta-analysis*. Russell Sage, New York (2009)
26. Armitage, C.J., Conner, M.: Efficacy of the Theory of Planned Behavior: A meta-analytic review. *British Journal of Social Psychology* 40, 471-499 (2001)
27. Mingers, J.: Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research* 12 (3), 240-259 (2001)
28. Mingers, J.: The paucity of multimethod research: a review of the information systems literature. *Information Systems Journal* 13, 233-249 (2003)
29. D'Arcy, J., Herath, T.: A review and analysis of deterrence theory in the IS security literature: making sense of the disparate findings. *European Journal of Information Systems* 20, 642-658 (2011)
30. Thompson, S.G., Sharp, S.J.: Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 18 (20), 2693-2708 (1999)
31. Robey, D., Markus, M.L.: Beyond Rigor and Relevance: Producing Consumable Research about Information Systems. *Information Resources Management Journal* 11 (1), 7-15 (1998)
32. Benbasat, I., Zmud, R.W.: Empirical research in information systems: the practice of relevance. *MIS Quarterly* 23, 3-16 (1999)

Experimental Evaluation of a Process Benchmarking Tool in a Green Business Process Management Context

Matthias Gräuler and Frank Teuteberg

University of Osnabrück, Accounting and Information Systems, Osnabrück, Germany
{matthias.graeuler, frank.teuteberg}@uos.de

Abstract. Using a combination of metamodels, ontologies, green performance indicators and metrics, we apply a novel approach in Semantic Business Process Benchmarking to the area of Green Business Process Management (Green BPM). Up to now, process benchmarking has mainly been a manual process; the approach described and empirically evaluated in this paper partially automates the time-consuming and costly process analyses while introducing more flexibility regarding varying terminology, level of abstraction and modeling notation. Also, overviews of literature relevant to the field of Green Semantic BPM and commonly applied metrics in a Green BPM context are given.

Keywords: Semantic Process Management, Ontologies, Sustainability, Green Business Process Management, Benchmarking

1 Introduction

Over the last two decades, many research efforts aimed at developing more energy efficient technologies, alternative energy sources and ecological end-of-life for products. We consider these efforts to be highly relevant and desirable in mitigating our collective impact on the environment, however, the possibilities offered by improving existing operational practices are often ignored. Business Process Management (BPM) offers an integrated, holistic approach to the management of sustainability change [1], which is needed to change these practices efficiently. By looking at the current state of an organization's business processes (BP) as well as possible improvements through comparison with other organizations' BP or reference processes proposed by researchers, considerable improvements can be implemented.

Business Process Benchmarking is commonly used to identify areas in which organizations can improve their efficiency [2], but it is primarily a time-consuming, manual process performed by domain experts. Therefore, as Drew pointed out in 1997, "ways must be found for doing [benchmarking] faster, more effectively and economically, without sacrificing rigour [sic] or integrity of approach [sic]" [3]. Benchmarking primarily suffers from two difficulties that need to be overcome:

- There are many different modeling notations for BPs (e.g. Event-Driven Process Chains (EPC) [4], Business Process Model and Notation (BPMN) [5]), which can-

not simply be compared because of syntactic incompatibility and varying syntactic richness.

- Even if a process or a process step describes the same chain of activities in two different models, in the same language and perspective, the terminology and semantics of the models may differ, which prohibits a direct comparison [6]. This may stem from the fact that process models can be created from various perspectives at different levels of abstraction [7], because models are created for varying purposes, not all of them requiring indication of each atomic activity, and different organizational entities use different terminology for describing the same domain.

With the help of semantic BPM, researchers try to overcome several issues with BPM (e.g. work done by domain experts, mostly interpreting unstructured information) by providing a common terminological reference point [8-9]. Up to now, there is no generally acknowledged meta-model covering all aspects of process models [7].

In this paper, we follow the suggestion to annotate BPs with corresponding effects of individual activities on the environment (e.g. emissions, waste), which are accumulated along the process flow [6], [10-13]. Using this approach, two or more (sub-)processes can be benchmarked to find ways to improve efficiency. The proposed approach must not be confused with life cycle assessment (LCA) [14], as it is by no means intended to replace but supplement LCA. During the course of our research, we acquired practical examples of BPs in different modeling notations, annotated them with semantic information and demonstrated how process benchmarking can be performed semi-automatically with only a small degree of manual modifications using the software package SEMAT. To achieve the necessary degree of rigor, we started with a systematic literature review [15] to identify the current state of Semantic and Green BPM and also appropriate metrics applicable to Green BPM. To achieve our goals, we laid out a research agenda with the following research questions:

RQ 1: What is the current status quo of Semantic and Green BPM?

RQ 2: What metrics are applied to Green BPM?

RQ 3: Does SEMAT support Green BPM by assisting queries?

This paper is structured as follows: Section 2 details the literature review and lists metrics for Green BPM found during our literature review. In section 3, the proposed benchmarking process is explained in detail and an example is given. Section 4 gives an overview over the design for the experimental validation of the tool, whose results will be shown in section 5. The paper closes with the 6th section, concluding the paper and providing starting points for future research.

2 Literature Review and Related Work

As we wanted to focus on high-quality literature to maximize the reliability of our findings [15-16], we examined the top 20 journals ranked by the AIS as well as the proceedings of the A-ranked conferences ICIS, ECIS, AMCIS, ACIS, PACIS accord-

ing to the ERA conference rating and additionally WI. We used the following search terms in varying combinations: *Benchmarking, Business, Green, Management, Modeling, Ontology, Process, Semantic, Sustainable and Sustainability* to investigate the current works on Green BPM as well as the progress made in the field of Semantic BPM underlying our research. We read title and abstract of each paper and determined whether it was relevant for our research or not (i.e. at least two of the following seven criteria have to be met). We also conducted a forward and backward search to identify more relevant literature. Table 1 summarizes the subset of results we consider to be relevant. In total, we identified 31 papers relevant to our research. The prominence of a certain topic is indicated by the number of asterisks, with a maximum of three asterisks. If a topic is not focused in a publication, it is denoted with a minus. The topics used for comparison to our approach are:

- *Business Processes (BP)*: Does the paper focus on business processes?
- *Benchmarking (Ben)*: Does the paper propose a process benchmarking approach to analyze and compare individual process models efficiently and effectively?
- *Ecological and/or Social Sustainability (Sust)*: Does the paper focus on ecological and/or social sustainability?
- *Semantic Approach (SA)*: Does the paper employ a semantic approach to analyze business processes?
- *Research Agenda (RA)*: Does the paper propose a research agenda or directions for future research?
- *Evaluation Approach (EA)*: Do the authors present an evaluation approach?
- *Focuses on...*: On which problem domain does the study focus?

The results of the literature review are displayed in table 1. The results were loosely clustered by the focused topics; due to the different dimensions examined in the review, a distinct clustering was neither possible nor desirable. As can be seen from our literature review, the combination of research topics (i.e. BPM, Benchmarking, Green IS and the Semantic Web) is largely unexplored, which necessitates this combination.

Three of the papers found in the literature review were written by the same research group [10], [17-18]. As can be seen in table 1, these works are closely related to our research. Similar to our approach, they propose annotating process models with semantic effects regarding e.g. energy efficiency or resource consumption and using the software tool ProcessSEER [19] to obtain the cumulative effects at the end-event. However, in contrast to our method, their approach only supports BPMN and, to our knowledge, there are no efforts to change this. Furthermore, their approach remains largely invalidated. At the time of writing, we were unable to find related papers published after 2010, which indicates that their research in this area has ceased.

While analyzing the literature, we found several Green BPM metrics (cf. table 2). This list of metrics is not meant to be exhaustive, it is rather intended to illustrate that Green BPM should not simply focus on single aspects but needs to be a holistic approach including interdependencies among different metrics. We focused on the social and ecological dimensions of the “triple bottom-line”, omitting the economic perspective, because contrary to the other two dimensions, this perspective has already been explored to a much more mature extent [20].

Table 1. Results of the Literature Review

Ref	BP	Ben	Sust	SA	RA	EA	Focuses on...
[10]	***	***	***	***	-	-	...extending existing BPM technology to enable organizations to be informed about their processes' carbon footprint
[18]	***	***	**	***	-	-	...an algebraic framework for green BPM
[21]	***	*	***	-	***	-	...using BPM techniques to leverage Green IT initiatives
[22]	***	***	-	***	-	-	...an approach for pattern-based process model analysis
[6]	***	***	-	***	*	*	...semantic interoperability of BP models
[9]	***	**	-	***	*	*	...the semantic interoperability of BPs
[23]	***	**	-	**	-	-	...semi-automatic checking of semantically analyzable BP models
[24]	***	*	-	***	-	-	...compliance checking in financial institutions
[25]	***	*	-	-	**	-	...a knowledge-based system aiding process redesign
[17]	***	-	***	*	***	-	...issues and a research agenda for green BPM
[1]	***	-	***	-	***	-	...call for action for investigating BPM to create more sustainable organizations
[26]	***	-	***	-	**	-	...research agenda for green BPM
[27]	**	-	***	-	***	-	...various areas of Green IT, e.g. Green BPM
[20]	**	-	**	-	**	-	...economic sustainability in BPM
[28]	***	-	***	-	-	**	...proposal of an approach for measuring CO ₂ emissions during BP execution
[29]	***	-	***	-	-	-	... the evaluation and comparison of process designs
[12]	***	-	-	***	*	-	...combination of semantic web services and BPM to create the concept of semantic BPM
[30]	***	-	-	**	**	-	...the necessity of methodological elaborations of BPM
[31]	***	-	-	*	**	-	...inter-organizational BP design
[32]	***	-	-	***	-	*	...partially automatic planning and modeling BPs
[7]	***	-	-	***	-	-	...usage of an ontology framework to reduce complexity of e.g. process modeling
[11]	***	-	-	***	-	-	...ontologically representing the business and IS perspective on BPs, and on translating between these two perspectives
[13]	***	-	-	***	-	-	...semantic annotation of EPCs to specify the semantics of individual process model elements
[33]	***	-	-	**	-	-	...domain-specific semantic BP modeling language for banks
[34]	***	-	-	***	-	-	...process verification using process logic
[35]	***	-	-	***	-	-	...analyzing process mining
[36]	***	-	-	***	-	-	...a domain ontology based approach to support BP design
[37]	***	-	-	***	-	-	...an integrated model for inter-organizational BP integration
[38]	***	-	-	**	-	-	...finding similarities and contradictions in BP models
[39]	***	-	-	-	*	*	...BPM success
[40]	-	-	***	-	-	*	...sustainable performance measurement at airlines
[41]	*	*	**	-	-	**	...description of a tool for managing material flow networks

Hoesch-Klohe and Ghose [18] and previously Teuteberg et al. [6], [42] propose to annotate each activity in a business process model and to accumulate the values along the process chain, taking the different possible paths of process execution into account. The metrics listed in table 2 can be attached to any construct of any modeling language describing an action (e.g. functions in EPCs or activities in UML-AD and BPMN). Some of them can easily be quantified (e.g. Energy Consumption, Waste Generation), whereas others can only be measured in a qualitative scale (e.g. Environmental Performance) [18].

Table 2. Metrics for Green BPM

Dim	Metric / Explanation / Example	Reference(s)
Env	<u>Air Quality</u> indicated by e.g. Air Quality Index	[18]
Env	<u>Congestion</u> leads to unnecessary consumption of resources	[40]
Env	<u>Emissions of greenhouse gases, ozone depleting substances or other emissions</u> in e.g. CO ₂ -equivalents	[1, 10, 17, 18, 26–28, 40, 43]
Env	<u>Energy Efficiency/Consumption</u> in e.g. kWh/unit	[1, 17, 21, 26–29, 43]
Env	<u>Environmental Performance</u> : Qualitative measure, representing a variety of measures; performance could range from A to D	[18]
Env	<u>Fuel Efficiency/Consumption</u> in e.g. km/100 liter fuel	[21, 28, 40]
Env	<u>Odour emission</u> in e.g. olf or decipol	[44]
Env	<u>Paper Consumption</u> in e.g. sheets/employee	[1, 27, 28]
Env	<u>Radiation</u> in e.g. sievert	[44, 45]
Env	<u>Waste Generation</u> in e.g. kg/unit	[18, 28, 40, 43]
Env	<u>Water Consumption</u> in e.g. liter/unit	[18, 26, 43]
Env	<u>Water Discharge</u> in liters	[43]
Env,	<u>Noise Generation</u> measureable in e.g. decibel	[40, 44]
Soc	<u>Probability of accidents/casualties</u>	[44]
Soc	<u>Training and Development</u> required for new employees	[40]
Soc	<u>Workforce size</u> indicates the number of employees needed	[43]

3 Green Process Benchmarking Approach – An Example

Figure 1 illustrates the actions performed during the benchmarking process using an example from the printing industry and applying some of the metrics introduced in table 2. This example was selected because it describes the production of photobooks, a rapidly growing market, and provides information about CO₂ emissions. Furthermore, the used production process could be inferred from the descriptions.

We used a process described by Hausmann [46] to create two very simple process models: one using EPC and one using BPMN. We annotated each step with the effects on the environment (i.e. energy consumption, emissions). The process models are transferred into a machine-readable form, such as a Comma Separated Values (CSV) file, so they can easily be imported by the software created for this benchmarking approach: SEMAT. Although essentially accomplishing the same outcome, both processes exhibit varying degrees of abstraction and different terminology. They are also modeled in an entirely different notation. To overcome the problems associated with this kind of benchmarking endeavor, we use a combination of methods.

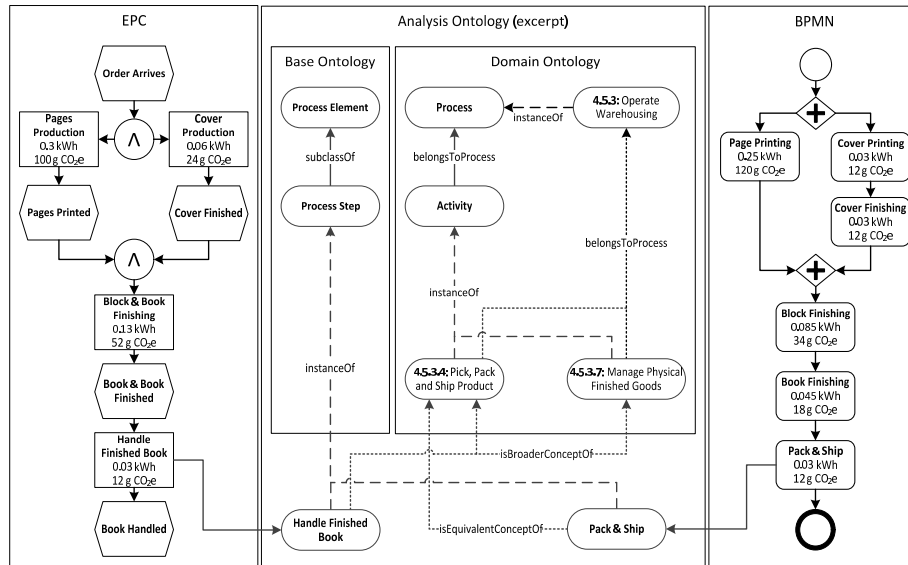


Fig. 1. Example for the Benchmarking Process

First, a domain ontology describing the benchmarking domain has either to be created or, in case this benchmarking approach has been used in the specific domain before, selected from a repository (Step 1). For the example in figure 1, we used the Process Classification Framework [47] that was transformed into the desired artifact. Then, a base ontology has to be created or reused (Step 2). The base ontology must be able to represent all relevant constructs of the used modeling languages. Furthermore, it can be reused if it provides support for the modeling languages used in the process models to be benchmarked. Figure 1 shows that the element types “function” (EPC) and “task” (BPMN) are syntactically equivalent and, therefore, mapped to the base ontology’s concept “Process Step”. Even though domain and base ontologies are merged into one analysis ontology later, at this point, they should be kept separate because the domain ontology can be reused regardless of the used modeling languages and the base ontology can be reused in any domain. The third step introduces the process models into the benchmarking process. A domain expert maps the elements of the process models to instances of the domain ontology (Step 3). In contrast to Höfferer (2007) [9], who suggested only one mapping type, we use three mapping types: one each if the element is equivalent to the domain concept, if it is represented in a broader or in a narrower sense. By resorting to a domain experts’ knowledge for mapping, we are able to overcome the problems raised by inconsistent terminology and different degrees of abstraction. Subsequently, the sustainability performance information needs to be captured and inserted into the ontology (Step 4). Using the specifically designed tool SEMAT, the process models are imported from text-files and transformed into an instance ontology of the base ontology. This instance ontology is then mapped to the base ontology using the mappings made in step three and, subsequently, merged into a single analysis ontology (Step 5). The analysis ontology can easily

be queried using query languages such as SPARQL [48] and by using predefined metrics supplied by SEMAT (Step 6).

Figure 2 demonstrates how the results of the analysis are displayed. The selected model (the process modeled in EPC notation from figure 1) and metrics are highlighted in the panel at the top. The table at the bottom shows the actual process performance and accumulates them along the process flow, enabling analysts to quickly assess process performance. As can be seen, any manufacturing process is generalized into the concept of “Produce Product”, which was taken from the PCF taxonomy.

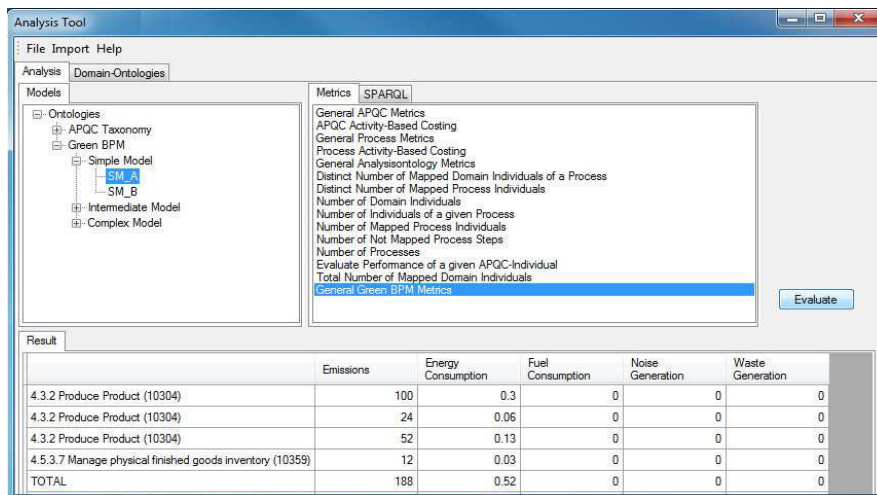


Fig. 2. Screenshot of the analysis results presented by SEMAT

4 Evaluation Concept

The previously described benchmarking process can be split into two parts: One that has to be performed once per process (steps 1-3) and one that has to be carried out every time the process’s performance information changes (steps 4-6). It should be noted that steps 1-3 are more time-consuming than steps 4-6, especially when an appropriate domain or base ontology has to be created. The scope of this research is limited to the evaluation of the advantageousness of our proposed methodology in respect of querying the data. Therefore, we decided to only evaluate step six, as this is the most frequent step and does not require extensive domain or process knowledge. It must be taken into account that in a manual comparison steps 1 to 5 are unnecessary and, have thus been ignored. Therefore, both groups operate within the same context and boundaries. This step was examined using an experimental setup.

Prior to the experiment (can be requested from the corresponding author), all test persons were screened whether they possessed sufficient knowledge about the modeling notations used later in the experiment; also, specific information, such as *Computer Self-Efficacy (Comp. SE)* and a self-assessment of *BPM Knowledge*, were collected using 7-point Likert-scales. Then, the test persons were randomly divided into two

groups that received different treatments. Group A (Manual) was given three more complex process models, each in a different notation (i.e. EPC and BPMN) and with annotated emissions, resource and energy consumption. Group B (SEMAT) was given a short introduction to our software used in the proposed benchmarking process and then received the same treatment as group A, except that they used analysis results from the software to solve the same tasks on identical process models.

The constructs *Comp. SE* and *BPM Knowledge* are based on a self-assessment by the test persons. According to the self-efficacy theory, the expectations of personal efficacy determine whether a certain behavior will be initiated, how much effort will be expended and how long it will be sustained in the face of obstacles [49]. Higgins and Compeau found that *Comp. SE* has a significant impact on individuals' expectations of the outcomes of using computers, their emotional reactions to computers as well as their actual computer use [50]. Another theory that was applied in our research model is the Theory of Task-Technology-Fit (TTF), which asserts that IS are means for users to complete tasks (i.e. summarizing process performance). The higher the TTF, the better the performance of the users (i.e. fast completion, high accuracy) [51]. The variables *Perceived Usefulness* (Perc. Usef.) and *Perceived Ease Of Use* (Perc. EoU) were taken from the Technology Acceptance Model and its successors [52]. These variables are assumed to be fitting, as they are commonly used to empirically analyze an IT artifact. It was considered whether the Expectation Confirmation Theory [53-54] applies in the context of the experiment. However, we decided not to include expectations and their (dis-)confirmation because the majority of the test persons have experience with BPM tools, but not to an extent in which an evaluation of these variables can be considered useful and worthwhile. Using this approach, we were able to test the following hypotheses:

- H1 Test persons who think of themselves as more computer proficient will...
 - H1a ...be more accurate when using SEMAT
 - H1b ...finish tasks faster when using SEMAT
- H2 Test persons who indicated that they have extensive knowledge of BPM will...
 - H2a ...be more accurate when not using SEMAT
 - H2b ...be more accurate when using SEMAT
 - H2c ...finish tasks faster when not using SEMAT
 - H2d ...finish tasks faster when using SEMAT
- H3 Test persons using SEMAT for the analysis will...
 - H3a ...be more accurate than those who did not use SEMAT
 - H3b ...finish tasks faster than those who did not use SEMAT
- H4 The test persons' rating for Perc. EoU will be higher if they...
 - H4a ...are more accurate when using SEMAT
 - H4b ...finish tasks faster when using SEMAT
- H5 The test person's rating for Perc. Usef. will be higher if they...
 - H5a ...are more accurate when using SEMAT
 - H5b ...finish tasks faster when using SEMAT
- H6 The Perc. EoU will have a positive effect on Perc. Usef. when using SEMAT
- H7 The test person's satisfaction will be positively related with...

H7a ...Perc. Usef. when using our SEMAT

H7b ...Perc. EoU when using our SEMAT

These hypotheses and the respective constructs are displayed in figure 3. All constructs in squares are measured by 7-point Likert-scales using at least four items to measure each construct. The constructs in oval shapes are measured during the experiment. Whereas most constructs' values were calculated using the arithmetic mean of the items measuring them (equally weighed), the variable *Time* was measured by the time needed to fulfill the assigned tasks. Respectively, *Accuracy* was measured by the number of correctly solved tasks.

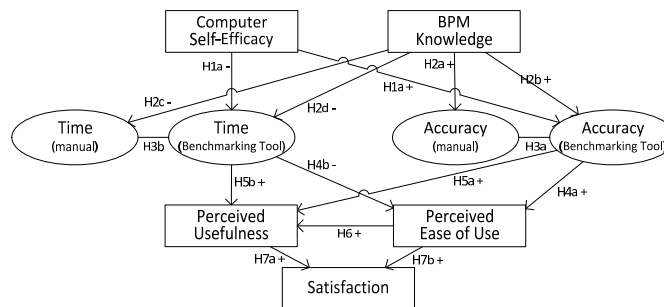


Fig. 3. Research Model

5 Results

In total, we collected 24 datasets from the participants, 18 of them (under-)graduate students and 6 Ph.D. candidates. 11 participants were assigned to the group performing a manual analysis, 13 to the group using SEMAT. Table 3 gives an overview of the descriptive statistics as well as reliability coefficients for each construct measured by Likert-scales. To display sufficient convergent validity, three criteria should be met: A minimum of (1) 0.7 for Cronbach's Alpha [55]; (2) 0.7 for composite reliability [56]; and (3) 0.5 for Average Variance Extracted (AVE) [57]. Our data demonstrate compliance with these criteria. Additionally, we conducted a factor analysis for each construct. All item loadings were above the .5 threshold with eigenvalues above 1.0 indicating good convergent validity. There were no correlations above .9, indicating that multicollinearity is not an issue [58].

To examine hypotheses H1 and H2, we used the correlations of several factors with the *Time* needed to complete a task and the *Accuracy* of the results; these are shown in table 4. In H1 we hypothesized that persons with a higher rating in *Comp. SE* will (H1a) be more accurate and (H1b) finish tasks faster when using SEMAT. We were not in a position to demonstrate these relationships, therefore, both hypotheses were rejected. However, this may indicate that *Comp. SE* only has a minor influence on benchmarking performance using SEMAT, which is desirable and can be assisted with further improvements regarding usability.

Table 3. Descriptive Statistics and Reliability Coefficients

	Group	Items	Mean	SD	Alpha	Comp. Rel.	AVE
Comp. SE	Manual	4	6.205	.706	.827	.895	.682
	SEMAT		5.846	.681			
BPM Knowledge	Manual	4	4.522	1.186	.872	.912	.723
	SEMAT		4.885	1.368			
Perc. Usef.	Manual	5	4.900	.939	.832	.885	.609
	SEMAT		5.508	.755			
Perc. EoU	Manual	7	5.571	.774	.872	.909	.592
	SEMAT		5.956	.770			
Satisfaction	Manual	4	4.975	1.204	.895	.928	.763
	SEMAT		5.558	.693			

Table 4. Correlations of various constructs with *time* and *accuracy*

Construct	Group	Comp.SE	BPM Knowledge		Perc. EoU	Perc. Usef.
		SEMAT	SEMAT	Manual	SEMAT	SEMAT
Time Total	Corr.	.068	-.023	-.159	-.159	-.230
	Sign.	.787	.928	.528	.528	.365
Accuracy Total	Corr.	-.292	-.309	.654	.358	.092
	Sign.	.224	.197	.008	.130	.703

Equally, we were unable to support H2, which indicated that people with more *BPM Knowledge* will finish the tasks faster and more accurately. Only H2a could be supported by a strong correlation ($p < .01$). Accordingly, these results suggest that *BPM Knowledge* only plays a minor role when using SEMAT, and therefore a lower degree of knowledge is necessary compared to performing manual benchmarking, where a strong correlation is present. While this is desirable as well, it implies that our tool – or BPM software in general – should assist the user wherever possible and mitigate the probability of human error.

Further, in H3 we hypothesized that persons using SEMAT will be more accurate and faster than those who did not. The descriptive statistics for *Time* and *Accuracy* shown in table 5 indicate that in fact, SEMAT users are on average 14.3% faster while scoring 6.7% higher. These results imply that SEMAT – and perhaps other BPM tools – increase efficiency and the quality of benchmarking results. We expect a greater difference between mean values with increasing task complexity. However, a T-test revealed that these differences in mean-values are not statistically significant and therefore H3 was rejected.

In H4 and H5 we hypothesized that, when using SEMAT, Time and Accuracy are positively related with the Perc. EoU and Perc. Usef. To test these hypotheses, we examined the correlations of said constructs; the results can be reviewed in table 4. Although the correlations imply that the hypotheses are in fact true, these correlations are not statistically significant. Subsequently, these hypotheses were rejected as well.

Table 5. Descriptive statistics for *Time* and *Accuracy*

	Group	Time Task 1	Time Task 2	Time Task 3	Time Total	Accuracy Task 1	Accuracy Task 2	Accuracy Task 3	Accuracy Total
Mean	Manual	396.400	209.930	675.408	1255.232	5.182	8.455	18.455	32.091
	SEMAT	309.454	177.132	606.720	1076.300	5.846	8.923	19.615	34.385
SD	Manual	119.604	42.641	121.712	221.511	1.834	1.809	4.083	6.156
	SEMAT	130.623	61.341	137.136	264.322	0.555	0.277	3.841	4.407

Table 6. Correlations among the constructs *Perc. Usef.*, *Perc. EoU* and *Satisfaction* for persons that used SEMAT for the analysis

		Perc. EoU	Satisfaction
Perc. Usef. SEMAT	Corr.	.622	.823
	Sign.	.004	.000
Perc. EoU SEMAT	Corr.	1.000	.677
	Sign.	-	.002

Further, we hypothesized that people’s *Perc. EoU* will have a positive effect on *Perc. Usef.* (H6). This hypothesis was tested by means of an investigation into the constructs correlations. As can be seen in table 6, there is a high positive correlation, with a statistical significance at the .005-level. Therefore, this hypothesis can be accepted. The last hypothesis suggests that a person’s *Satisfaction* will be positively related with the *Perc. Usef.* (H7a) and *Perc. EoU* (H7b), provided they used SEMAT for the analysis. Again, table 6 shows the correlations and their significance used for testing these hypotheses. Hypothesis H7a is supported by the strong correlation between the constructs *Perc. Usef.* and *Satisfaction* (.823; $p < .001$) and hypotheses H7b by the correlation between *Perc. EoU* and *Satisfaction* (.677; $p < .005$).

We used regression analysis to calculate the R^2 values. As to be expected from the previously reported results, most R^2 values were unsatisfactory and/or statistically insignificant. However, we found an R^2 value of .598 ($p < .01$) with *Perc. EoU* as the independent and *Perc. Usef.* as the dependent variable and an R^2 value of .839 ($p < .001$) with *Perc. EoU* and *Perc. Usef.* as the independent variables and *Satisfaction* as the dependent variable.

Additionally, we gathered the participants’ opinions on SEMAT. The participants found that the software was easy to understand (38.5%), very helpful to determining process information (30.8%) and helped them to concentrate on what is important to accomplish the tasks (23.1%). On the other hand, they indicated that they would appreciate more extensive support with aggregating process information (30.8%) and automated checks whether the entered information is valid (23.1%).

6 Conclusion and Future Research

In this paper, we describe an approach for benchmarking sustainability-related metrics using specially engineered software. We show our approach to evaluate SEMAT in an

experiment with 24 participants and share the gathered data. Regarding the population and size of the sample, even though all participants had training in BPM, it would have been beneficial to target BPM professionals in a natural use-setting with our experiment. While a sample size of 24 is sufficient to make statistically relevant statements, a bigger sample always improves statistical power [59]. Another way to improve the validity of the results would be to add more levels of task complexity. Also, in future research it would be beneficial to draw comparisons with other established benchmarking approaches.

SEMAT is helpful to improve sustainability efforts, but still possesses some issues to address, one of them being the accurate handling of measurements. Even if metrics are applied consistently, some organizations may capture, e.g. fuel efficiency in miles per gallon whereas others may capture it in liters per 100 kilometers. To ensure correct benchmarking results, these measurements should automatically be translated without introducing the risk of human error. This can be achieved by another ontology that is not only able to perform the aforementioned calculations but also to, e.g., translate any greenhouse gas into CO₂ equivalents.

The creation of the domain and base ontology itself is rather time-consuming. Therefore, ways need to be found to accelerate these processes and make them less prone to human error. This can be achieved by maintaining a centralized base ontology that is appropriate for a multitude of notations and languages. This paper only examines the benefits of querying the analysis ontology, which is only one out of six steps in the methodology. Once the mentioned improvements are implemented, the entire benchmarking process should be evaluated. We show that, on average, benchmarking efforts using our tool are less time-consuming than doing the work manually, although the differences are not significant enough and should therefore be repeated, addressing the issues mentioned above.

Also, our research has several implications for the creation of future (Green) BPM tools. Currently, there is no feature implemented that can tell whether, e.g., a process that produces fewer emissions (metric A) is preferable to a process that produces less noise (metric B) (i.e. allowing preference rankings between different metrics A and B). Such a preference ranking varies in each organization and location and must currently be decided manually. In future, Green BPM software enabling this decision support and containing appropriate formulae and thresholds for preference rankings could be implemented. We also show that *Perc. Usef.* and *Perc. EoU* have a significant impact on *Satisfaction*. Therefore, developers of (green) BPM software should focus on creating easy to use, well documented software that does not distract the users with unneeded information. Furthermore, this software should be easily accessible for less experienced analysts. The variety of metrics that can be applied to Green BPM shows that software tools must exhibit a certain degree of flexibility in regard to what information can be annotated to process models.

One of the shortcomings of any BPM measure, the rather time-consuming collection of process information, could be addressed by integrating various data sources. Some of which may be used to generate and/or update process performance information, further increasing efficiency of the benchmarking approach.

Acknowledgements

The authors would like to thank the participants in the experiment as well as the other project members, specifically Ms. Marita Imhorst and Mr. Martin Kluth, who provided valuable insights, help and substantive feedback during the research process.

This work is part of the project IT-for-Green (Next Generation CEMIS for Environmental, Energy and Resource Management). The IT-for-Green project is funded by the European regional development fund (grant number W/A III 80119242).

References

1. Seidel, S., Vom Brocke, J., Recker, J.: Call for Action: Investigating the Role of Business Process Management in Green IS. *Sprouts* 11 (4), 1–7 (2011)
2. Watson, G.H.: *Strategic Benchmarking - How To Rate Your Company's Performance Against The World's Best*. Wiley and Sons, New York (1993)
3. Drew, S.A.W.: From Knowledge to Action: The Impact of Benchmarking on Organizational Performance. In: *Long Range Planning* 30 (3), 427–441 (1997)
4. Scheer, A.-W.: *ARIS - Business Process Modeling*. Springer, Berlin (2000)
5. Object Management Group: *Business Process Model and Notation (BPMN) - Version 2.0*
6. Teuteberg, F., Kluth, M., Smolnik, S., Ahlemann, F.: Semantic Benchmarking of Process Models - An Ontology-Based Approach. In: *ICIS 2009 Proceedings*, paper 89. AIS (2009)
7. Markovic, I., Hasibether, F., Jain, S., Stojanovic, N.: Process-Oriented Semantic Business Modeling. In: *9th Wirtschaftsinformatik 2009*, pp. 683-693 (2009)
8. Thomas, O., Fellmann, M.: Semantic Business Process Management: Ontology-Based Process Modeling Using Event-Driven Process Chains. *International Journal of Interoperability in Business Information Systems* 2, 29–44 (2007)
9. Höfferer, P.: Achieving Business Process Model Interoperability Using Metamodels and Ontologies. In: *European Conference on Information Systems St. Gallen* (2007)
10. Hoesch-Klohe, K., Ghose, A., Lê, L.-S.: Towards Green Business Process Management. In: *IEEE International Conference on Services Computing*, pp. 386–393 (2010)
11. Hepp, M., Roman, D.: An Ontology Framework for Semantic Business Process Management. In: *Proceedings of Wirtschaftsinformatik 2007 Karlsruhe* (2007)
12. Hepp, M., Leymann, F., Bussler, C., Domingue, J., Wahler, A., Fensel, D.: Semantic Business Process Management: A Vision Towards Using Semantic Web Services for Business Process Management. In: *IEEE International Conference on e-Business Engineering*, pp. 535–540 (2005)
13. Thomas, O., Fellmann, M.: Semantic Process Modeling - Design and Implementation of an Ontology-Based Representation of Business Processes. *Business Information Systems Engineering* 2, 15–28 (2009)
14. International Organization for Standardization: *ISO 14040: Environmental Management - Life Cycle Assessment - Principles and Framework*, Geneva (2006)
15. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly* 26, 13–23 (2002)
16. Levy, Y., Ellis, T.J.: A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science: International Journal of an Emerging Transdiscipline* 9, 181–212 (2006)

17. Ghose, A., Hoesch-Klohe, K., Hinsche, L., Le, L.-S.: Green Business Process Management: A Research Agenda. *Australasian Journal of Information Systems* 16, 103–117 (2009)
18. Hoesch-Klohe, K., Ghose, A.: Carbon-Aware Business Process Design in Abnoba. In: LNCS, Vol. 6470, pp. 551–556. Springer (2010)
19. Hinge, K., Ghose, A., Koliadis, G.: Process Seer: A Tool for Semantic Effect Annotation of Business Process Models. In: IEEE International EDOC Conference Auckland (2009)
20. Cleven, A., Winter, R., Wortmann, F.: Process Performance Management as a Basic Concept for Sustainable Business Process Management – Empirical Investigation and Research Agenda. In: LNBIP, Vol. 66, pp. 479–488. Springer (2011)
21. Houy, C., Reiter, M., Fettke, P., Loos, P., Geb, D.: Towards Green BPM – Sustainability and Resource Efficiency through Business Process Management. In: LNBIP, Vol. 66, pp. 501–510. Springer (2011)
22. Becker, J., Weiß, B., Winkelmann, A.: Automatic Identification of Structural Process Weaknesses – Experiences with Semantic Business Process Modeling in the Financial Sector. In: *Wirtschaftsinformatik, Zürich* (2011)
23. Becker, J., Räckers, M., Winkelmann, A.: Pattern-Based Semi-Automatic Analysis of Weaknesses in Semantic Business Process Models in the Banking Sector. In: *European Conference on Information Systems, Pretoria* (2010)
24. Becker, J., Bergener, P., Delfmann, P., Eggert, M., Weiß, B.: Supporting Business Process Compliance in Financial Institutions – A Model-Driven Approach. In: *Wirtschaftsinformatik* (2011)
25. Nissen, M.E.: Redesigning Reengineering Through Measurement-Driven Inference. *MIS Quarterly* 22, 509–535 (1998)
26. Nowak, A., Leymann, F., Mietzner, R.: Towards Green Business Process Reengineering. In: LNCS, Vol. 6568, pp.187–192. Springer (2011)
27. Loos, P., Nebel, W., Marx Gómez, J., Hasan, H., Watson, R.T., vom Brocke, J., Seidel, S., Recker, J.: Green IT: A Matter of Business and Information Systems Engineering?. *Business & Information Systems Engineering* 3 (4), 245–252 (2011)
28. Recker, J., Rosemann, M., Gohar, E.R.: Measuring the Carbon Footprint of Business Processes. In: LNBIP, Vol. 66, pp. 511–520. Springer (2011)
29. Kundisch, D., Herrmann, P., Meier, C.: Sustainable Process Management - Status Quo and Perspectives. In: LNBIP, Vol. 46, pp. 94–106. Springer (2010)
30. Recker, J.: Towards an Understanding of Process Model Quality. *Methodological Considerations*. In: *European Conference on Information Systems, Göteborg* (2006)
31. Legner, C., Wende, K.: The Challenges of Inter-Organizational Business Process Design - A Research Agenda. In: *European Conference on Information Systems, St. Gallen* (2007)
32. Heinrich, B., Bolsinger, M., Bewernik, M.: Automated Planning of Process Models : The Constructions of Exclusive Choices. In: *International Conference on Information Systems, Phoenix* (2009)
33. Becker, J., Weiß, B., Winkelmann, A.: Transferring a Domain-Specific Semantic Process Modeling Language: Findings from Action Research in the Banking Sector. In: *European Conference on Information Systems, Pretoria* (2010)
34. Bi, H., Zhao, L.: Process Logic for Verifying the Correctness of Business Process Models. In: *International Conference on Information Systems, Turku* (2004)
35. De Medeiros, A., van der Aalst, W.M.P., Karla, A., Pedrinaci, C.: Semantic Process Mining Tools: Core Building Blocks. In: *European Conference on IS* (2008)
36. Hua, Z., Zhao, J.L., Storey, V.C.: Exploring a Domain Ontology Based Approach to Business Process Design. In: *International Conference on IS, St. Louis* (2010)

37. Otto, B., Wäsch, J.: A Model for Inter-Organizational Business Process Integration. In: *Wirtschaftsinformatik, Dresden* (2003)
38. Simon, C., Mendling, J.: Integration of Conceptual Process Models by the Example of Event-driven Process Chains. In: *Wirtschaftsinformatik, Karlsruhe* (2007)
39. Hailemariam, G., vom Brocke, J.: What Is Sustainability in Business Process Management? A Theoretical Framework and Its Application in the Public Sector of Ethiopia. In: *LNBIP, Vol. 66*, pp. 489–500. Springer (2011)
40. Alemayehu, W., Vom Brocke, J.: Sustainability Performance Measurement – The Case of Ethiopian Airlines. In: *LNBIP. 66*, pp. 467–478. Springer (2011)
41. Joschko, P., Page, B., Wohlgemuth, V.: Combination of job oriented simulation with ecological material flow analysis as integrated analysis tool for business production processes. In: *Proceedings of the 2009 Winter Simulation Conference*, pp 1456–1465. IEEE (2009)
42. Teuteberg, F., Kluth, M., Ahlemann, F., Smolnik, S.: Semantic Process Benchmarking to Improve Process Performance. *Benchmarking: An International Journal* (2013)
43. Global Reporting Initiative: Sustainability Reporting Guidelines 3.1. Global Reporting Initiative (2011)
44. De Haes, U.: Towards a Methodology for Life Cycle Impact Assessment. SETAC-Europe, Brussels (1996)
45. Jolliet, O., Margni, M., Charles, R., Humbert, S., Payet, J., Rebitzer, G., Rosenbaum, R.: IMPACT 2002+: A New Life Cycle Impact Assessment Methodology. *The International Journal of Life Cycle Assessment* 8, 324–330 (2003)
46. Hausmann, M.: Sustainability of the CEWE PHOTOBOOK. In: *International Conference on Digital Printing Technologies and Digital Fabrication*, pp. 736–740 (2011)
47. APQC: APQC Process Classification Framework (PCF)
48. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF - W3C Recommendation
49. Bandura, A.: Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Advances in Behaviour Research and Therapy* 1, 139–161 (1978)
50. Higgins, C.A., Compeau, D.R.: Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Quarterly* 19, 189–211 (1995)
51. Goodhue, D.L.: Understanding User Evaluations of Information Systems. *Management Science* 41, 1827–1844 (1995)
52. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 319–340 (1989)
53. Oliver, R.: Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *Journal of Applied Psychology* 62, 480–486 (1977)
54. Oliver, R.: A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research* 1, 460–469 (1980)
55. Kline, P.: *The Handbook of Psychological Testing*. Routledge, London (1998)
56. Hair, J.F.J., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L.: *Multivariate Data Analysis*. Pearson Prentice Hall, Upper Saddle River (2006)
57. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18, 39–50 (1981)
58. Burns, R.B., Burns, R.A.: *Business Research Methods and Statistics using SPSS*. SAGE Publications, London (2008)
59. Cohen, J.: *Statistical Power Analysis for the Behavioural Sciences*. Lawrence Erlbaum, New York (1988)

Nachhaltigkeit in IT-Organisationen - Ein Forschungsrahmen für das Nachhaltige Informationsmanagement

Koray Ereğ, Fabian Löser, und Rüdiger Zarnekow

Technische Universität Berlin, Fachgebiet IuK-Management, Berlin, Germany
{koray.erek, f.loeser, ruediger.zarnekow}@tu-berlin.de

Abstract. Das Leitbild einer nachhaltigen Entwicklung hat sich als ein wichtiges Grundprinzip unternehmerischen Wirtschaftens etabliert. Die integrative Berücksichtigung von ökonomischen, ökologischen und sozialen Aspekten in den Wertschöpfungsstrukturen von Organisationen führte zu dem, was heute unter Nachhaltigkeitsmanagement in Unternehmen verstanden wird. Steigende Energieverbräuche in der Betriebs- und Nutzungsphase von Informationstechnologien einerseits und die immer wichtiger werdende Rolle von Informationssystemen als Enabler nachhaltiger Geschäftsprozesse in Unternehmen andererseits haben Nachhaltigkeitsthemen stärker in das Blickfeld der Wirtschaftsinformatik gerückt und deuten auf einen Paradigmenwechsel hin. Nach wie vor mangelt es jedoch an einer theoretischen und konzeptionellen Grundlage, welche diese beiden Sichtweisen sinnvoll integriert, um hieraus entsprechende Forschungs- und Handlungsfelder zu identifizieren. Dieser Beitrag adressiert diese Forschungslücke und stellt mit dem Modell eines Nachhaltigen Informationsmanagements einen Strukturierungsrahmen zur ganzheitlichen Adaption von Nachhaltigkeit in Forschung und Praxis der Wirtschaftsinformatik zur Verfügung. Das entwickelte Framework stellt das Ergebnis einer mehrjährigen Forschung in diesem Themengebiet dar.

Keywords: Nachhaltigkeit, Green IT, Green IS, Informationsmanagement

1 Ausgangssituation und Problemstellung

Die Bedeutung von Informationstechnologien (IT)¹ für Unternehmen nimmt kontinuierlich zu. Nahezu jeder Geschäftsprozess wird heutzutage IT-gestützt ausgeführt [1]. Dieser branchenübergreifende Effekt wird u. a. durch eine zunehmende Internationalisierung und Digitalisierung der Wertschöpfungsstrukturen von Unternehmen weiterhin verstärkt [2]. Mit der globalen Vernetzung und der damit einhergehenden starken IT-Durchdringung in den Unternehmensprozessen nehmen aber auch gleichzeitig der Einfluss und die Bedeutung der IT auf Ökonomie, Ökologie und Gesellschaft stetig

¹ Die Abkürzung „IT“ wird in diesem Beitrag als Sammelbegriff für alle Informations- und Kommunikationstechnologien verwendet

zu. Die ökologischen Herausforderungen in der IT werden seit einigen Jahren in der Öffentlichkeit und Wissenschaft unter dem Begriff „Green IT“ diskutiert. Sie ergeben sich insbesondere aufgrund des rasanten Anstiegs des Energieverbrauchs von IT-Systemen und der zum Betrieb notwendigen Infrastruktur durch immer komplexere Geschäftsanwendungen [3-4].

Demgegenüber sehen sich IT-Organisationen mit gekürzten IT-Budgets sowie der Forderung der Unternehmensleitung konfrontiert, den Wertbeitrag der IT zur Nachhaltigkeitsstrategie des Unternehmens darzulegen und messbar zu machen [5]. Insofern bezieht sich Nachhaltigkeit im Informationsmanagement nicht nur auf Green IT-Aspekte, sondern muss auch die Potenziale von Informationssystemen (IS) zur Unterstützung nachhaltiger Prinzipien in den Geschäftsprozessen und -produkten berücksichtigen (Green IS). Allerdings konzentrieren sich aktuelle Nachhaltigkeitsmaßnahmen von IT-Organisationen lediglich auf isolierte und in weiten Teilen unkoordinierte technische (Energie-)Effizienzmaßnahmen mit verengtem Blick auf Kosteneinsparungen (Bottom-Up-Vorgehen). Diese Herangehensweise greift jedoch langfristig zu kurz und muss durch ein ganzheitliches Vorgehen mit strategischen Zielvorgaben für ein nachhaltiges Management, welches die unterschiedlichen Sichtweisen integrativ berücksichtigt, ergänzt werden (Top-Down-Vorgehen). Die Berücksichtigung von Nachhaltigkeitsaspekten mit dem Ziel der Gestaltung und Umsetzung eines nachhaltigen IT-Managementsystems stellt daher für viele Organisationen nach wie vor eine große Herausforderung dar. Demzufolge liegt es nahe, die Frage aufzugreifen, welche Anstöße der Nachhaltigkeitsansatz für die zukünftige Forschung in der Wirtschaftsinformatik (WI) leisten kann und wie dieser effizient umgesetzt werden kann. Vor diesem Hintergrund setzt sich dieser Beitrag zum Ziel, einen Forschungsrahmen für Nachhaltigkeit in der WI abzuleiten. Im Fokus steht dabei die Erarbeitung eines praxisnahen Modells für ein Nachhaltiges Informationsmanagement (NIM), welches eine ganzheitliche Umsetzung und Steuerung des Leitbildes der Nachhaltigkeit im Informationsmanagement gewährleistet. Daher gelangen wir zu folgenden Forschungsfragen, welche durch diesen Beitrag adressiert werden:

- Wie ist der Status quo der Nachhaltigkeitsorientierung in der Praxis zu beurteilen und welche Anforderungen lassen sich hieraus für das NIM ableiten?
- Wie kann Nachhaltigkeit im Informationsmanagement ganzheitlich umgesetzt werden und welche Forschungsfelder können identifiziert werden?

Ziel des Beitrags ist es somit, der Forschung und Praxis einen Strukturierungsrahmen für Nachhaltigkeit zur Verfügung zu stellen, aktuelle Defizite im Umsetzungsvorgehen in der Praxis aufzuzeigen und relevante Forschungsfelder zur Nachhaltigkeit für die WI zu identifizieren. Zu diesem Zweck erfolgt in Kapitel zwei eine definitorische Abgrenzung der relevanten Begriffe und Konstrukte. Zudem wird ein Überblick über den aktuellen Forschungsstand gegeben. Daraufhin werden in Kapitel drei das zugrundeliegende Forschungsdesign beschrieben sowie darauf aufbauend die Anforderungen an das NIM abgeleitet. In Kapitel vier wird das Modell eines NIM zunächst gesamthaft vorgestellt und anschließend die identifizierten Forschungs- und Handlungsfelder (HF) anwendungsnah erläutert. Der Beitrag schließt mit einem Fazit und einem Ausblick.

2 Grundlagen und Stand der Forschung

2.1 Nachhaltigkeitsmanagement und Nachhaltiges Informationsmanagement

Managementansätze im Bereich der Nachhaltigkeit basieren auf dem politisch-gesellschaftlich getriebenen Drei-Säulen-Modell der Nachhaltigkeit (engl. „Triple Bottom Line“) und gehen von der Annahme aus, dass nachhaltigkeitsorientiertes Wirtschaften neben ökonomischen, auch ökologische und soziale Aspekte beinhalten muss [6]. Durch die simultane Betrachtung aller drei Säulen der Nachhaltigkeit wird somit ein ganzheitliches Nachhaltigkeitsmanagement in Unternehmen ermöglicht, das die Interessen interner und externer Stakeholder in einem ausgeglichenen Maße berücksichtigt und somit eine dauerhafte Fortführung der Geschäftstätigkeit ermöglicht.

Diesem Ansatz folgend wird der vorliegenden Arbeit das folgende Grundverständnis eines nachhaltigen Informationsmanagements (NIM) zugrunde gelegt: *Das NIM adressiert eine adäquate Integration ökonomischer, ökologischer und sozialer Belange in das Management von IT-Organisationen mit dem Ziel der Gestaltung und Umsetzung eines ganzheitlichen Nachhaltigkeitsmanagements.* Das Ziel eines NIM ist es somit, das Grundkonzept der betrieblichen Nachhaltigkeit ganzheitlich und anwendungsorientiert in die Managementaktivitäten von IT-Organisationen zu integrieren, bestehende Ansätze des Informationsmanagements dahingehend weiterzuentwickeln und bedarfsorientiert neue Managementkonzepte einzuführen.

2.2 Green IT und Green IS

Seit einigen Jahren hat die IT-Industrie ihren Beitrag zum Umwelt- und Ressourcenschutz mit medienwirksamen Begriffen wie „Green IT“ [7-8], „Green Information Systems (Green IS)“ [9], „Environmental Sustainability of IT“ [10] oder auch „IT-for-Green“ [11] versehen. Diesen Nachhaltigkeitsdiskussionen liegen grundsätzlich zwei unterschiedliche Sichtweisen zugrunde [7], [10-12]: Einerseits wird die IT als Objekt des Umweltschutzes betrachtet, indem der zunehmende Energieverbrauch der betriebenen IT-Infrastruktur und somit die von der IT induzierten klimaschädlichen CO₂-Emissionen thematisiert werden („Green IT“). Andererseits wird im Rahmen von „IT-for-Green“ oder auch „Green Business“ der Beitrag von IS zur Unterstützung der Nachhaltigkeitsziele von Unternehmen betrachtet und folglich untersucht, inwiefern durch den Einsatz innovativer IS, wie z. B. intelligente (Echtzeit-)Steuerung von Logistikprozessen, die Umweltbelastungen in den (Kern-)Geschäftsprozessen des Unternehmens reduziert werden können [11]. Letzteres betont die Funktion der IT als Enabler zur Bewältigung der Nachhaltigkeitsherausforderungen in Unternehmen, die in Abhängigkeit der Branche unterschiedlich ausgeprägt sein kann [13].

Die Identifikation von Optimierungspotenzialen zur Steigerung der Energieeffizienz der eingesetzten Technologien nimmt derzeit eine zentrale Rolle hinsichtlich der Umsetzung von Green IT-Initiativen ein. Harmon und Auseklis zählen Kostenreduktions- und Performance-Ziele als treibende Motive einer Green IT-Adaption in IT-Organisationen und deuten daher Green IT als *„the practice of maximizing the efficient use of computing resources to minimize environmental impact“* und bestäti-

gen, dass diese Betrachtungsweise nur einen Teilbereich des übergeordneten Leitbilds der Nachhaltigkeit repräsentiert [14]. Watson et al. [9] bemängeln den rein technologiefokussierten Gedanken einer Green IT. Sie ziehen die Terminologie Green IS vor und argumentieren, dass die Rolle der IT als Enabler nachhaltiger Geschäftsprozesse in Unternehmen („IT-for-Green“) einen zentralen Bestandteil der Nachhaltigkeitsdiskussionen in der IT bilden muss. Dementsprechend legt Elliot [10] beide Sichtweisen zugrunde, indem er ökologische Nachhaltigkeit in der IT als „[...] activities to minimize the negative impacts and maximize the positive impacts of human behavior on the environment through the design, production, application, operation, and disposal of IT and IT-enabled products and services throughout their life cycle“ definiert [10]. Diese Definition, welche Green IT und Green IS integriert, entspricht dem Verständnis des NIM in dieser Arbeit.

2.3 Umweltauswirkungen der IT

Die strategischen Anforderungen an eine IT-Organisation und die Bereitstellung von IT-Ressourcen hängen wesentlich davon ab, welche Rolle der IT im Unternehmen beigemessen wird: Ist IT ein strategischer Erfolgsfaktor für die Differenzierung im Wettbewerb oder hat IT eine reine Unterstützungsfunktion mit dem Ziel, effizientere Geschäftsprozesse zu ermöglichen [15]? Diese grundsätzliche Rollenfestlegung ist sowohl in ökonomischer als auch ökologischer Hinsicht von strategischer Bedeutung und determiniert die Leitlinien der IT-Nachhaltigkeitsstrategie. Der IT-Durchdringungsgrad in den Geschäftsprozessen eines Unternehmens, die Bedeutung für das Kerngeschäft sowie die Verwendung von IT in den Endprodukten des Unternehmens bemessen hierbei die Höhe der IT-Ausgaben. Darüber hinaus reflektiert die Bedeutung der IT für das Kerngeschäft des Unternehmens das Ausmaß der IT-induzierten CO₂-Emissionen. So wird der IT in informationsintensiven Branchen (Finanzdienstleistungen, Medien, Bildung etc.) eine höhere Bedeutung beigemessen als bspw. in energieintensiven Branchen (industrielle Fertigung, Chemie etc.), in der sie primär eine geschäftsunterstützende Funktion einnimmt [13]. Demzufolge ist es notwendig und sinnvoll, eine Differenzierung bezüglich der IT-bedingten Umweltauswirkungen vorzunehmen: Die IT-Umweltauswirkungen ersten Grades („1st degree environmental impact“) beschreiben die direkten negativen Effekte, welche durch Produktion, Betrieb und Entsorgung von IT verursacht werden. Dazu zählen der Ressourceneinsatz für die Herstellung von IT-Hardware, der durch den Betrieb der IT-Infrastruktur entstehende Verbrauch elektrischer Energie, welche nicht regenerativ bzw. CO₂-neutral erzeugt wurde, sowie die negativen Folgen der Entstehung von Elektroschrott [16]. Dieser mit Green IT adressierte Themenkomplex umfasst sowohl den Betrieb von Rechenzentren (RZ) und die in der Büroumgebung (BU) verwendete IT als auch die IT-Beschaffung und -Entsorgung. In informationsintensiven Branchen ist dieser Anteil der direkten IT-bezogenen CO₂-Emissionen sehr hoch (Abbildung 1).

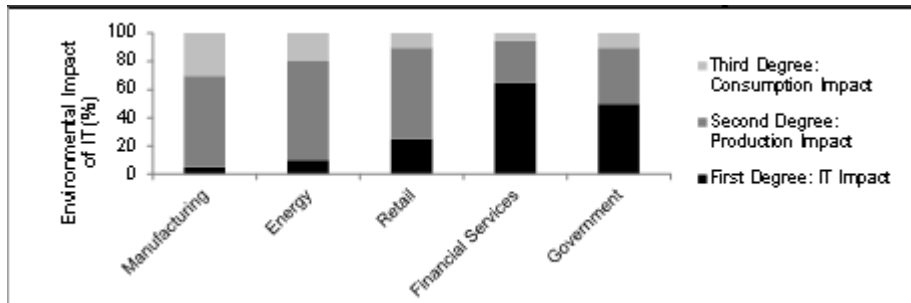


Abb. 1. Branchenspezifische Ausprägungen IT-bezogener Umweltauswirkungen [13]

Die IT-bezogenen Umweltauswirkungen zweiten Grades („2nd degree environmental impact“) beziehen sich auf die Produktions- und Geschäftsprozesse innerhalb des Unternehmens und werden durch „Green IS“ oder „IT-for-Green“ adressiert [9], [11], [17]. Diese spielen vor allem in der industriellen Produktion eine große Rolle. Im Gegensatz zu den Auswirkungen ersten Grades handelt es sich hierbei um positive Effekte, und zwar in Form von IT-unterstützten Effizienzsteigerungen in den internen Prozessen. Schließlich manifestieren sich im „3rd degree environmental impact“ die Umweltauswirkungen in der Nutzungsphase der Endprodukte bzw. der Dienstleistungen durch den Kunden. Diese Kategorie ist jedoch nur für Unternehmen relevant, bei denen die IT einen Bestandteil des Endproduktes bildet, wie z. B. Online-Banking [3].

3 Forschungsdesign und -beitrag

Dem Forschungsprozess dieser Arbeit liegt das Paradigma der angewandten Wissenschaft sowie der Forschungsrahmen des Business Engineering zugrunde [18]. Demzufolge beginnt der Forschungsprozess primär in der Praxis, ist auf die Entwicklung eines Artefakts (Modell eines NIM) gerichtet und endet mit der Ableitung von (Handlungs-)Empfehlungen für die Praxis und Wissenschaft. Ein wichtiges Kennzeichen der vorliegenden Forschungsarbeit ist das explorative Vorgehen, bei dem ein induktives Forschungsdesign gewählt wurde. Die vorherrschenden Methoden waren Desk-Research, Experteninterviews und Fallstudienforschung. Den im Rahmen dieser Arbeit gewonnenen Erkenntnissen liegt somit folgende empirische Datenbasis zugrunde, die im Zeitraum von 2009 bis 2011 geschaffen wurde [19-21]:

- Vorstudie anhand von 18 Experteninterviews mit CIOs und IT-Managern
- Erhebung von acht Fallstudien mit IT-Organisationen
- Durchführung einer schriftlichen Befragung mit 116 CIOs und IT-Managern

Als erster Schritt wurde ein systematischer Literatur-Review zu Green IT, zum betrieblichen Nachhaltigkeitsmanagement und zum Informationsmanagement durchgeführt. Ziel war es, ein theoretisches Vorverständnis zu erlangen und Konzepte im Bereich der unternehmerischen Nachhaltigkeit zu identifizieren.

3.1 Ergebnisse der Vorstudie

Im Anschluss an die Begutachtung der Literatur wurde eine explorative (Vor-)Studie zum Status quo einer Green IT durchgeführt, um Erkenntnisse über relevante Handlungsfelder (HF) einer Green IT in der Praxis zu gewinnen und Defizite im Umsetzungsvorgehen aufzudecken (Stand der Praxis). Dabei wurden 18 themenzentrierte Experteninterviews mit CIOs, IT-Managern sowie Umwelt- und Nachhaltigkeitsbeauftragten durchgeführt. Die Unternehmen, in denen die befragten Personen tätig waren, gehörten entweder der IT-Branche an oder waren in den Kerngeschäftsprozessen besonders stark mit IT durchdrungen. Alle Unternehmen verfügten seit geraumer Zeit über Lösungsansätze im Bereich Green IT bzw. Nachhaltigkeit. Durch die Verwendung eines strukturierten Interviewleitfadens wurden alle Experten zu den gleichen Themenfeldern befragt. Die Gespräche wurden aufgezeichnet, transkribiert und anschließend einer detaillierten Inhaltsanalyse unterzogen.

Die Analyse ergab, dass sich aktuelle Nachhaltigkeitsmaßnahmen von IT-Organisationen vorrangig auf isolierte und unkoordinierte technische Einzelmaßnahmen beschränken. Eine Integration der Initiativen in das Umwelt- oder Nachhaltigkeitsmanagementsystem des Unternehmens zur Sicherstellung eines ganzheitlichen Vorgehens ist bislang eher die Ausnahme. Insbesondere mangelte es nach wie vor an konkreten Zielgrößen zur Nachhaltigkeit in den IT-Organisationen. Ein etabliertes Nachhaltigkeitsmanagement, das die drei Säulen der Nachhaltigkeit sinnvoll integriert und eine Abgrenzung zu einer Green IT erkennen lässt, war in den IT-Organisationen daher nicht gegeben. Vielmehr standen häufig der Business Case und das öffentliche Interesse für das Unternehmen im Fokus aller Bemühungen.

3.2 Ergebnisse der explorativen Fallstudien

Die Fallstudie als empirische und praxisorientierte Forschungsmethode eignet sich insbesondere, um aktuelle und schwer abgrenzbare Phänomene kontextabhängig zu untersuchen, und ermöglicht eine Tiefenbetrachtung der Wirkungszusammenhänge in der Praxis [22]. Es werden mehrere Datenerhebungstechniken eingesetzt, um den Sachverhalt möglichst vollständig und aus unterschiedlichen Perspektiven zu erfassen [22-23]. Während Einzelfallstudien („single case“) ein (einzigartiges) Problem in einem Fall intensiv untersuchen, werden bei Mehrfachfallstudien („multiple case“) gleiche Sachverhalte und Problemstellungen in unterschiedlichen Umgebungen detailliert analysiert [23]. Letztere gelten als robuster, da sie einer Replikationslogik folgen.

Bei den in dieser Forschungsarbeit verwendeten Fallstudien handelt es sich um Mehrfachfallstudien, welche als geeignete Methode identifiziert wurde, um den State-of-the-Art von Green IT in unterschiedlichen IT-Organisationen zu untersuchen. Dazu wurden Unternehmen ausgewählt, die in ihrer Branche als führend im Bereich der Nachhaltigkeit angesehen werden. Um Unterschiede zwischen den verschiedenen Branchen sowie zwischen IT-Organisationen analysieren zu können, wurden IT-Dienstleister (IT-DL) unterschiedlicher Größe untersucht. Das Spektrum der betriebenen Arbeitsplatzsysteme reichte von 800 bis über 100.000 Desktop-PCs. Es wurden sowohl öffentliche IT-DL analysiert als auch privatwirtschaftliche, interne IT-DL von

Unternehmen der Software-, Medien-, Chemie-, Technologie- und Finanzdienstleistungsbranche (s. Tabelle 1).

Tabelle 1. Eckdaten zu den untersuchten Unternehmen (U) im Rahmen der Fallstudienanalyse

	U 1	U 2	U 3	U 4	U 5	U 6	U 7	U 8
Branche	IT-DL kommunale Verwaltung	Chemie	Medien	Technologie	Verkehrsbetrieb	Finanzdienstleistungen	Software	IT-DL des Bundes
Umsatz in Mio. € (2011)	107,61	1.103	3.190	4.200	148	33.228	14.233	k.A.
Mitarbeiter in (2011)	480	6.381	12.885	30.000	1.933	100.996	59.420	2.400
Anzahl RZ	1	3	5	26	2	18	26	2

Die Datenerhebung erfolgte anhand von semi-strukturierten Interviews mit Experten aus den jeweiligen Unternehmen auf Basis eines Leitfadens. Als Grundlage für die Erstellung des Interviewleitfadens diente der konzeptionelle Bezugsrahmen des integrierten Informationsmanagements (IIM). Das IIM stellt die zentralen Managementprozesse eines IT-DL auf Basis eines Source-Make-Deliver-Ansatzes dar, wobei die IT-Governance die Leistungserstellung strukturiert und regelt [1]. Das IIM-Modell betrachtet somit den gesamten Wertschöpfungsprozess, einschließlich der Schnittstellen zu Lieferanten und Kunden.

Um den State-of-the-Art von Green IT in den Fallstudien untersuchen zu können, wurden IT-Verantwortliche zu den Bereichen Einkauf, IT Operations sowie Vertriebs- und Kundenbeziehungsmanagement befragt. Dabei wurde detailliert untersucht, welche Bedeutung einer Green IT in den jeweiligen Bereichen beigemessen wird und welche Maßnahmen umgesetzt wurden. Darüber hinaus wurde analysiert, inwiefern die IT-Organisation einen Beitrag dazu leistet, die negativen Umweltauswirkungen des Unternehmens durch Reengineering von Produktions- und Geschäftsprozessen zu verringern („2nd degree“). Ferner wurde untersucht, ob die IT-Organisation durch „grüne“ Innovationen die ökologischen Eigenschaften der Endprodukte bzw. Dienstleistungen des Unternehmens verbessert („3rd degree“). Die Interviews wurden durch eine Sekundäranalyse von unternehmensinternen Projekt- und Betriebsunterlagen ergänzt. Sämtliche Gespräche wurden aufgenommen und anschließend transkribiert. Die dokumentierten Fallstudien wurden durch die befragten Interviewpartner anschließend validiert. Die daraus resultierenden Ergebnisse wurden für eine vergleichende Fallstudienanalyse („cross case analysis“) herangezogen, um gezielt die Ausprägung eines nachhaltigkeitsorientierten Informationsmanagements in den Unternehmen vergleichen und bewerten zu können (s. Tabelle 2).

Tabelle 2. Gegenüberstellung der Ergebnisse der explorativen Fallstudien

Umweltauswirkungen der IT	Wertschöpfungskette der IT-Organisation	Berücksichtigung von Nachhaltigkeitsaspekten und Umsetzung von Green-IT/IS-Maßnahmen	U1	U2	U3	U4	U5	U6	U7	U8	
Erster Grad	Governance	Nachhaltige IT Governance	◐	◑	◒	◓	◔	◕	◖	◗	◘
		Strategisches Alignment	◐	◑	◒	◓	◔	◕	◖	◗	◘
	Source	Lieferantenmanagement	◑	◒	◓	◔	◕	◖	◗	◘	◙
		Einkauf von IT Hardware und Dienstleistungen	◐	◑	◒	◓	◔	◕	◖	◗	◘
	Make	Produktionsstrategie	◐	◑	◒	◓	◔	◕	◖	◗	◘
		Portfolio nachhaltiger IT-Produkte und DL	◑	◒	◓	◔	◕	◖	◗	◘	◙
		IT-Systeme Büroumgebung	◐	◑	◒	◓	◔	◕	◖	◗	◘
		IT-Systeme RZ	◐	◑	◒	◓	◔	◕	◖	◗	◘
		Infrastruktur RZ	◐	◑	◒	◓	◔	◕	◖	◗	◘
		Monitoring des IT-Ressourcenverbrauchs	◑	◒	◓	◔	◕	◖	◗	◘	◙
	Deliver	Externe Kommunikation	◐	◑	◒	◓	◔	◕	◖	◗	◘
		Interne Kommunikation	◐	◑	◒	◓	◔	◕	◖	◗	◘
		Schulung von Mitarbeitern	◑	◒	◓	◔	◕	◖	◗	◘	◙
		Kooperationen (Netzwerke)	◑	◒	◓	◔	◕	◖	◗	◘	◙
Zweiter Grad	Geschäftsbereich des Unternehmens	Business Process Reengineering (BPR)	◑	◒	◓	◔	◕	◖	◗	◘	
		Nutzung von Umweltinformationssystemen	◑	◒	◓	◔	◕	◖	◗	◘	
		Interdisziplinäre Zusammenarbeit mit Abteilungen	◑	◒	◓	◔	◕	◖	◗	◘	
Dritter Grad	Externer Markt	Umweltfreundliche Endprodukte	◑	◒	◓	◔	◕	◖	◗	◘	
		Umweltfreundliche Dienstleistungen	◐	◑	◒	◓	◔	◕	◖	◗	◘

Die Ergebnisse des Fallstudienvergleichs zeigen, dass IT-Organisationen, die eine führende Rolle im betrieblichen Nachhaltigkeitsmanagement einnehmen, unterschiedliche Ansätze verfolgen. Den höchsten Umsetzungsgrad von Green IT findet man im RZ. Dies ist nicht erstaunlich, denn das RZ verzeichnet den größten Stromverbrauch und entsprechend lassen sich hier enorme (Kosten-)Einsparungen erzielen. Die wichtigste Erkenntnis ist allerdings, dass sich eine sehr große Anzahl und Vielfalt von Maßnahmen in den unterschiedlichen Wertschöpfungsstufen Govern, Source, Make und Deliver sowie im BPR und der Entwicklung innovativer, umweltfreundlicher Produkte und Dienstleistungen finden lassen – aber dennoch keines der Unternehmen alle Bereiche des NIM hinreichend und ganzheitlich adressiert.

Die Ergebnisse bestätigen zudem die Erkenntnis der Vorstudie, dass es bislang an Managementkonzepten fehlt, die ein ganzheitliches NIM unterstützen, indem sie die vielfältigen Handlungsfelder und Einsatzbereiche von Green IT und Green IS identifizieren und Empfehlungen für die Umsetzung entsprechender Maßnahmen aufzeigen. Die Quick-Wins im Bereich RZ wurden bereits von vielen Unternehmen umgesetzt – hingegen ist es den meisten IT-Verantwortlichen und -Managern unklar, wie sie durch ein NIM über die bloße Senkung operativer Kosten hinaus einen Beitrag zur Wettbewerbsdifferenzierung leisten können. Ergänzend zur Fallstudienenerhebung wurde eine breit angelegte schriftliche Befragung zur Relevanz und zum Umsetzungsstand einer Green IT durchgeführt, um die Validität der Ergebnisse zu erhöhen. Zudem wurden fortlaufend Diskussionen mit verschiedenen Expertengruppen aus Forschung und

Praxis geführt, um die Ergebnisse kritisch zu hinterfragen. Die aus den empirischen Untersuchungen gewonnenen Erkenntnisse wurden schließlich zur konzeptionellen und inhaltlichen Ausgestaltung des NIM-Modells verwendet.

4 Modell des Nachhaltigen Informationsmanagements (NIM)

Das nachfolgend vorgestellte Modell stellt eine Antwort auf die beschriebenen Herausforderungen dar, indem es eine anwendungsorientierte Integration der Nachhaltigkeit (Ökonomie, Ökologie und Soziales) sicherstellt, um eine Grundlage für die Weiterentwicklung bestehender und die Einführung neuer Managementkonzepte im Informationsmanagement zu schaffen. In Analogie zum IIM-Modell werden die drei Wertschöpfungsstufen Source (IT-Beschaffung), Make (IT-Produktion) und Deliver (IT-Vertrieb und -Kommunikation) unterschieden und, in Anlehnung an den Forschungsrahmen des Business Engineering, die drei Gestaltungsebenen Strategie (strategische Zielvorgaben), Prozesse (planerische Aufgaben) und operative Umsetzung zugrunde gelegt [1], [18]. Die nachhaltige IT-Governance ist dem Wertschöpfungsprozess der IT-Organisation übergeordnet und stellt eine grundlegende strategische Schnittstelle zum Gesamtunternehmen dar (s. Abbildung 2).

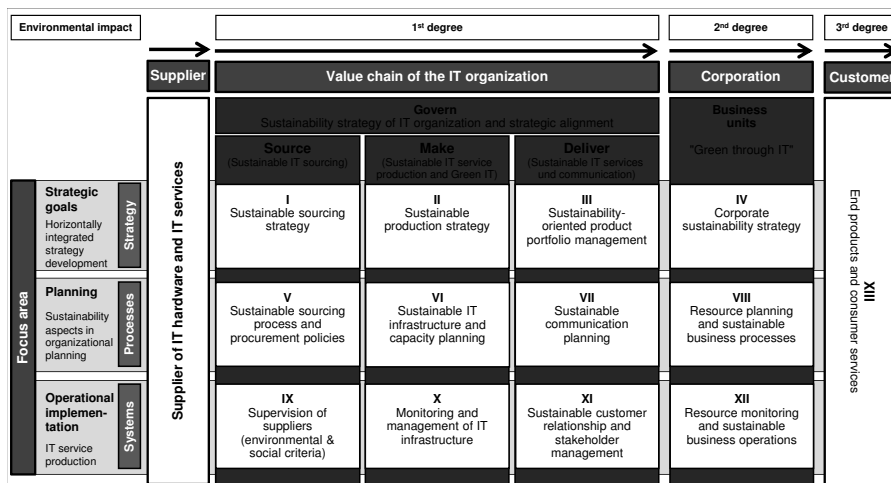


Abb. 2. Modell des Nachhaltigen Informationsmanagements (NIM)

Das NIM adressiert die mit der unternehmerischen Nachhaltigkeit in Zusammenhang stehenden Handlungsfelder innerhalb der IT-Organisationen, die IT-gestützten Geschäftsprozesse des Unternehmens sowie – für den Fall dass IT ein Bestandteil dieser ist – die am Markt offerierten Endprodukte und Dienstleistungen. Aus der Matrixstruktur ergeben sich somit 13 konkrete Handlungsfelder (HF I-XIII). Durch die Integration ökologischer, ökonomischer und sozialer Nachhaltigkeitskriterien in jedem dieser Handlungsfelder können Nachhaltigkeitsmaßnahmen identifiziert, geplant und implementiert werden – und zwar auf ganzheitlicher Basis unter Berücksichtigung strategischer Aspekte. Damit sollen IT-Organisationen und Wissenschaftler dabei

unterstützt werden, die in der Praxis und Forschung bislang vernachlässigten Handlungsfelder der Nachhaltigkeit zu analysieren, um somit die Grundlage für ein ganzheitliches und nachhaltiges Informationsmanagement mit einem breiten Spektrum an möglichen Maßnahmen zu schaffen. In den folgenden Sektionen werden die zentralen Gestaltungsebenen des NIM vorgestellt.

4.1 Nachhaltigkeit in den Wertschöpfungsprozessen der IT-Organisation

Die nachhaltige IT-Governance stellt die Schnittstelle zwischen dem Unternehmen und der IT-Organisation dar und ist für eine konsistente, strategische Nachhaltigkeitsausrichtung der IT-Organisation verantwortlich. In Anlehnung an die Definition des IT Governance Institute [24] hat sie zum Ziel, mittels Vorgaben und Leitlinien ein nachhaltiges Management der IT innerhalb der Organisation sicherzustellen, indem die Rahmenbedingungen für die darunterliegenden Gestaltungsebenen und Wertschöpfungsstufen des Informationsmanagements definiert, die nachhaltigkeitsbezogenen Risiken identifiziert und bewertet, das Ressourcenmanagement verantwortet und ein effektives Nachhaltigkeits-Controlling umgesetzt wird.

Im Rahmen der Beschaffung (Source) steht das nachhaltigkeitsorientierte Management der Lieferantenbeziehungen für den Einkauf von IT-Produkten (Hardware, Software und Services), die für die Leistungserstellung einer IT-Organisation benötigt werden, im Vordergrund. Die Beschaffungsstrategie bildet hierbei die Basis für die darunter liegenden Ebenen der Beschaffungsplanung und des operativen Beschaffungsmanagements. Ferner wird hier die Fertigungstiefe des IT-Leistungserbringers festgelegt, welche Leistungen durch die IT-Organisation erstellt werden und welche IT-Leistungen möglicherweise von einem externen Dienstleister fremdbezogen werden. Auf Prozessebene (HF V) werden konkrete Zielsetzungen für die Nachhaltigkeitskriterien des Beschaffungsprozesses definiert. Die nachhaltige Einkaufsplanung wird durch die im Lastenheft der IT-Services spezifizierten Nachhaltigkeitsanforderungen determiniert und dient als Basis für die Auswahl der Lieferanten. Auf operativer Ebene (HF IX) wird schließlich der Einkauf von nachhaltigen IT-Produkten durchgeführt und hinsichtlich der Vertragsbedingungen überwacht und evaluiert.

Die Produktionsstrategie (HF II) determiniert die internen Abläufe und den Output des IT-Leistungserstellungsprozesses und definiert somit die Rahmenbedingungen und generellen Leitlinien der IT-Leistungserbringung. Im Vordergrund steht der Endkundennutzen, also die effektive Erfüllung der Nachhaltigkeitsansprüche des Unternehmens an die IT-Organisation. Einen zentralen Aspekt innerhalb der Definition der Produktionsstrategie ist die Festlegung langfristiger Nachhaltigkeitsziele und Rahmenbedingungen für die Produktionsinfrastruktur. Neue zusätzliche Einflussfaktoren sind z. B. die technische und organisatorische Umsetzung von Virtualisierung und Cloud Computing oder der Umgang mit der stark steigenden Leistungsdichte der IT-Systeme. Neben den bekannten ökonomischen und technischen Anforderungen müssen die Komponenten der IT-Hardware, die Stromversorgung/-verteilung und Klimatisierung in den RZ sowie Netzwerke, Drucker, Speicher- und Arbeitsplatzsysteme in der Büroumgebung (BU) kritisch hinsichtlich der festgelegten Nachhaltigkeitskriterien analysiert werden. Im Rahmen der Produktionsplanung (HF VI) bietet vor allem

die Erhöhung der Energie- und Ressourceneffizienz im RZ und in der BU ein großes Potenzial. In diesem Kontext sind konkrete Produktionskapazitäten festzulegen, wobei eine Abwägung zwischen Anschaffungskosten und Betriebskosten sowie der Umweltverträglichkeit erfolgen muss. Der erforderliche Ressourceneinsatz sollte mit den Anforderungen im Rahmen der Leistungsplanung abgeglichen und mit den ökologischen Zielen der Nachhaltigkeitsstrategie in Einklang gebracht werden. Um Optimierungspotenziale im RZ identifizieren zu können, sollten bestimmte Energieeffizienzkennzahlen festgelegt werden. Die operative Steuerung des Produktionsprozesses (HF X) muss die Wirksamkeit der implementierten Nachhaltigkeitsmaßnahmen überprüfen und deren Effizienz in Bezug auf Wirtschaftlichkeit und Umwelt sicherstellen. Die Steuerung der Kapazitätsauslastung und die Lastverteilung im RZ im Rahmen des IT-Leistungserstellungsprozesses basieren auf einem kontinuierlichen Monitoring der Energiebedarfe von Serversystemen, Netzwerkkomponenten und Kühlung, wodurch die Analyse der festgelegten Energieeffizienzkennzahlen ermöglicht wird. Das Monitoring der Ressourcenbedarfe und die dynamische Steuerung der Kapazitäten ermöglichen eine deutliche Reduktion des Stromverbrauchs und der operativen Kosten.

In Rahmen des Deliver-Prozesses werden die Geschäftsbeziehungen zwischen der IT-Organisation und den Geschäftsbereichen definiert und der Vertrieb nachhaltiger IT-Services gesteuert. Die vom Kunden (i.d.R. Fachbereiche) geforderten Nachhaltigkeitsmerkmale müssen in technische Nachhaltigkeitsanforderungen der IT-Produkt- und Serviceerstellung transformiert werden. Ziel der nachhaltigen Vertriebs- und Kommunikationsstrategie (HF III) ist somit die aktive Gestaltung und Positionierung eines IT-Service-Portfolios unter Berücksichtigung von Nachhaltigkeitskriterien sowie die entsprechende Ausgestaltung des Marketing-Mixes. Die IT-Organisation kann sich für eine energie- und kosteneffiziente strategische Ausrichtung entscheiden, welche zu gesteigerter Nachhaltigkeit der Services bei niedrigen Kosten führt, oder für eine Differenzierungsstrategie, bei der sich die IT-Services durch ihre besonders ausgeprägten Nachhaltigkeitseigenschaften von den Produkten der Wettbewerber unterscheiden, wobei die gesteigerten Produktionskosten durch Premium-Preise kompensiert werden können. Im Rahmen der Planung (HF VII) sollten Kenngrößen definiert werden, welche einerseits die Anforderungen des Kunden in konkrete Zielvorgaben transformieren und andererseits die Leistungserstellung der IT-Organisation hinsichtlich der Nachhaltigkeitsaspekte transparent machen. Die Kommunikation der implementierten Nachhaltigkeitsmaßnahmen und der reduzierten CO₂-Emissionen spielen eine gewichtige Rolle. Zum einen wird durch die Wahrnehmung der Nachhaltigkeitseigenschaften der IT-Produkte für den Leistungsabnehmer ein zusätzlicher Mehrwert geschaffen, da dieses Engagement gegenüber internen und externen Stakeholdern kommuniziert werden kann. Andererseits haben die Anwender, welche die IT-Services im Rahmen ihrer beruflichen Tätigkeit nutzen, einen maßgeblichen Einfluss auf den Stromverbrauch – denn durch einen bewussten, verantwortlichen Umgang mit IT-Ressourcen können enorme Stromeinsparungen erzielt werden. Schließlich ist zur Erfüllung der Kundenanforderungen eine nachhaltigkeitsorientierte Steuerung des IT-Service-Portfolios erforderlich, die eine kontinuierliche Überwachung und Anpassung der konkreten Nachhaltigkeitskriterien und -ziele der IT-

Leistungen voraussetzt (HF XI). Dies betrifft bspw. auch die Integration von Nachhaltigkeitsmaßnahmen in den Nachhaltigkeitsbericht des Unternehmens. Zudem sollte das NIM einen integralen Bestandteil des Customer Relationship Managements (CRM) bilden.

4.2 Nachhaltigkeit in den Geschäftsprozessen des Unternehmens

Auf strategischer Ebene muss eine Ausrichtung der IT-Strategie an der Nachhaltigkeitsstrategie des Unternehmens erfolgen (HF IV). Diese kann, je nach Positionierung des Unternehmens am Markt, unterschiedliche Ziele verfolgen. Bieker [25] differenziert an dieser Stelle zwischen Effizienz-, Innovations-, Transformations- und Glaubhaftigkeitsstrategien. Eine Effizienzstrategie kann durch Erhöhung der IT-Energieeffizienz und Reduktion operativer IT-Kosten unterstützt werden. Durch zielgerichtete Investitionen und Aufbau von Know-how können die technologische Innovationskraft des Unternehmens gestärkt und nachhaltigkeitsbezogene Produkt- und Dienstleistungsinnovationen gefördert werden. Durch eine grundlegende Veränderung des Geschäftsmodells und durch ein Reengineering interner Prozessabläufe kann ein Unternehmen nachhaltig transformiert werden, um neue Märkte zu adressieren. Durch ein ungewöhnliches Engagement zur Vermeidung von Emissionen sowie Erhöhung der Transparenz und Stärkung der externen Beziehungen kann eine Stakeholderorientierte Glaubhaftigkeitsstrategie durch den Einsatz von IT gefördert werden.

Das NIM adressiert ebenfalls die Abläufe des Unternehmens auf planerischer Ebene (HF VIII). Durch eine Erweiterung von ERP-Systemen um Nachhaltigkeitsaspekte kann die Effizienz sämtlicher Geschäftsabläufe durch die Vermeidung von Überkapazitäten und den flexiblen Einsatz von Ressourcen verbessert und dadurch negative Umweltauswirkungen reduziert und Kosten gesenkt werden. Selbiges gilt für eine grundlegende Neuentwicklung von Geschäftsprozessen, welche die mögliche Automatisierungs- und Unterstützungsfunktionen der IT optimal nutzen. Beispiele sind Telekonferenzsysteme, intelligente Gebäude- und Flottenmanagement-Systeme oder Materialmanagement-Datenbanken für Einkauf und Recycling.

Durch betriebliche Umweltinformationssysteme (BUIS) können Ressourcenverbräuche und Emissionen kalkuliert, kontrolliert und optimiert werden (HF XII). Zudem wird sowohl bei Mitarbeitern als auch bei Kunden ein Bewusstsein für Umweltbelange geschaffen und Unternehmen, die eine Vorreiterrolle einnehmen, können sich vom Wettbewerb differenzieren. Zu beachten ist an dieser Stelle, dass die Handlungsfelder IV, VIII und XII die internen Produktions- und Geschäftsprozesse adressieren.

4.3 Nachhaltigkeit in Endprodukten und Dienstleistungen des Unternehmens

Die in Abschnitt 4.1 beschriebenen Handlungsfelder wurden bereits in zahlreichen Studien untersucht und eine Vielzahl der identifizierten Maßnahmen findet in Unternehmen bereits Anwendung. Die in Abschnitt 4.2 beschriebenen Handlungsfelder, welche eine enge Zusammenarbeit zwischen der IT-Organisation und den Geschäftsbereichen voraussetzen, werden nur von wenigen Unternehmen erkannt und in Angriff genommen. Auch in der Forschung wurde das Feld „Green IS“ bislang kaum empirisch untersucht – wenngleich viele Wissenschaftler die großen Potenziale er-

kennen. Das Handlungsfeld XIII hingegen ist in Wissenschaft und Praxis weitestgehend unbekannt, nur vereinzelte wissenschaftliche Arbeiten (z. B. [16], [13]) weisen auf die Existenz dieses Handlungs- und Forschungsfeldes hin. Nach unserer Ansicht existiert hier eine Forschungslücke und für Unternehmen eröffnen sich an dieser Stelle durch eine Verflechtung von IT-Know-how und der Entwicklung und dem Design nachhaltiger Produkte und Dienstleistungen mannigfaltige Möglichkeiten. Ein Beispiel für die Reduktion der Umweltauswirkungen einer Dienstleistung, welche durch IT reduziert werden können, ist das Online-Banking, welches CO₂-Emissionen und Papierverbrauch reduziert. Smart Homes, welche den Energieverbrauch von Privathäusern durch Sensoren und dynamische Steuerung von Licht und Heizung drastisch senken können, stellt ein Beispiel für ein Produkt dar, dessen Umweltauswirkungen durch IT reduziert werden. Weitere Beispiele sind intelligente Verkehrsleitsysteme oder die Software für Elektroautos (im Opel Ampera sorgen bspw. zehn Millionen Zeilen Code dafür, dass über 100 computergesteuerte Komponenten funktionieren).

Tabelle 3. Relevante Forschungsthemen für das NIM

1st degree	Govern	- Erweiterung etablierter Konzepte (Green ITIL, Nachhaltigkeit in Cobit, Sustainability-IT-BSC) - Entwicklung von Green IS Strategien - Definition von KPIs und Kennzahlen zu Nachhaltigkeit
	Source	- TCO-Betrachtungen und Life Cycle Analysis - Ökologischer Fußabdruck der Herstellung von IT-Komponenten
	Make	- Methodik zur Ermittlung des Carbon Footprints von IT-Services (Ökobilanzierung) - Erarbeitung von Messkonzepten für den Energieverbrauch von IT-Ressourcen (RZ und BU)
	Deliver	- Definition von IT-Servicekatalogen unter Berücksichtigung von Kenngrößen zu Nachhaltigkeit (z. B. Definition von „Green Service Level Agreements“)
2nd degree		- Unterstützung unternehmensweiter Nachhaltigkeitsziele durch IS (Strategic Green IS Alignment) - Green Business Process Reengineering (Green BPR) und Green Business Process Management - Betriebliche Umweltinformationssysteme (BUIS) - IT-gestützte Nachhaltigkeitstransformation - Kooperation und Identifikation von Synergien zwischen IT-Organisation und den Fachbereichen
3rd degree		- Innovative Maßnahmen und Möglichkeiten zur Reduktion des Fußabdrucks von Produkten und Dienstleistungen unterschiedlicher Unternehmen - Identifikation von Einsatzgebieten sowie Bestimmung und Vergleich der Potenziale nachhaltiger IT in unterschiedlichen Branchen

5 Fazit und Ausblick

Die zunehmende Bedeutung von Nachhaltigkeitsaspekten in IT-Organisationen lässt sich, wie in dieser Arbeit aufgezeigt, auf zwei Entwicklungsströmungen zurückführen: Einerseits steigen der IT-Ressourcenbedarf und die Energiepreise für den Betrieb von IT-Infrastrukturen und machen die Implementierung von Maßnahmen zur Verringerung der operativen Kosten wirtschaftlich interessant. Andererseits lässt sich aktuell eine wachsende Kundennachfrage nach nachhaltigen (IT-)Produkten beobachten und so kann durch den intelligenten Einsatz von IT ein wichtiger Beitrag zur Steigerung der ökologischen Nachhaltigkeit sowohl in den Geschäfts- und Produktionsprozessen als auch in den Produkten und Dienstleistungen eines Unternehmens geleistet werden. Im empirischen Teil der Arbeit wurde hierzu der State-of-the-Art in IT-Organisationen explorativ analysiert. Die Untersuchungen haben verdeutlicht, dass es im Informationsmanagement an einer theoretischen und konzeptionellen Grundlage zur Nachhaltigkeit fehlt. Obwohl erste Ansätze zum Umweltschutz und zur Ressour-

ineffizienz existieren, mangelt es bislang noch an klaren Strategien und Vorgehensweisen, aus denen sich ein entsprechendes Nachhaltigkeitsmanagement für die gesamte IT-Wertschöpfungskette ableiten lässt. Es werden zwar in vielen IT-Organisationen Nachhaltigkeitsinitiativen auf operativer Ebene gestartet, meist jedoch ohne Berücksichtigung der strategischen Relevanz. Eine Harmonisierung dieser Maßnahmen mit den strategischen Nachhaltigkeitszielen des Unternehmens ist erforderlich, um die ökonomischen, ökologischen und sozialen Nachhaltigkeitsziele erreichen zu können. Zu diesem Zweck wurde das NIM-Modell erarbeitet, das als Grundlage für die Implementierung eines ganzheitlichen Nachhaltigkeitsmanagements in IT-Organisationen dienen soll. Das Modell differenziert zwischen 13 konkreten Handlungsfeldern und eignet sich somit als praxisorientiertes Analyseinstrument zur Identifikation möglicher Nachhaltigkeitsmaßnahmen entlang spezifischer Wertschöpfungsphasen und Organisationsebenen. Für die weiterführende Forschung in diesem Themenbereich stellt das erarbeitete NIM-Modell einen ganzheitlichen Strukturierungsansatz dar und zeigt eine Vielzahl möglicher Forschungsfelder auf. Für die angewandte Forschung ergibt sich damit die Notwendigkeit, weiterhin an innovativen und aussagekräftigen (Management-)Konzepten und Standards zur Umsetzung und Unterstützung von Nachhaltigkeitsgrundsätzen zu arbeiten. So wird betont, dass „[...] sich die Wirtschaftsinformatik der Umwelt- und Nachhaltigkeitsdiskussion stellen [...]“ und hierbei sowohl ihren theoretisch-methodischen als auch praktisch-gestalterischen Beitrag zur Lösung der aktuellen Herausforderungen in der Praxis leisten muss [11].

Literatur

1. Zarnekow, R., Brenner, W., Pilgram, U.: Integriertes Informationsmanagement. Strategien und Lösungen für das Management von IT-Dienstleistungen. Springer, Berlin (2005)
2. Straube, F., Doch, S., Borkowski, S., Nagel, A.: Kundenorientierung und Nachhaltigkeit als Treiber für ein Innovatives Logistikcontrolling. Controlling 21, 431–437 (2009)
3. Buchta, D., Eul, M., Schulte-Croonenberg, H.: Strategisches IT-Management. Wert steigern, Leistung steuern, Kosten senken. Gabler, Wiesbaden (2009)
4. Global e-Sustainability Initiative (GeSi): The Boston Consulting Group SMART 2020 Addendum Deutschland: Die IKT Industrie als Treibende Kraft auf dem Weg zu Nachhaltigem Klimaschutz, <http://www.gesi.org/LinkClick.aspx?fileticket=X7m82qhz%2F60%3D&tabid=60>
5. Gartner Inc.: Gartner Highlights Key Predictions for IT Organizations and Users in 2010 and Beyond, <http://www.gartner.com/it/page.jsp?id=1278413>
6. Elkington, J.: Cannibals with Forks: The Triple Bottom Line of 21st Century Business. Capstone, Oxford (1997)
7. Molla, A.: The Reach And Richness Of Green IT: A Principal Component Analysis. In: ACIS Proceedings, pp. 754–764. AIS Electronic Library (2009)
8. Murugesan, S., Gangadharan, G.R.: Harnessing Green IT: Principles and Practices. IEEE Press, New York (2008)
9. Watson, R.T., Boudreau, M.C., Chen, A.J.: Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community. MIS Quarterly 34, 23–38 (2010)

10. Elliot, S.: Transdisciplinary Perspectives on Environmental Sustainability: A Resource Base and Framework for IT Enabled Business Transformation. *MIS Quarterly* 35 (1), 197-236 (2011)
11. Loos, P., Nebel, W., Marx Gómez, J., Hasan, H., Watson, R.T., Brocke, J., Seidel, S., Recker, J.: Green IT: Ein Thema für die Wirtschaftsinformatik?. *Wirtschaftsinformatik* 53, 239–247 (2011)
12. Melville, N.: Information Systems Innovation for Environmental Sustainability. *MIS Quarterly* 34 (1), 1-21 (2010)
13. Mingay, S., Di Maio, A.: Defining the Environmental Value of IT (2007)
14. Harmon, R.R., Demirkan, H., Auinger, A., Reinoso, M.: From Green Computing to Sustainable IT: Developing a Sustainable Service Orientation. In: 43rd Hawaii International Conference on System Sciences (2010)
15. Melville, N., Kraemer, K.: Review - Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly* 28, 283–322 (2004)
16. Hilty, L.M., Arnfalk, P., Erdmann, L., Goodman, J., Lehmann, M., Wäger, P.A.: The Relevance of Information and Communication Technologies for Environmental Sustainability – A Prospective Simulation Study. *Environmental Modelling & Software* 21, 1618–1629 (2006)
17. Nedbal, D., Wetzlinger, W., Auinger, A., Wagner, G.: Sustainable IS Initialization Through Outsourcing: A Theory-Based Approach. In: 17th Americas Conference on Information Systems (AMCIS) (2011)
18. Österle, H., Winter, R. (eds.): *Business Engineering*. Springer, Berlin (2003)
19. Ereğ, K., Löser, F., Schmidt, N.-H., Zarnekow, R., Kolbe, L. M.: Green IT Strategies: A Case Study-Based Framework for Aligning Green IT with Competitive Environmental Strategies. In: PACIS Proceedings (2011)
20. Ereğ, K., Schmidt, N.-H., Zarnekow, R., Kolbe, L. M.: Green IT im Rahmen eines nachhaltigen Informationsmanagements – Status-quo und Handlungsempfehlungen für die Praxis. *HMD* 274, 18-27 (2010)
21. Schmidt, N.-H., Ereğ, K., Kolbe, L. M., Zarnekow, R.: Examining the Contribution of Green IT to the Objectives of IT Departments. *Australian Journal of Information Systems (AJIS)* 171, 127-140 (2010)
22. Dube, L., Pare, G.: Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly* 27, 597–635 (2003)
23. Eisenhardt, K.M.: Building Theories from Case Study Research. *Academy of Management Review* 14, 532–550 (1989)
24. IT Governance-Institute: Board Briefing on IT Governance Illinois, USA (2005)
25. Bieker, T.: Sustainability Management with the Balanced Scorecard Institute for Economy and the Environment, St. Gallen (2005)

Kennzahlenbasierte Erfolgsmessung von Green IT-Maßnahmen – Eine empirische Analyse zum aktuellen Stand in Forschung und Praxis

Nicky Opitz¹, Florian Thies¹, Koray Ereke², Lutz M. Kolbe¹, und Rüdiger Zarnekow²

¹ Georg-August-Universität Göttingen, Göttingen, Germany
{nopitz, fthies, lkolbe}@uni-goettingen.de

² Technische Universität Berlin, Berlin, Germany
{koray.erek, ruediger.zarnekow}@tu-berlin.de

Abstract. In diesem Beitrag wird der Umsetzungsstand von Green IT-Kennzahlen in IT-Organisationen verschiedener Größe und Branchen untersucht. Zu diesem Zweck wurden aktuelle Kennzahlen aus der Literatur entnommen und auf Grundlage eines Bezugsrahmens den (Prozess-)Bereichen der IT-Beschaffung, der IT-Produktion und des IT-Vertriebs systematisch aufbereitet. Die Erfassung des Status quo in der Praxis wurde mittels einer Befragung von insgesamt 94 CIOs und IT-Führungskräften aus einer Stichprobe von 822 Unternehmen analysiert. Die Ergebnisse deuten darauf hin, dass Unternehmen den Erfolg von Green IT-Maßnahmen kennzahlbasiert erfassen und kontrollieren können, wobei der Einsatz solcher Kenngrößen von der Unternehmensgröße, der Nachhaltigkeitsorientierung und der Green IT-Expertise abhängig sind.

Keywords: Green IT, KPI, Kennzahlen, Erfolgsmessung, Nachhaltigkeitscontrolling

1 Einleitung

Mit der globalen Vernetzung und der damit einhergehenden starken IT-Durchdringung in den Unternehmensprozessen nimmt auch gleichzeitig der Einfluss der IT auf Ökonomie, Ökologie und Gesellschaft stetig zu. Die ökologischen Herausforderungen in der IT werden seit einigen Jahren in der Öffentlichkeit und Wissenschaft unter dem Begriff „Green IT“ diskutiert. Sie ergeben sich insbesondere aufgrund des rasanten Anstiegs des Energieverbrauchs von IT-Komponenten, inklusive der zum Betrieb notwendigen Infrastruktur, und den damit verbundenen klimaschädlichen CO₂-Emissionen. Durch u. a. immer komplexere Geschäftsanwendungen, die leistungsfähigere Server erfordern, und die sich intensivierende Nutzung des Internets in den vergangenen Jahren wird dieser Trend bekräftigt [5].

In Deutschland betrug der Anteil des IT-bedingten Stromverbrauches am Gesamtstromverbrauch bereits im Jahr 2001 ca. 7,1 % (38 Terrawattstunden – TWh). Dabei

ist dieser Anteil in den letzten Jahren kontinuierlich auf heute über 10,5 % (55,4 TWh) gestiegen.

Demgegenüber sehen sich IT-Organisationen mit gekürzten IT-Budgets sowie der Forderung der Unternehmensleitung konfrontiert, den Wertbeitrag der IT zur Nachhaltigkeitsstrategie des Unternehmens darzulegen und messbar zu machen [3]. Infolgedessen gewinnen Fragen der ressourceneffizienten IT-Leistungserstellung im Rahmen einer Green IT-Strategie deutlich mehr an Bedeutung als noch vor einigen Jahren. Nicht zuletzt bedingt durch die hohen Kosteneinsparpotenziale, aber auch durch die zunehmende Aufmerksamkeit der Politik und Wissenschaft konnten sich somit bereits eine Reihe von Green IT-Maßnahmen in der Praxis etablieren. Kernthema dieses Beitrags ist die auf Kennzahlen basierende Erfolgsmessung solcher Maßnahmen. Im Einzelnen sollen folgende Fragen beantwortet werden:

- Welche Green IT-Kennzahlen lassen sich in der Literatur bekannt und wie lassen sie sich systematisieren?
- Welche dieser Kennzahlen werden in Unternehmen tatsächlich erhoben?
- Welchen Einfluss spielen externe Faktoren wie Unternehmensgröße oder Green IT-Expertise auf die Messung solcher Kennzahlen?

Dazu wurden zunächst Kennzahlen aus der Literatur identifiziert und darauf aufbauend ein Fragebogen entwickelt, an 830 verschiedene Unternehmen versandt und mittels deskriptiver Statistik und Hypothesentests ausgewertet.

2 Forschungsstand

2.1 Green IT und Green IS

Seit einigen Jahren hat die IT-Industrie ihren erkannten Beitrag zum Umwelt- und Ressourcenschutz mit medienwirksamen Begriffen wie „Green IT“ [20], „Green Information Systems (IS)“ [27], „Environmental Sustainability of IT“ [9] oder auch „IT-for-Green“ [19] versehen.

Harmon und Auseklis zählen Kostenreduktions- und Performance-Ziele als treibende Motive einer Green IT-Adaption in IT-Organisationen und deuten daher Green IT als „the practice of maximizing the efficient use of computing resources to minimize environmental impact“ [15]. Dieser Auffassung wird im Rahmen dieser Arbeit entsprochen und daher folgende Green IT-Definition zugrunde gelegt:

Green IT umfasst sämtliche Maßnahmen und Lösungen, die zu einer effizienteren Nutzung der Ressource Energie durch die betriebene IT-Infrastruktur und zu einer umweltfreundlicheren Produktion und Verwertung von IT-Hardware beitragen, einschließlich der begleitenden Aktivitäten zu deren Steuerung und Kommunikation [30].

In jüngerer Zeit setzt sich jedoch vermehrt eine breiter gefasste Sicht in der Praxis und Wissenschaft durch [27]. So präferieren Watson et al. den umfassenderen Ansatz und bemängeln ebenfalls den rein technologiefokussierten Gedanken einer Green IT. Sie ziehen den Begriff Green IS vor und argumentieren, dass die Rolle der IT als

Enabler nachhaltiger Geschäftsprozesse in Unternehmen („IT-for-Green“) einen zentralen Bestandteil der Nachhaltigkeitsdiskussionen in der IT bilden muss.

In dieser Arbeit wird vorrangig Green IT als ökologisch nachhaltiger Ansatz zur Bewältigung der Nachhaltigkeitsherausforderungen in IT-Organisationen betrachtet. Die Rolle der IT zur Unterstützung von Nachhaltigkeit in den Geschäftsprozessen eines Unternehmens ist nicht Gegenstand der Untersuchung.

2.2 Modell des integrierten Informationsmanagements als Bezugsrahmen

Im Zuge einer zunehmenden Dienstleistungsorientierung im Informationsmanagement haben sich die Interaktionsmodelle zwischen IT-Organisationen (Leistungserbringer) und den Fachbereichen (Leistungsabnehmer) verändert und weiterentwickelt. Damit einhergehend ist von einem neuen Aufgaben- und Rollenverständnis in den IT-Organisationen die Rede, in welchem diese sowohl als unternehmensinterne als auch auf dem freien Markt agierende externe Einheiten in Erscheinung treten können. Im Rahmen ihrer Geschäftstätigkeit können sie IT-Produkte (Hardware, Software, Dienstleistungen) für interne (in der Regel Fachbereiche) bzw. externe Kunden anbieten. Anstelle einer traditionell projekt- und aufgabenbasierten Zusammenarbeit (Plan-Build-Run) tritt daher vermehrt eine marktorientierte Kunden-Lieferanten-Beziehung in Erscheinung, bei der IT-Organisationen verstärkt die Rolle eines Dienstleisters einnehmen [16]. Auf Grundlage dieser Kunden-Lieferanten-Beziehung ist es möglich, etablierte Referenzmodelle für das Supply-Chain-Management auch auf das Informationsmanagement zu übertragen. Das Modell eines Integrierten Informationsmanagements (IIM) folgt dieser Idee und stellt die zentralen Managementprozesse von IT-Organisationen auf Basis eines Source-Make-Deliver-Ansatzes dar, die zur Herstellung und Nutzung von IT-Produkten erforderlich sind [29], [16]. Das IIM-Modell legt somit eine wertschöpfungskettenorientierte Sicht zugrunde, einschließlich der Schnittstellen zu Lieferanten und Kunden (vgl. Abbildung 1). Dabei wird der Prozess der IT-Leistungserbringung als integrierter Fertigungsprozess betrachtet.

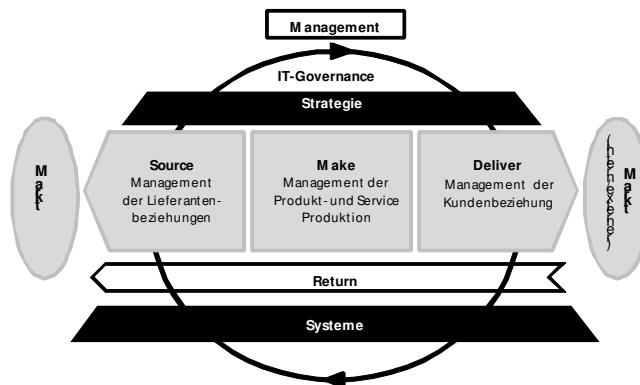


Abb. 1. Das Modell des Integrierten Informationsmanagements [29], [16]

2.3 Status quo zu Green IT-Kennzahlen

Im Vorfeld der Befragung wurde ein umfangreicher Katalog möglicher Green IT-Kennzahlen erstellt. Dafür wurde das Phasenmodell zur Literaturanalyse nach Fettke [11] adaptiert, welches im Rahmen der Review-Forschung zur Abbildung des State-of-the-Arts entwickelt wurde. Um zu einer systematischen Kategorisierung zu gelangen, wurden die einzelnen Kennzahlen in das Modell des integrierten Informationsmanagements eingeordnet (vgl. Abbildung 1), auf dessen Kategorien der Fragebogen schlussendlich auch beruhte.

Die Ergebnisse sind in den folgenden Tabellen 1-4 zusammengefasst. Einige genannte Kennzahlen sind in der Forschung noch nicht einheitlich benannt und/oder wurden im Rahmen dieser Forschungsarbeit aus bestehenden Ansätzen hergeleitet.

Im Bereich Make, also in der Beschaffung von IT-Gütern wie Hardware- oder Softwareressourcen und IT-Dienstleistungen ist die Erforschung einer kennzahlenbasierten Messung hinsichtlich ökologischer Nachhaltigkeit noch im Anfangsstadium. Es konnten drei mögliche Kennzahlen ermittelt werden (vgl. Tabelle 1). Das aus der klassischen Betriebswirtschaftslehre bekannte Konzept der Total Cost of Ownership (TCO) kann durch Berücksichtigung von Energiekosten oder verursachte Emissionen in den Green IT-Kontext überführt werden. Die Ansätze „Anteil erneuerbarer Energie“ und „Anteil ökologisch Nachhaltiger IT-Komponenten“ sind in dieser Form in der Literatur als Kennzahl nicht benannt, lassen sich aber aus vorhandenen Gedankengängen ableiten.

Tabelle 1. Green IT-Kennzahlen aus dem Bereich Source

Kennzahl	Abk.	Beschreibung/Formel	Quelle(n)
Total Cost of Ownership	TCO	Lebenzyklusbetrachtung der Gesamtkosten	[4]
Anteil erneuerbarer Energien im IT-Betrieb (in %)	-	$\left(\frac{\text{Einsatz erneuerbarer Energie in der IT (kWh)}}{\text{Gesamter Energiebedarf der IT (kWh)}} \right) * 100$	-
Anteil ökologisch nachhaltiger IT-Komponenten (in %)	-	$\left(\frac{\text{ökologisch nachhaltige IT - Komponenten}}{\text{Gesamtzahl IT - Komponenten}} \right) * 100$	-

Während wie gerade beschrieben im Bereich Source nur wenige Kennzahlen aus der Literatur extrahiert werden konnten, gibt es eine hohe Bandbreite an Messgrößen im Make-Bereich des integrierten Informationsmanagements. Dies könnte vor allem darauf zurückzuführen sein, dass hier die IT-Komponenten betrieben werden und somit Verbrauchs- und Emissionsmessungen direkt durchgeführt werden können. Durch die Fülle von Kennzahlen wird hier im Text nicht näher auf einzelne eingegangen, in Tabelle 2 und Tabelle 3 sind die verwendeten Kennzahlen in die Funktionsbereiche Rechenzentrum (RZ) und Büroumgebung unterteilt.

Tabelle 2. Green IT-Kennzahlen aus dem Bereich Make (Rechenzentrum)

Kennzahl	Abk.	Beschreibung/Formel	Quelle(n)
Power Usage Effectiveness	PUE	$\frac{\text{Gesamter Energieverbrauch eines Rechenzentrums (kWh)}}{\text{Gesamter Energieverbrauch der IT-Komponenten (kWh)}}$	[8]
Data Center Infrastructure Efficiency	DCIE	$\frac{\text{Gesamter Energieverbrauch der IT-Komponenten (kWh)}}{\text{Gesamter Energieverbrauch eines Rechenzentrums (kWh)}}$	[5]
Data Center energy Productivity	DCeP	$\frac{\text{Geleisteter Output des Rechenzentrums}}{\text{Gesamtenergieverbrauch des Rechenzentrums}}$	[14], [8]
Energy Reuse Efficiency	ERE	$\frac{\text{Gesamter Energieverbrauch des RZ} - \text{Wiederverwendete Energie}}{\text{Energieverbrauch der IT-Komponenten}}$	[21]
Carbon Usage Effectiveness	CUE	$\frac{\text{Menge CO2-Equivalent aus Energieverbrauch RZ}}{\text{Energieverbrauch der IT-Komponenten im RZ}}$	[4]
Free Cooling Potential	-	Gibt darüber Auskunft, wie viel kWh jährlich durch den Einsatz von Freiluftkühlung an Stelle konventioneller Kühlung eingespart werden kann.	[24]
Deployed Hardware Utilization Ratio	DH-UR	Auslastungsgrad von z. B. Servern, Storage	[24]
Standard Power Evaluation Corporation power	SPEC _{power}	$\frac{\sum \text{Rechenleistung}}{\sum \text{Leistungsaufnahme}}$	[22]

Tabelle 3. Green IT-Kennzahlen aus dem Bereich Make (Büroumgebung)

Kennzahl	Abk.	Beschreibung/Formel	Quelle(n)
Anteil sensibilisierter Mitarbeiter	-	$\frac{\text{Mitarbeiter, die an Green IT-Sensibilisierung teilgenommen haben}}{\text{Anzahl Mitarbeiter im Unternehmen}}$	[30], [5]
Menge Hardware zu Anzahl Mitarbeiter	-	IT-Geräte sollen effizient genutzt werden (Auslastung). In vielen Unternehmen erfolgt die Erweiterung der Gerätelandschaft aber eher unsystematisch.	[26]
Hardware Compute Load	HCL	$\frac{\text{Leistungsaufnahme am Netzteil} - \text{Verlust durch Umwandlung}}{\text{Leistungsaufnahme interner Lüfter}}$	[23]
Erfolg von Green-IT Projekten	-	$\frac{\text{Anzahl erfolgreicher Green IT-Projekte}}{\text{Anzahl Green IT-Projekte}}$	[2]

Für den Bereich Deliver, der die Handlungsfelder Absatz oder Vertrieb sowie Kommunikation umfasst, konnten aus der Literatur keine konkret messbaren Kennzahlen für den Green IT-Bereich abgeleitet werden. Für die interne und externe Kommunikation von Green IT-Bestrebungen eignen sich jedoch Green IT-Benchmarks bzw. Zertifizierungen, von denen einige Beispiele in Tabelle 4 beschrieben sind. Diese wurden auch in der Befragung verwendet.

Tabelle 4. Green IT-Kennzahlen aus dem Bereich Deliver

Kennzahl	Beschreibung ¹	Quelle(n)
Energieeffizienz im Rechenzentrum vom TÜV Rheinland	100 Seiten umfassender Kriterienkatalog bewertet das Engagement hinsichtlich Minimierung des Energieeinsatzes.	[17]
RZ-Benchmarking GreenIT BB Netzwerk	30 Parameter zur Bewertung der Ressourceneffizienz, kann punktuell oder dauerhaft durchgeführt werden.	[13]
Datacenter Benchmarking Technische Universität Berlin	Ermöglicht das Aufdecken von Einsparpotenzialen in Rechenzentren anhand von sieben Kennwerten. Dadurch geringer Aufwand.	[28]
Blauer Engel - Umweltzeichen Energiebewusster Rechenzentrumsbetrieb	Dient zur Auszeichnung von Rechenzentren, deren Betreiber eine langfristig ausgerichtete Strategie zur Optimierung der Energie- und Ressourceneffizienz verfolgen.	[18]
Green-IT-Certification DEKRA	Berücksichtigt zusätzlich zu konkreten Maßnahmen auch Managementprozesse und Konzepte.	[10]

¹ Die Beschreibung der Benchmarks basiert zum Teil auf den Angaben der Anbieter und wurde von den Autoren im Rahmen dieses Beitrags nicht validiert.

3 Forschungsmethodik

3.1 Vorgehen bei der Befragung

Für die Unternehmensbefragung wurden Entscheidungsträger und IT-Verantwortliche von in Deutschland ansässigen Unternehmen für eine schriftliche Umfrage ausgewählt. Die Befragung erfolgte postalisch, umfasste zusätzlich zum Fragebogen ein personalisiertes Anschreiben sowie einen unfrankierten Rückumschlag. Durch das persönliche Anschreiben sowie den kompakt gehaltenen Fragebogen, der aus einer doppelseitig bedruckten DIN-A4-Seite bestand, wurde eine hohe Rücklaufquote angestrebt.

Die Stichprobengröße umfasste insgesamt 830 in Deutschland ansässige - zum Großteil aber international tätige - Unternehmen, bei denen der Fragebogen je nach vorhandener Datenlage an den Chief Information Officer (CIO) bzw. IT-Verantwortlichen oder an die Geschäftsführung adressiert wurde. Um ein möglichst breites Spektrum von Unternehmen in der Stichprobe zu erfassen, wurden sowohl große Unternehmen als auch kleine und mittelständische Unternehmen unterschiedlicher Branchen einbezogen, die sich wie folgt aufteilen:

- 500 Großunternehmen, darunter je 20 Banken und Versicherungen
- 330 kleine und mittelständische Unternehmen (KMU), darunter 50 Rechenzentren

Die Anzahl der Unternehmen innerhalb der beiden Gruppen wurden zufällig gewählt. Es wurde sichergestellt, dass alle Bereiche abgedeckt sind und keine Branche bewusst ausgeschlossen wird [7].

Zur Beantwortung des Fragebogens waren etwa fünf Minuten nötig. Diese Zeitspanne wurde im Vorfeld der Umfrage durch einen Pretest evaluiert. Der Versand der 830 Fragebögen erfolgte am 21. Dezember 2011 und es wurde im Anschreiben um eine Rücksendung des Fragebogens bis zum 21. Januar 2012 gebeten. Die letzten in die Auswertung aufgenommenen Antworten trafen am 20. Februar 2012 ein.

Der Fragebogen war in fünf Abschnitte unterteilt. Im ersten Abschnitt wurden allgemeine Daten zum Thema Nachhaltigkeit abgefragt, beispielsweise wie das Unternehmen sein Engagement beim Thema Umweltschutz im Allgemeinen und bei Green IT im Speziellen einschätzt. Außerdem wurde nach einer Selbsteinschätzung zur Expertise im Bereich Green IT gefragt sowie versucht zu erfassen, wo im Unternehmen die Verantwortung für Green IT-Maßnahmen angesiedelt ist. Abschließend wurde erhoben, ob das Unternehmen bereits Green IT-Maßnahmen umgesetzt hat. In den nächsten drei Abschnitten wurden nach den oben genannten Kennzahlen, gegliedert nach Source, Make und Deliver, gefragt. Insbesondere wurde der Frage nachgegangen, ob die genannten Kennzahlen erhoben werden. Ferner wurde ein freies Textfeld geschaffen, indem weitere, nicht im Fragebogen aufgelistete Kennzahlen eingetragen werden konnten. Im letzten Abschnitt wurden allgemeine Unternehmensdaten wie Branche, Mitarbeiterzahl und Umsatz erhoben.

3.2 Hypothesen

Zusätzlich zur deskriptiven Datenanalyse über den Bekanntheitsgrad und die tatsächliche Messung von Green IT-Kennzahlen wurden einige induktive Hypothesen auf Basis der gewonnenen Erkenntnisse aus der Literaturanalyse formuliert und getestet:

H1: Je größer ein Unternehmen gemessen an der Mitarbeiterzahl ist, desto mehr Green IT-Maßnahmen wurden bereits umgesetzt.

Hypothese H1 stützt sich auf die Erkenntnis, dass große Unternehmen als Vorreiter bei der Umsetzung von Green IT-Maßnahmen gelten.

H2: Unternehmen, die ein eigenes RZ betreiben, erheben mehr Green IT-Kennzahlen als Unternehmen, die kein eigenes RZ betreiben.

Hypothese H2 basiert auf der Feststellung, dass es vorwiegend Green-IT-Kennzahlen gibt, die sich im Rahmen der Energieeffizienzmessung von RZ erheben lassen.

H3: Je höher die Selbsteinschätzung der Green IT-Expertise eines Unternehmens ist, desto mehr Green IT-Kennzahlen werden erhoben.

Aufgrund der Komplexität der Erhebung einiger Green IT-Kennzahlen sowie der geringen Thematisierung von Kennzahlen in der Theorie wird vermutet, dass ein positiver Zusammenhang zwischen der Anzahl der erhobenen Green IT-Kennzahlen eines Unternehmens und der Selbsteinschätzung der Green IT-Expertise eines Unternehmens besteht.

H4: Je höher die Selbsteinschätzung der Green-IT-Expertise der Unternehmen ist, desto weniger Herausforderungen werden in Bezug auf Green-IT-Kennzahlen gesehen.

Analog zu H3 wird in H4 ein negativer Zusammenhang zwischen der Selbsteinschätzung der Green-IT-Expertise eines Unternehmens und der Anzahl der Herausforderungen, die das Unternehmen bezüglich der Erhebung von Green IT-Kennzahlen sieht, vermutet.

3.3 Datenauswertung

Im Rahmen dieses Beitrags werden vorwiegend deskriptive Statistiken eingesetzt und um induktive Verfahren zum Testen der im vorherigen Abschnitt aufgestellten Hypothesen ergänzt. Bei dieser Untersuchung handelt es sich um eine Querschnittstudie zu einem festen Zeitpunkt. Anhand dieser werden Aussagen über die Gesamtheit der untersuchten Unternehmen bzw. über die zugrundeliegende Grundgesamtheit getroffen [1].

4 Ergebnisse

4.1 Stichprobe

Von den 830 versendeten Fragebögen konnten acht Fragebögen nicht zugestellt werden. Daher verringert sich der Stichprobenumfang auf 822 Unternehmen, von denen 94 geantwortet haben. Damit beläuft sich die Rücklaufquote auf 11,44 %. Da nicht alle Fragebögen vollständig ausgefüllt wurden, wird der prozentuale Anteil an Enthaltungen jeweils im Rahmen der Auswertung einzelner Fragen angegeben.

Die Branchenzugehörigkeit, der Jahresumsatz 2010 und die Mitarbeiteranzahl der teilnehmenden Unternehmen ist in Tabelle 5 dargestellt. Die Gliederung der verschiedenen Branchen orientierte sich dabei an der Branchenklassifikation des Statistischen Bundesamtes [6]. Wie aus der Tabelle ersichtlich, dominieren größere Unternehmen die Stichprobe. Da sich auch unter den 830 angeschriebenen Firmen 500 große befanden, repräsentieren die Rückläufer diese ursprünglich angeschriebene Gruppe sehr gut. Diese Stichprobenverteilung enthält somit keine Anzeichen für eine Schweißverzerrung – zumindest was die Unternehmensgröße angeht.

Tabelle 5. Stichprobenübersicht

Branche	Jahresumsatz 2010 in Millionen Euro		Anzahl Mitarbeiter		
	%	%	%	%	
Verarbeitendes Gewerbe	35,1	< 5	4,3	< 50	8,5
Handel	14,9	< 10	6,4	< 250	18,1
Verkehr und Kommunikation	12,8	< 50	10,6	< 1.000	7,5
Kredit- und Versicherungsgewerbe	10,6	< 500	10,6	< 5.000	19,2
Energie- und Wasserversorgung	6,4	< 1.000	6,4	< 15.000	16,0
Sonstige	10,6	≥ 1.000	51,2	≥ 15.000	27,6
Enthaltungen: 9,6 %		Enthaltungen: 10,6 %		Enthaltungen: 3,2 %	

4.2 Deskriptive Statistiken

Nachhaltigkeit und Green IT. Zum Thema Umweltschutz im Allgemeinen gaben 43,6 % der Unternehmen an, einen dezidierten Umweltbeauftragten zu haben. Bei der Selbsteinschätzung zum Thema Umweltschutz gaben 21,2 % der Befragten an, dieses Thema genieße „hohe“, bei weiteren 39,4 % sogar eine „sehr hohe“ Priorität. Hierzu ist anzumerken, dass eine Selbsteinschätzung zu sozialadäquatem Antwortverhalten führen kann. Für den Bereich Green IT ist in 29,8 % der Fälle der CIO verantwortlich, in 18,1 % ein anderer IT-Mitarbeiter, in 7,5 % der Umweltbeauftragte und in nur 6,4 % ein hierfür vorgesehener Green IT-Beauftragter. Erwähnenswert ist, dass bei 14,9 % der befragten Unternehmen, niemand für diesen Bereich verantwortlich ist. In Bezug auf die Green IT-Expertise schätzen diese 24,5 % als „durchschnittlich“, 25,3 % als „hoch“ und 6,4 % als „sehr hoch“ ein. 87,2 % der teilnehmenden Unternehmen gaben an, dass in ihrem Unternehmen bereits Green IT-Maßnahmen umgesetzt wurden, während dies in 12,8 % der Unternehmen bisher nicht der Fall ist. Als Gründe hierfür wurden Kosteneinsparungen mit 26,6 % am häufigsten mit „sehr hoher“ Relevanz angegeben.

Kennzahlen im Bereich Source: Im Bereich der Beschaffung wird der Anteil ausgetauschter IT-Komponenten mit 46,81 % am häufigsten erhoben (vgl. Abbildung 1). Des Weiteren ermitteln 42,55 % der Unternehmen den TCO von IT-Komponenten, während der Anteil eingesetzter erneuerbarer Energien zum Betrieb der IT vergleichsweise nur bei 24,47 % der Unternehmen erhoben wird.

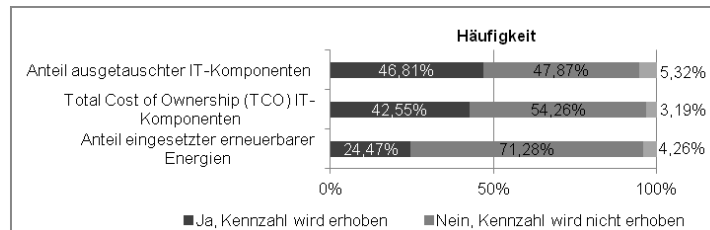


Abb. 2. Einsatz Green IT-Kennzahlen im Bereich Source

Als „sonstige“ Kennzahlen im Bereich Beschaffung wurde jeweils von einem Unternehmen die Leistungsaufnahme von IT-Geräten, der statistische Energieverbrauch bzw. dessen Reduzierung sowie die Erhebung von Kennzahlen genannt, sofern geeignete Werte von Herstellern mitgeliefert werden.

Kennzahlen im Bereich Make. In Abbildung 2 ist die Ausprägung der erhobenen Green IT-Kennzahlen im Bereich der Rechenzentren (Make) dargestellt. Die Auslastung der Server wird bei 56,4 % der teilnehmenden Unternehmen erhoben und ist damit die am häufigsten erhobene Green IT-Kennzahl im Bereich RZ. Unternehmen, welche die Auslastung der Server erfassen, ermitteln zu 79,3 % die Auslastung des Storage (Festplattenspeicher etc.), zu 77,4 % die Auslastung der CPU und zu 75,5 % die Auslastung des Arbeitsspeichers.

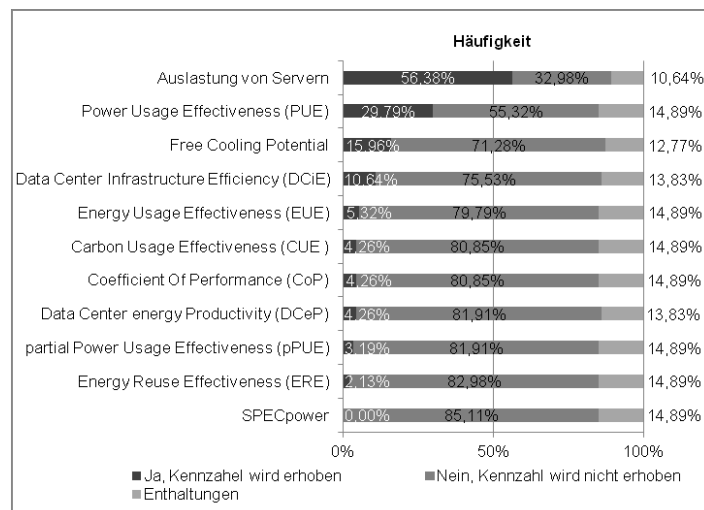


Abb. 3. Einsatz Green IT-Kennzahlen im Bereich Make (Rechenzentrum)

Die PUE-Kennzahl wird von 29,8 % der Unternehmen ermittelt und ist damit die am zweithäufigsten erhobene Green IT-Kennzahl im RZ-Bereich, gefolgt vom Free Cooling Potential, welches bei 16,0 % und der Data Center Infrastructure Efficiency, welche bei 10,6 % der teilnehmenden Unternehmen ermittelt werden. Alle weiteren Green IT-Kennzahlen kommen nur in weniger als sechs Prozent der Unternehmen zum Einsatz.

Abbildung 3 zeigt den Praxiseinsatz von Green IT-Kennzahlen im Bereich Büro-umgebung. In 68,09 % der teilnehmenden Unternehmen wird die Anzahl der Hardware beispielsweise die Anzahl an Druckern in einer Kennzahl erfasst und 21,28 % der Unternehmen halten die Anzahl erfolgreich abgeschlossener Green IT-Projekte in einer Kennzahl fest. Der Anteil an sensibilisierten Mitarbeitern hinsichtlich eines ressourcenschonenden Umgangs wird bei 20,21 % der Unternehmen in einer Kennzahl festgehalten. Im Gegensatz dazu, dass die Auslastung von Servern bei 56,38 % der Unternehmen erfasst wird, erheben nur 19,15 % der Unternehmen die Auslastung des IT-Equipments in der Büroumgebung. Dabei wird zu 83,33 % die Auslastung des Arbeitsspeichers, zu 77,78 % die Auslastung der CPU und zu 61,11 % die Auslastung des Storage erhoben.

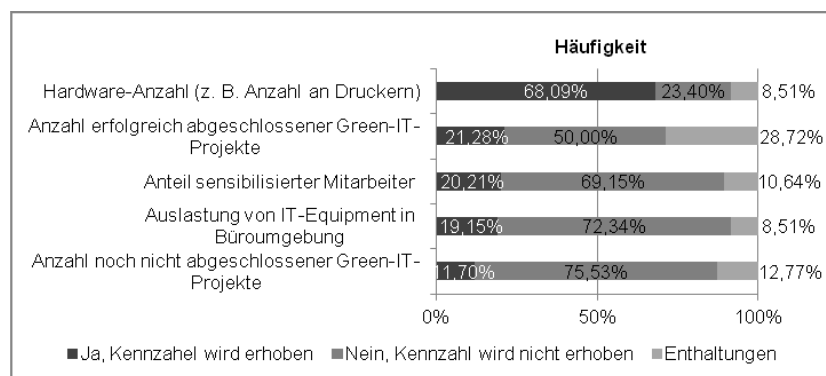


Abb. 4. Einsatz Green IT-Kennzahlen im Bereich Make (Büroumgebung)

Kennzahlen im Bereich Deliver. Zu Green IT-Kennzahlen im Bereich des IT-Vertriebs und der IT-Kommunikation (Deliver) wurde zunächst gefragt, ob diese im Nachhaltigkeitsbericht aufgeführt werden. Nur bei 14,9 % der Unternehmen war dies der Fall. Ein Anteil von 5,3 % der Unternehmen enthielt sich bei dieser Frage. Des Weiteren wurde evaluiert, ob Green IT-Kennzahlen für Marketingzwecke zum Einsatz kommen. Dies ist nur bei 12,8 % der Unternehmen der Fall, während die Mehrheit der Unternehmen in Höhe von 83,0 % keine Green IT-Kennzahlen für Marketingzwecke einsetzt. Ein Anteil von 4,2 % der Unternehmen enthielt sich bei dieser Frage. Sonstige Green IT-Kennzahlen, die sich dem Bereich Absatz zuordnen lassen, wurden nicht genannt.

Des Weiteren wurde evaluiert, ob die Unternehmen an Green IT-Benchmarks sowie Zertifizierungsangeboten teilnehmen. An Green IT-Benchmarks nehmen 11,7 % der befragten Unternehmen teil. Die Mehrheit (27,3 %) dieser Unternehmen

nimmt am GreenIT RZ-Benchmarking der GreenIT-BB teil, 18,2 % am Benchmark energieeffizientes RZ vom TÜV Rheinland und 36,4 % nehmen an sonstigen Benchmarks teil. Ein Green IT-Zertifikat besitzen 7,5 % der befragten Unternehmen. Davon sind 42,9 % der Unternehmen durch das Zertifikat „Energieeffizientes Rechenzentrum“ vom TÜV Rheinland ausgezeichnet. Unter sonstigen Zertifikaten wurde einmal die Zertifizierung durch die Deutsche Energie-Agentur, einmal die Auszeichnung durch einen Green IT Best Practice Award sowie einmal die Zertifizierung nach der Umweltmanagementnorm ISO 14001 genannt.

4.3 Hypothesen-Tests

H1: Diese Hypothese vermutet einen positiven Zusammenhang zwischen den Merkmalen Mitarbeiterzahl und der Anzahl der bereits umgesetzten Green IT-Maßnahmen. Da sowohl die Evaluierung der Anzahl der Mitarbeiter als auch die Anzahl der bereits umgesetzten Green IT-Maßnahmen im Unternehmen auf Basis einer Ordinalskala erhoben wurde und die Anzahl der Mitarbeiter der Unternehmen nicht annähernd normalverteilt ist, erfolgte der Test dieser Hypothese durch die sogenannte Rangkorrelation auf Basis der Spearman-Korrelation [12].

Das Ergebnis der Korrelationsanalyse von H1 ist in Tabelle 6 abgebildet und der Korrelationskoeffizient r in Höhe von 0,536 belegt, dass ein positiver Zusammenhang mittlerer Stärke zwischen den beiden Variablen besteht, welcher zudem hochsignifikant ist (Signifikanzwert $p < 0,05$).

Tabelle 6. Korrelation - Anzahl Mitarbeiter und Anzahl Green IT-Maßnahmen

			Anzahl Mitarbeiter	Anzahl Green-IT-Maßnahmen
Spearman-Rho	Anzahl Mitarbeiter	Korrelationskoeffizient	1,000	0,536**
		Sig. (2-seitig)		0,000
		N	91	91
	Anzahl Green IT-Maßnahmen	Korrelationskoeffizient	0,536**	1,000
		Sig. (2-seitig)	0,000	
		N	91	91

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Obwohl der Zusammenhang zwischen der Anzahl der Mitarbeiter eines Unternehmens und der Anzahl der bereits durchgeführten Green IT-Maßnahmen nicht sehr stark ist, kann die Hypothese H1 durch diese statistische Analyse unterstützt werden. Anzumerken ist an dieser Stelle jedoch, dass die Anzahl der bereits durchgeführten Green-IT-Maßnahmen kein alleiniger Indikator für das Green IT-Bestreben von Unternehmen ist.

H2: Zum Test von H2 wurde überprüft, ob die Nullhypothese $H_{2(0)}$ (Unternehmen, die ein eigenes RZ betreiben, erheben **nicht** mehr Green-IT-Kennzahlen als Unternehmen, die kein eigenes RZ betreiben) abgelehnt werden kann. Dazu wurde der T-Test für unabhängige Stichproben angewandt.

Tabelle 7. T-Test - Anzahl Green IT-Kennzahlen bei Unternehmen mit und ohne RZ

Anzahl-Green-IT-Kennzahlen	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit				
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz
Varianzen sind gleich	1,389	0,242	2,400	90	0,018	2,683	1,118
Varianzen sind nicht gleich			3,128	24,479	0,005	2,683	0,858

Die Ergebnisse des Hypothesen-Tests sind in Tabelle 7 dargestellt, welche aus den beiden Zeilen „Varianzen sind gleich“ und „Varianzen sind nicht gleich“ besteht. Der Levene-Test gibt zunächst darüber Auskunft, ob sich die Varianz der beiden Gruppen unterscheidet, während der T-Test eine Aussage darüber liefert, ob die Mittelwerte der beiden betrachteten Gruppen signifikant voneinander abweichen. Der Levene-Test ist nicht signifikant, es wird also der Bereich „Varianzen sind gleich“ für den T-Test verwendet. Dieser ist auf mit 0,018 dem Niveau $p < 0,05$ signifikant, so dass $H_{2(0)}$ abgelehnt und kann H_2 angenommen werden kann.

H3: Diese Hypothese vermutet einen positiven Zusammenhang zwischen der Selbsteinschätzung der Green IT-Expertise der Unternehmen sowie der Anzahl der Green IT-Kennzahlen, die ein Unternehmen erhebt. Analog zum Test von H_1 wurde für die statistische Analyse die sogenannte Rangkorrelation auf Basis der Spearman-Korrelation durchgeführt. Das Ergebnis der Korrelationsanalyse von H_3 ist in Tabelle 8 abgebildet und der Korrelationskoeffizient in Höhe von 0,465 zeigt, dass ein positiver Zusammenhang von mittlerer Stärke zwischen den beiden Variablen besteht und mit $p < 0,05$ zudem hochsignifikant ist.

Tabelle 8. Korrelation - Green-IT-Expertise und Green IT-Kennzahlen

			Selbsteinschätzung der Green-IT-Expertise	Anzahl der Herausforderungen bzgl. Green-IT-Kennzahlen
Spearman-Rho	Selbsteinschätzung der Green-IT-Expertise	Korrelationskoeffizient	1,000	0,146
		Sig. (2-seitig)		0,258
	N		62	62
	Anzahl der Herausforderungen bzgl. Green-IT-Kennzahlen	Korrelationskoeffizient	0,146	1,000
		Sig. (2-seitig)	0,258	
	N		62	62

Durch das Ergebnis der Korrelationsanalyse wird Hypothese H_3 unterstützt, sodass sich die Vermutung über einen positiven Zusammenhang zwischen der Selbsteinschätzung der Green-IT-Expertise eines Unternehmens und der Anzahl der Green-IT-Kennzahlen die ein Unternehmen erhebt, gestärkt wird.

H4: Diese Hypothese vermutet einen negativen Zusammenhang zwischen der Selbsteinschätzung der Green IT-Expertise der Unternehmen sowie der Anzahl der Herausforderungen bzgl. Green IT-Kennzahlen, die ein Unternehmen sieht. Analog zum Test von H_1 und H_3 wurde für die statistische Analyse die sogenannte Rangkorrelation auf Basis der Spearman-Korrelation durchgeführt. Der vermutete Zusammenhang lässt sich durch statistische Auswertung der gegebenen Antworten nicht unter-

stützen, da der errechnete Korrelationskoeffizient $r = 0,146$ einen sehr schwachen positiven Zusammenhang deutet, welcher nicht signifikant ($p = 0,258$) ist.

Zusammenfassend werden die Hypothesen 1 - 3 durch die Stichprobe gestützt, Hypothese 4 kann abgelehnt werden.

5 Fazit und Ausblick

Es kann mit einigen Einschränkungen postuliert werden, dass eine kennzahlenbasierte Erfolgsmessung von Green IT-Maßnahmen in Unternehmen auf breiter Basis stattfindet. Es gibt dabei eine große Anzahl möglicher Kenngrößen, die sich durch das Modell der Integrierten Informationsmanagements systematisieren lassen, wodurch Forschungsfrage 1 beantwortet wird. Zu beachten ist aber, dass Kennzahlen im Bereich Make am weitesten verbreitet sind. Die Beantwortung der Forschungsfrage 2 ist differenziert zu führen. Die Unternehmen kennen und messen grundsätzlich fast alle genannten Kennzahlen, wobei einzelne Kennzahlen nur selten erfasst und andere, wie der PUE sehr verbreitet sind. Kennzahlen aus dem Bereich Make dominieren auch in der Praxis. Unternehmen mit mehr Mitarbeitern, eigenem Rechenzentrum oder höherer Green IT-Expertise setzen mehr Green IT-Maßnahmen um und messen diese mit Kennzahlen. Diese Beobachtung beantwortet die Forschungsfrage 3.

Bei den genannten Ergebnissen sind einige Einschränkungen zu beachten. Die Datenbasis besteht ausschließlich aus Unternehmen mit Hauptgeschäftssitz in Deutschland. Somit ist die internationale Vergleichbarkeit der Ergebnisse nicht unbedingt gegeben. Ferner fehlen für einen Gesamtüberblick Behörden und Nichtregierungsorganisationen als weitere Entitäten in der Stichprobe. Außerdem basieren die Kennzahlen im Bereich Deliver nur auf Zertifizierungsmöglichkeiten. Für eine gesteuerte Erfolgsmessung in Vertrieb und Kommunikation müssen Konzepte aus dem klassischen Marketing wie der Werbewirkungsforschung für Green IT adaptiert werden und daraus entsprechende Kennzahlen abgeleitet werden.

Dieser Beitrag enthält einige Implikationen für Forschung und Praxis. Es konnte gezeigt werden, dass es im Bereich Make sowohl im Rechenzentrum als auch in der Büroumgebung fortgeschrittene Konzepte zur Erfolgsmessung gibt. Es müssen in erster Linie in den Bereichen „Source“ und Deliver Steuerungsinstrumente für den Erfolg von Green IT entwickelt werden. Für die Praxis bleibt festzustellen, dass Green IT-Kennzahlen ein Instrument zur Erfolgsmessung z. B. im Nachhaltigkeitscontrolling darstellen können.

Für die weitere Forschungsarbeit gilt es herauszufinden, ob die Messung des Erfolgs von Green IT-Maßnahmen mit Hilfe geeigneter Kennzahlen Auswirkungen auf die Unternehmensperformance haben und in welchem Umfang dieses nachgewiesen werden kann.

Literatur

1. Atteslander, P.: Methoden der empirischen Sozialforschung, 13. Auflage, Schmidt, Berlin (2010)
2. Bachour, N., Chasteen, L.: Optimizing the Value of Green IT Projects within Organizations. In: Institute of Electrical and Electronics Engineers (IEEE) (ed.), Proceedings of 2nd Green Technologies Conference, Grapevine, Texas, pp. 1-10 (2010)
3. BearingPoint: CIO Snapshot 2009, <http://www.bearingpointconsulting.com/SID-53D596F1-46B5DFD0/de-de/download/CIO.pdf>
4. Belady, C., Rawson, A., Pflueger, J., Cader, T.: Green Grid Data Center Power Efficiency Metrics PUE and DCIE. The Green Grid, Beaverton (2008)
5. Buchta, D., Eul, M., Schulte-Croonenberg, H.: Strategisches IT-Management – Wert steigern, Leistung steuern, Kosten senken. 3. Auflage, Gabler, Wiesbaden (2009)
6. Destatis (o. V.): Gliederung der Klassifikation der Wirtschaftszweige. Ausgabe 2008, Statistisches Bundesamt, Wiesbaden (2008)
7. Diekmann, A.: Empirische Sozialforschung – Grundlagen, Methoden, Anwendungen. 16. Auflage, Rowohlt, Reinbek (2006)
8. Dumitru, J., Fagarasan, I., Iliescu, S., Hady Said, Y., Ploix, S.: Increasing Energy Efficiency in Data Centers using Energy Management. In IEEE/ACM (ed.), International Conference on Green Computing and Communications, Chengdu, Sichuan, China, pp. 159-165 (2011)
9. Elliot, S.: Transdisciplinary Perspectives on Environmental Sustainability - A Resource Base and Framework for IT-Enabled Business Transformation. MIS Quarterly 35 (1), 197-236 (2011)
10. Experton Group (o. V.): Green IT – DEKRA Certification zertifiziert Green IT, <http://www.experton-group.de/press/releases/pressrelease/article/green-it-dekra-certification-zertifiziert-green-it.html>
11. Fettke, P.: State-of-the-Art des State-of-the-Art - Eine Untersuchung der Forschungsmethode „Review“ innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik 04, 257–266 (2006)
12. Field, A.: Discovering Statistics Using SPSS. 3rd ed., Sage Publications, Los Angeles (2009)
13. GreenIT BB (o. V.): RZ-Benchmarking – Ressourcen- und Energieeffizienz erkennen – Von Best Practice profitieren – Optimierungspotenziale identifizieren. GreenIT BB (eds.), Berlin (2010)
14. Haas, J., Monroe, M., Pflueger, J., Ouchet, J., Snelling, P., Rawson, A., Rawson, F.: Proxy Proposals for measuring Data Center Productivity. The Green Grid (ed.), Beaverton, 2009
15. Harmon, R., Auseklis, N.: Sustainable IT Services: Assessing the Impact of Green Computing Practices. In: Proceedings of the PICMENT 2009, pp. 1707-1717. IEEE (2009)
16. Hochstein, A., Brenner, W., Uebernickel, F.: Operations Management and IS: Using the SCOR-Model to Source, Make, and Deliver IS Services. In: Proceedings of the 12th Americas Conference on Information Systems (AMCIS 2006), Acapulco, Mexico (2006)
17. Huttenloher, R., Zertifizierung von Rechenzentren – Energieeffizienz darf keine Momentaufnahme sein – Zertifizierungsaufwand rentiert sich schon binnen eines Jahres. Rechenzentren und Infrastruktur 1, 10-12 (2009)

18. Köhn, M.: Ein neuer Blauer Engel ist geboren – das Umweltzeichen „Energiebewusster Rechenzentrumsbetrieb“ (RAL-ZU 161). In: greenletter – Informationen zur Green-IT-Initiative des Bundes Ausgabe 3, 5-6 (2012)
19. Loos, P.: Green IT: Ein Thema für die Wirtschaftsinformatik?. *Wirtschaftsinformatik* 53 (4), 239-247 (2011)
20. Molla, A.: The Reach And Richness Of Green IT: A Principal Component Analysis. In: *Proceedings of the 20th Australasian Conference on Information Systems*, Melbourne, Australia, pp. 754-764 (2009)
21. Patterson, M., Tschudi, B., Vangeet, O., Cooley, J., Azevedo, D.: ERE: A Metric for Measuring the Benefit of Reuse Energy from a Data Center. *The Green Grid*, Beaverton (2010)
22. Patterson, D.A., Hennessy, J.L.: *Computer Organization and Design - The Hardware/Software Interface*. 4th ed., Morgan Kaufmann Publishers, Waltham (2011)
23. Pichler, M.: *Nachhaltige IT im Rechenzentrum – Entwicklung und Darstellung eines Modells zur Messbarkeit von Effizienz im Rechenzentrum*. Diplomica, Hamburg (2009)
24. Stanley, J.R., Brill, K.G.: *Four Metrics Define Data Center “Greenness”: Enabling users to quantify energy consumption initiatives for environmental sustainability and “bottom line” profitability*. Uptime Institute (ed.), Santa Fe (2007)
25. The Green Grid (o. V.), *Green Grid Metrics: Describing Datacenter Power Efficiency – Technical Committee*. White Paper, The Green Grid (eds.), Beaverton (2007)
26. T-Systems (o. V), *White Paper – Nachhaltigkeit und ICT – Der Weg zum Grünen Business*. T-Systems (eds.), Frankfurt am Main (2011)
27. Watson, R.T., Boudreau, M.C., Chen, A.J.: *Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community*. *MIS Quarterly* 34 (1), 22-38 (2010)
28. Wilkens, M.: *Data Center Benchmarking und Blauer Engel für Rechenzentren*. Technische Universität Berlin - Fakultät für Wirtschaft und Management (eds.), Berlin (2011)
29. Zarnekow, R., Brenner, W., Pilgram, U.: *Integriertes Informationsmanagement: Strategien und Lösungen für das Management von IT-Dienstleistungen*. Springer, Berlin (2005)
30. Zarnekow, R., Kolbe, L.M., Ere, K., Schmidt, N.-H.: *Studie: Nachhaltigkeit und Green IT in IT-Organisationen – Status quo und Handlungsempfehlungen*. In: Zarnekow, R.: *Research Papers in Information Systems Management*. Band 1, Berlin (2010)

Modellgestützte Analyse und Optimierung der Energieeffizienz betrieblicher Informations- und Kommunikationstechnik

Steffen Vaupel und Jörg Leukel

Universität Hohenheim, Lehrstuhl Wirtschaftsinformatik 2, Stuttgart, Germany
{steffen.vaupel, joerg.leukel}@uni-hohenheim.de

Abstract. Die Energieeffizienz von Informationssystemen (IS) wird zu einer wichtigen Kennzahl für Entscheider im Informationsmanagement. Derzeitige Unternehmensarchitekturmodelle (EAM) unterstützen allerdings nur unzureichend die automatisierte Bestimmung der Energieeffizienz und die Bewertung von Maßnahmen zur Erhöhung der Energieeffizienz. Der Grund liegt darin, dass diese Modelle auf die Planung und Dokumentation von Informationssystemen ausgerichtet sind und hierzu die energieeffizienzrelevanten Informationen nicht ausreichend erfasst werden. Folglich ist es notwendig, derzeitige EAM-Ansätze um eine Methode zur Energieeffizienzanalyse und -optimierung zu erweitern. Die vorliegende Arbeit schlägt eine Methode für die modellgestützte Energieeffizienzanalyse und -optimierung im Rahmen von EAM vor. Die Evaluation durch ein Simulationsexperiment mit einem Prototyp zeigt die Wirksamkeit und Nützlichkeit der entwickelten Methode.

Keywords: Domänenspezifische Modellierungssprachen (DSML), Green-IT, Simulation, Unternehmensarchitekturmodelle (EAM)

1 Einführung

Die Energieeffizienz von Informationssystemen erfährt zunehmende Aufmerksamkeit des IT-Managements (Green-IT). Hierzu tragen vor allem der angestiegene Gesamtenergieverbrauch der IT und die dadurch entstehenden Energiekosten bei. Zusätzlich werden IT-Ressourcen mittlerweile als eine von drei Hauptfaktoren für die Entwicklung von nachhaltigen Unternehmensstrategien angesehen [1]. Maßnahmen zu Green-IT spannen ein weites Handlungsfeld auf und betreffen sowohl IT-basierte Sachgüter und Dienstleistungen, Geschäftsprozesse, Planungs- und Steuerungsaufgaben in Unternehmen als auch die IT-Ressourcen im engeren Sinne [2]. Letztere können als Menge der unmittelbaren IT-Energieverbraucher Gegenstand von vielfältigen Green-IT-Maßnahmen sein (z. B. Serverkonsolidierung, dynamische Lastverteilung und Rechenzentrumsoptimierung unter Einbeziehung von Klimatisierung und Stromversorgung). Diese IT-Ressourcen sind Teil von betrieblichen IS und wir bezeichnen sie

im Folgenden mit betrieblich genutzter Informations- und Kommunikationstechnik (IKT).

Green-IT-Strategien für betrieblich genutzte IKT basieren häufig auf Zielen zur Reduzierung des unternehmensweiten IKT-Energieaufwands. In der Typologie von Green-IT-Strategien gehört diese Strategie zur Vermeidung, Steuerung und Ökoeffizienz [3]. Notwendige Voraussetzung ist die Ermittlung der Energieeffizienz von gegebener oder geplanter IKT. Die Energieeffizienz ist das Verhältnis von Nutzen zu dem Energieaufwand einer Periode. Wird der Energieaufwand ermittelt und bei gleichbleibendem Nutzenniveau reduziert, so erhöht sich die Energieeffizienz. Während der Energieaufwand durch Messung oder Schätzung relativ gut ermittelt werden kann, ist dies für den Nutzen schwieriger: Der Nutzen entsteht durch die Nutzung der IKT für betriebliche Aktivitäten eines oder mehrerer Geschäftsprozesse. Aufgrund der Schichtenarchitektur von IS z. B. Prozess-, Software-, Hardwareschicht, ist es nicht unmittelbar möglich, den auf der Prozessschicht entstehenden Nutzen den Elementen der IKT verursachungsgemäß zuzurechnen. Zudem ist der Nutzen, welchen die Anwender erzielen, von einer eher abstrakten Natur. Ein direktes Maß des Nutzens kann nicht abgeleitet werden. Infolgedessen ist eine vertikale Effizienzberechnung aus Nutzen und Energieaufwand kaum durchführbar.

Aufgrund der Zuordnungsproblematik sind schichtenspezifische Ansätze erfolgversprechender. Daher wird in dieser Arbeit die Analyse und Optimierung auf die IKT eingeschränkt, die unter Aufrechterhaltung eines vorgegebenen Nutzens gleichmäßig ausgelastet ist. Dann kann eine relative Bewertung der IKT-Objekte hinsichtlich ihrer energetischen Attribute erfolgen. Zielsetzung der Arbeit ist die (1) Entwicklung einer Methode zur modellgestützten Energieeffizienzanalyse und –optimierung im Rahmen von EAM und die (2) Überprüfung der Methode mit Hilfe eines Prototyps, um die Wirksamkeit und Nützlichkeit der entwickelten Methode aufzuzeigen. Im Sinne der gestaltungsorientierten Forschung ist das Artefakt die entwickelte Methode, die zur Evaluation in einem Simulationsexperiment verwendet wird.

Die Intention der Methode ist es, dass IT-Management als Zielgruppe bei operativen Entscheidungen zu unterstützen. Die Optimierungsschritte entsprechen potentiellen operationalen Schritten bzw. stellen eine Auswahl von Handlungsempfehlungen dar. Die Methode dient zunächst der Optimierung der betrieblich genutzten IKT, die lokal beim Anwenderunternehmen betrieben wird. Ein weiteres Einsatzszenario besteht in der Migration von Anwendungen zwischen zwei IKT, z. B. zwischen der lokalen IKT (on-premise) und der zentralen IKT eines Cloud-Anbieters (off-premise). Aus den Energieeffizienzangaben der Cloud-Anbieter (z. B. Power Usage Effectiveness, PUE) ist es möglich, homogene Vergleichsmodelle zu konstruieren. Die entwickelte Methode ist folglich analog beim Cloud-Anbieter anzuwenden und liefert dadurch Vergleichswerte für eine alternative Platzierung von Anwendungen.

Der Beitrag ist wie folgt gegliedert: In Kapitel 2 werden derzeitige EAM-Ansätze auf Eignung untersucht. In Kapitel 3 wird ein Modell betrieblich genutzter IKT eingeführt. Für dieses Modell werden in Kapitel 4 Analysefunktionen und in Kapitel 5 Optimierungsfunktionen vorgeschlagen und formal definiert. Die Evaluation erfolgt in Kapitel 6. Eine Schlussfolgerung wird in Kapitel 7 gezogen.

2 Verwandte Arbeiten

Modelle zur Beschreibung betrieblich eingesetzter IKT müssen für Analysezwecke eine formale Syntax besitzen, energiebedarfsrelevante Attribute beinhalten und energierelevante Auswertungsfunktion (z. B. Energiemetriken) bereitstellen oder zumindest die Erweiterung um solche Funktionen ermöglichen. Zum Zweck der Optimierung müssen Modifikationen der Instanzen möglich sein und zu einer automatischen Neuberechnung der Analyseergebnisse führen.

Die formale Syntax ist grundsätzlich notwendig um die Validität der Modellinstanzen zu gewährleisten bzw. zu überprüfen. Die automatisierte Instanzanalyse benötigt zudem formale Modelle. Die energiebedarfsrelevanten Attribute sowie Auswertungsfunktionen können, sofern existierende Ansätze zur IKT-Modellierung dies unterstützen, generisch spezifiziert werden. Nicht geeignet sind jedoch Ansätze, die ausschließlich Planungs- und Dokumentationsaufgaben unterstützen. Für die Energieeffizienzanalyse ist ein Arbeitsmodell notwendig, das iterativ verbessert und automatisch analysiert werden kann. Zusammenfassend bestehen die Anforderungen aus Designanforderungen des IKT Modells (formal, energiebedarfsrelevante Attribute und Auswertungsfunktionen) und Laufzeitanforderungen (Berechnung der Auswertungsfunktionen und Modifikationen).

Der Suchraum für Ansätze zu IKT-Modellierung wird zunächst auf die gesamte Unternehmensarchitektur (EAM) ausgeweitet, da die IKT nur einen Teil dieser darstellt. Es sind weit über siebzig Ansätze bekannt [4]. Hierzu zählen vor allem das Zachman-Framework [5], das Federal Enterprise Architecture Framework (FEAF) [6], das Treasury Enterprise Architecture Framework (TEAF) [7], The Open Group Architecture Framework (TOGAF) [8], Architecture of Integrated Information Systems (ARIS) [9], Multi-perspective Enterprise Modeling (MEMO) [10], ArchiMate [11], Semantic Object Model (SOM) [12] und ADOben [13]. Eine generelle Referenz für den Entwurf solcher Rahmenwerke kann der Generalised Enterprise Reference Architecture and Methodology (GERAM) [14] entnommen werden.

In [15] wird über die Relevanz von EAM in Unternehmen berichtet. Hierzu werden der Abbildungsbereich von EAM, die verfügbaren Analysetechniken und die unterstützten Szenarien und Benutzergruppen untersucht. Die Ergebnisse zeigen einen beträchtlichen Unterschied zwischen den Anforderungen und Erwartungen der Anwender gegenüber den existierenden EAM-Ansätzen. Insbesondere bei der automatisierten Analyse und Transformation von erstellten EAM liegt die Umsetzung noch weit hinter den Erwartungen der Benutzer zurück. Im Einzelnen bedeutet dies, dass Modellierungssprachen und Softwarewerkzeuge für die Transformation von EAM selbst weiterzuentwickeln sind. Die Ermittlung und Bewertung möglicher Soll-EAM und zugehöriger Entwicklungspfade sind derzeit nur in Ansätzen vorhanden.

Das Multi-perspective Enterprise Modeling (MEMO) Framework berücksichtigt bereits viele der gestellten Anforderungen. Dieser Ansatz vereinigt verschiedene Modelltypen im Zuge einer multiperspektivischen Unternehmensmodellierung. In Anbetracht der gewählten Zielsetzung ist das Teilmodell MEMO-OML (Object Modeling Language) von besonderem Interesse. MEMO-OML ist formal über ein (Meta-)Metamodell definiert, enthält typisierte Attribute und eine visuelle Repräsentation

[16]. Damit ist MEMO-OML eine domänenspezifische (visuelle) Sprache ((V)DSL). Dieser konzeptionelle Ansatz lässt sich für das angestrebte Ziel nutzen.

Ein weiterer Suchraum für Ansätze zur Modellierung der IKT hinsichtlich der Energieeffizienzanalyse und Optimierung bieten die *Betrieblichen Umweltinformationssysteme* (BUIS). Praktisch werden BUIS selbständig implementiert, in andere Informationssysteme integriert oder an vorhandene Informationssysteme angegliedert. Für viele Bereiche eines Unternehmens fehlen wesentliche umweltrelevante Daten (z. B. der IKT) als Datengrundlage für die Konstruktion eines BUIS. Ansätze für BUIS, die eine maschinelle Analyse und Transformation von IKT-Modellen ermöglichen, konnten nicht identifiziert werden.

Basierend auf der Analyse der bestehenden Arbeiten, insbesondere des Multi-perspective Enterprise Modeling (MEMO) Frameworks, wird der deklarative Ansatz für IKT-Modelle gewählt. Während in MEMO-OML ein eigenes (Meta-)Metamodell spezifiziert wird, soll das Metamodell der IKT-Modelle eine Instanz eines bereits gegebenen (Meta-)Metamodells sein (EMOF/Ecore). Zur Analyse und Optimierung sollen Methoden aus dem modellbasierten Software Engineering adaptiert werden. Hierbei handelt es sich im Einzelnen um Modellqualitätskriterien (Metriken) und Modelltransformationsregeln (Modell-Refactoring).

3 Physikalisches Architekturmodell (PAM)

Das Modell der IKT wird in diesem Abschnitt als deklaratives Modell beschrieben, gefolgt von einer alternativen Beschreibung als graphbasiertes mathematisches Modell. Das PAM ist ein Metamodell. PAM-Instanzen sind konkrete Modelle von IKT.

3.1 Deklaratives Modell

Dem Ansatz folgend wird ein deklaratives Metamodell der IKT als Beschreibung von physischen Architekturen (PAM) definiert. Das PAM Metamodell enthält sieben Entitäten: Room (R), Uninterruptible Power Supply (U), Cooling (T), Client Node (C), Server Node (S), Network Node (N) und Network Object Link (L). Abbildung 1 zeigt die Entitäten, ihre Assoziationen, Aggregationen, Vererbungsstrukturen und Kardinalitäten. Jede Entität (mit der Ausnahme von Room und Network Object Link) lässt sich energetisch sortieren d. h. für zwei typgleiche Instanzen kann das energieeffizientere Gerät ermittelt werden. Das PAM Metamodell stellt wiederum eine Instanz eines (Meta-)Metamodells dar. Dabei handelt es sich um das Ecore-Metamodell als Implementierung der (Essential) Meta Objects Facility ((E)MOF).

Entscheidend für die Auswahl des Ecore-Metamodells ist die modellgetriebene Entwicklungsunterstützung des Eclipse Modeling Frameworks (EMF). Ausgehend von dem Metamodell der IKT als Instanz des Ecore-Metamodells kann direkt domänenspezifischer Modell-Code insbesondere für Modelleditoren generiert werden.

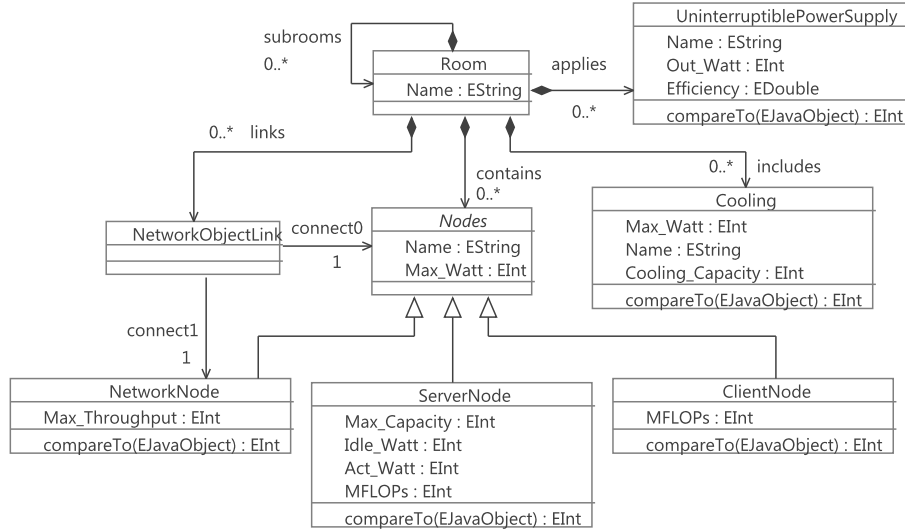


Abb. 1. PAM-Metamodell

3.2 Mathematisches Modell

Eine alternative Definition, die bei der Modellanalyse hilfreich ist, erfolgt als mathematisches Modell. Dabei werden die Modellelemente auf einen Graphen abgebildet.

$$G = (V, L, (connect^i)^{i=0,1}) \text{ mit } connect^0: L \rightarrow V_{CSN}, connect^1: L \rightarrow N \quad (1)$$

$$V = R \cup T \cup U \cup V_{CSN} \text{ mit } V_{CSN} = C \cup S \cup N \quad (2)$$

$$r \in R = ((Name \times String) \times (subrooms \times (\wp(R) \setminus r)) \times (links \times \wp(L)) \times (contains \times \wp(V_{CSN})) \times (includes \times \wp(T)) \times (applies \times \wp(U))) \quad (3)$$

$$t \in T = ((Name \times String) \times (Max_Watt \times Integer) \times (Cooling_Capacity \times Integer)) \quad (4)$$

$$u \in U = ((Name \times String) \times (Out_Watt \times Integer) \times (Efficiency \times Double)) \quad (5)$$

$$c \in C = ((Name \times String) \times (Max_Watt \times Integer) \times$$

$$(MFLOPs \times Integer) \tag{6}$$

$$s \in S = ((Name \times String) \times (Max_Watt \times Integer) \times (MFLOPs \times Integer) \times (Max_Capacity \times Integer) \times (Idle_Watt \times Integer) \times (Act_Watt \times Integer)) \tag{7}$$

$$n \in N = ((Name \times String) \times (Max_Watt \times Integer) \times (Max_Throughput \times Integer)) \tag{8}$$

$$l \in L \tag{9}$$

3.3 Implementierung

Durch das Graphical Modeling Framework (GMF) besteht die Möglichkeit einen visuellen Modelleditor zur Ansicht und Modellierung von PAM-Instanzen zu generieren. In Abbildung 2 ist eine visualisierte PAM-Instanz dargestellt. Die Definition einer PAM-Instanz kann direkt in dem visuellen Editor erfolgen. Das Softwaresystem hält das Diagramm (Visualisiere PAM-Instanz) und das logische Modell (PAM-Instanz) konsistent.

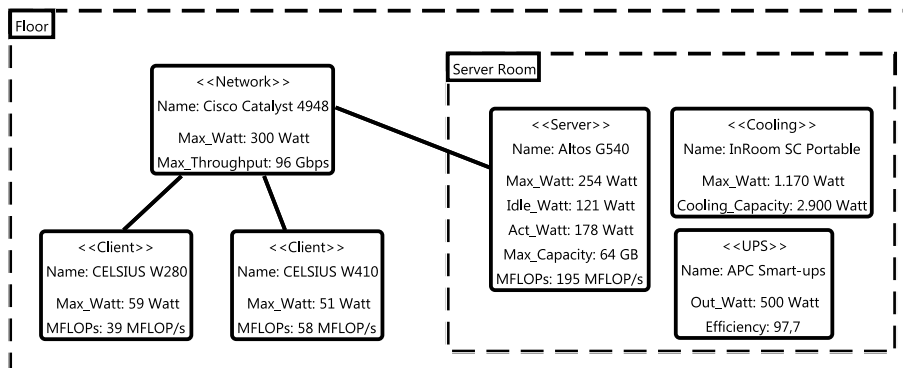


Abb. 2. Visualisierte PAM-Instanz (Beispiel)

4 Analysefunktionen

Die Analyse der PAM-Instanzen erfolgt über die Auswertung von Metrikfunktionen (oder verkürzt Metriken). Die Metriktypen werden vor der Definition der Metriken eingeführt. Das Ergebnis einer Metrik wird durch eine generische Statistikfunktion in die Klassen *Unterdurchschnittlich*, *Regulär*, *Überdurchschnittlich* eingestuft. So lassen sich abweichende Teile der Modellinstanz identifizieren. Abschließend wird auf die Implementierung der definierten Metriken eingegangen.

4.1 Metriktypen

Vier Metriktypen lassen sich nach Guerra et al. [17] unterscheiden: Modell-, Element-, Gruppen- und Pfadmetriken. Dabei sind die mit ‚Energy Efficient Ratio‘, ‚Operations per Watt‘, ‚Energy Consumption Ratio‘, ‚Capacity per Watt‘ und ‚Charge retention loss‘ bezeichneten Metriken Elementmetriken. Eine Modellmetrik stellt die ‚Power usage effectiveness‘ Metrik dar. Gruppen- und Pfadmetriken werden gegenwärtig nicht definiert. Die Resultate der Elementmetriken lassen sich anhand der Statistikfunktion einstufen.

4.2 Definition der Metriken

1. Energy Efficient Ratio (EER)

Die EER-Metrik ist das Verhältnis zwischen der maximal absorbierten Wärmeenergie (Cooling_Capacity) und der maximalen Leistungsaufnahme (Max_Watt) einer Klimakomponente. Die Metrik kann nur auf Klimakomponenten angewendet werden.

$$\begin{aligned} \pi_{EER} : T &\rightarrow \underline{Double} \\ \pi_{EER}(t) &= t.Cooling_Capacity/t.Max_Watt \end{aligned} \quad (10)$$

2. Operations per Watt

Die Operationen pro Watt stellen das Verhältnis der prozessorspezifischen maximalen Anzahl von Rechenoperationen (MFLOPs) und der maximalen Leistungsaufnahme (Max_Watt) dar. Die Metrik kann auf Server und Clients angewendet werden.

$$\begin{aligned} \pi_{MFLOPs/Max_Watt} : (C \cup S) &\rightarrow \underline{Double} \\ \pi_{MFLOPs/Max_Watt}(i) &= i.MFLOPs/i.Max_Watt \end{aligned} \quad (11)$$

3. Energy Consumption Ratio (ECR)

Die ECR-Metrik ist das Verhältnis zwischen dem maximalen Datendurchsatz (Max_Throughput) und der maximalen Leistungsaufnahme (Max_Watt) einer Netzwerkkomponente. Die Metrik kann nur auf Netzwerkkomponenten angewendet werden.

$$\begin{aligned} \pi_{Max_Watt/Max_Throughput} : N &\rightarrow \underline{Double} \\ \pi_{Max_Watt/Max_Throughput}(n) &= n.Max_Watt/n.Max_Throughput \end{aligned} \quad (12)$$

4. Capacity per Watt

Die Kapazität pro Watt stellt das Verhältnis der maximalen Speicherkapazität (Max_Capacity) und der minimalen Leistungsaufnahme (Idle_Watt) dar. Die Metrik kann nur auf Server angewendet werden.

$$\begin{aligned} \pi_{Max_Capacity/Idle_Watt}: S &\rightarrow \underline{Double} \\ \pi_{Max_Capacity/Idle_Watt}(s) &= s.Max_Capacity/s.Idle_Watt \end{aligned} \quad (13)$$

5. Charge Retention Loss (CRL)

Die CRL-Metrik beziffert die Ladeerhaltungsenergie einer unterbrechungsfreien Stromversorgung (USV). Die Metrik kann nur auf unterbrechungsfreie Stromversorgungen angewendet werden.

$$\begin{aligned} \pi_{CRL}: U &\rightarrow \underline{Double} \\ \pi_{CRL}(u) &= ((u.Out_Watt/u.Efficiency) * 100) - u.Out_Watt \end{aligned} \quad (14)$$

6. Power Usage Effectiveness (PUE)

Die PUE-Metrik ist das Verhältnis des gesamten maximalen Energiebedarfs und dem Energiebedarf, der ausschließlich von Informations- und Kommunikationskomponenten benötigt wird. Die Metrik kann nur auf Räume angewendet werden.

$$\begin{aligned} \pi_{PUE}: R &\rightarrow \underline{Double} \\ \pi_{PUE}(r) &= (\sum_{r_j \in r.(subrooms)^*} (\sum_{i \in r_j.contains()} i.Max_Watt \\ &\quad + \sum_{i \in r_j.includes()} i.Max_Watt \\ &\quad + \sum_{i \in r_j.applies()} \pi_{CRL}(i))) \\ &\quad / (\sum_{r_j \in r.(subrooms)^*} (\sum_{i \in r_j.contains()} i.Max_Watt)) \end{aligned} \quad (15)$$

4.3 Generische Statistikfunktion

Die generische Statistikfunktion stuft das Ergebnis einer Metrik anhand der mittleren absoluten Abweichung e vom Mittelwert $AVG(\Phi, M)$ aller Metriken dieser Typen ein. Die generische Statistikfunktion ist wie folgt definiert:

$$AVG(\Phi, M) = (\sum_{m \in M} \Phi) / \#M \quad (16)$$

$$e = \frac{1}{\#M} \sum_{m \in M} |\Phi(m) - AVG(\Phi, M)| \quad (17)$$

$$\min(\Phi, M) = AVG(\Phi, M) - (e/2) \quad (18)$$

$$\max(\Phi, M) = AVG(\Phi, M) + (e/2) \quad (19)$$

Dabei bezeichnet Φ die Metrik und M die Grundmenge der Metrik.

4.4 Implementierung.

Die Auswahl des Ecore-Metamodells zur Definition des PAM-Metamodells bietet die Möglichkeit von der Object Constraint Language (OCL) zur PAM-Instanzprüfung Gebrauch zu machen. Ein entsprechendes Laufzeitsystem ist im EMF/GMF vorhan-

den. Die Implementierung der Metriken in OCL erfolgt analog zu der mathematischen Definition (siehe Tabelle 1). Die Implementierung der generischen Statistikfunktion erfolgt analog mit OCL. Abbildung 3 zeigt eine visuell aufbereitete Modellanalyse in tabellarischer Form. Die Ergebnisse der Metriken werden dabei gemäß der Klassifikation gefärbt.

Object	Metric	Value	Min	Max	Unit
Datcenter (Diagram)	PUE	2.7798	n/a	n/a	n/a
Floor (Room)					
Cisco Catalyst 4948 (Network)	ECR	3.125	3.125	3.125	Watt/Gbps
CELSIUS W280 (Client)	Operations per Watt (Max)	0.661	0.7801	1.0182	MFLOPs/Watt
CELSIUS W410 (Client)	Operations per Watt (Max)	1.1373	0.7801	1.0182	MFLOPs/Watt
Server Room (Room)					
InRoom SC Portable (Cooling)	EER	2.4786	2.4786	2.4786	Cooling_Capacity/Watt
APC Smart-ups (UPS)	Charge retention loss	11.7707	11.7707	11.7707	Watt
Altos G540 (Server)	Operations per Watt (Max)	0.7677	0.7677	0.7677	MFLOPs/Watt
	Capacity per Watt (Idle)	0.5289	0.5289	0.5289	Capacity/Watt

Abb. 3. Visuell aufbereitete Analyse einer PAM-Instanz

Tabelle 1. Implementierung der Metriken

Metrik	OCL-Ausdruck
π_{EER}	<code>self.Cooling_Capacity / self.Max_Watt</code>
π_{MFLOPs/Max_Watt}	<code>self.MFLOPs / self.Max_Watt</code>
$\pi_{Max_Watt/Max_Throughput}$	<code>self.Max_Watt / self.Max_Throughput</code>
$\pi_{Max_Capacity/Idle_Watt}$	<code>self.Max_Capacity / self.Idle_Watt</code>
π_{CRL}	<code>((self.Out_Watt/self.Efficiency)*100) - self.Out_Watt</code>
π_{PUE}	<code>(self -> closure(subrooms).includes.Max_Watt -> sum() + self -> closure(subrooms).applies -> iterate (w; var:Real = 0 var + ((w.Out_Watt/w.Efficiency)*100)-w.Out_Watt))+ self -> closure(subrooms).contains.Max_Watt -> sum()) / self -> closure(subrooms).contains.Max_Watt -> sum()</code>

5 Optimierungsfunktionen

Die automatische Analyse eines IKT-Modells erlaubt es, Bestandteile mit geringer Energieeffizienz zu identifizieren. Der Anwender hat die Möglichkeit durch manuelle Transformation d. h. durch Veränderung der PAM-Instanz eine Simulation der Auswirkungen der vorgenommenen Änderungen durchzuführen. Dabei nutzt er die automatische Analyse wiederholt, um die Wirksamkeit seiner manuellen Änderungen zu bewerten. Die automatische Transformation stellt dem Benutzer einen Katalog vorgefertigter Transformationsregeln zur Verfügung.

5.1 Manuelle Transformation

Die manuelle Transformation kann als triviale Optimierungsmöglichkeit genutzt werden. Bei dieser Verfahrensweise stehen dem Benutzer alle Modifikationsmöglichkeiten der PAM-Instanz zur Verfügung. Ob die Modifikationen tatsächlich eine Optimierung der Energieeffizienz generiert haben, zeigt sich allerdings erst nach einer erneuten Modellanalyse. Der Anwender hat bei der manuellen Transformation die meisten Freiheitsgrade, muss aber den optimierenden Effekt seiner Modifikationen durch eine folgende Analyse sicherstellen.

5.2 Automatische Transformation

Die automatische Transformation verwendet ein Transformationssystem und einen Katalog von Metaregeln zur Optimierung. Die Regeln leiten sich aus den Operationen her, die im Rahmen des IKT-Managements (z. B. Beschaffung, Ablösung, Konsolidierung) vollzogen werden. Die Anforderung an die Metaregeln in dem Regelkatalog besteht in einer tatsächlichen Reduzierung des Energiebedarfs. Neben dieser *Optimierungsanforderung* muss außerdem die *Funktionsanforderung* weiterhin Gültigkeit behalten. Eine *Funktionsanforderung* kann sich auf den verfügbaren Speicher oder die Rechenleistung eines Gerätes beziehen.

Die Abbildung 4 zeigt eine Menge von Geräten (Knoten) sowie Optimierungs- und Funktionsanforderungen (Kanten) modelliert als Graph (O/F Graph). Dabei stellt der Kantentyp $\omega(c_i, c_j) = -1$ die energetische Ordnungsfunktion dar (Vergleiche Definition 20). Besteht zwischen zwei Geräten c_i, c_j eine Kante $\omega(c_i, c_j) = -1$ (O-Kante) so benötigt das Gerät c_i weniger Energie als das Gerät c_j . Existiert zusätzlich eine entgegengesetzte Kante (F-Kante) zwischen c_j und c_i so ist die *Funktionsanforderung* hinsichtlich eines Attributes oder einer Attributmenge erfüllt. Der Austausch von c_j durch c_i liefert eine äquivalente energieoptimierte PAM-Instanz.

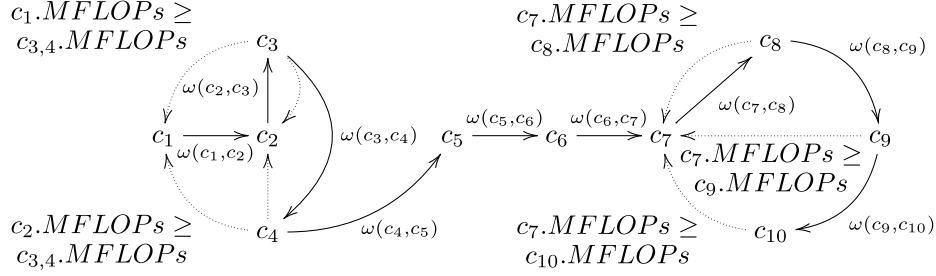


Abb. 4. Optimierungs- und Funktionsanforderungen (O/F Graph)

$$\omega_c(c_i, c_j) \mapsto \begin{cases} -1 & c_i.Max_Watt < c_j.Max_Watt \\ 0 & c_i.Max_Watt = c_j.Max_Watt \\ 1 & \omega_c(c_j, c_i) = -1 \end{cases} \quad (20)$$

In Abbildung wird das Gerät c_3 von den Geräten c_2 und c_1 sowohl optimiert als auch funktional abgedeckt. Um die Optimierung zu maximieren wird das nach der Ordnungsfunktion ω kleinste Gerät (c_1) gewählt.

Für beliebige PAM-Instanzen werden die O/F-Graphen dynamisch konstruiert. Fügt ein Anwender einen neuen Gerätetyp in die PAM-Instanz ein, so wird diese automatisch erfasst. Der Katalog von Metaregeln zur Optimierung setzt sich daher zunächst nur aus abstrakten Regeln zusammen, die für eine gegebene PAM-Instanz entsprechend instanziiert werden. Eine abstrakte Ersetzungsregel für einen Client c_i durch c_j (im Zeichen $[c_i \setminus c_j]$) ist in Definition 21 zu sehen.

$$\begin{aligned} & c_i, c_j \in C ([c_i \setminus c_j]) \\ & (\omega(c_j, c_i) = -1 \wedge c_j.MFLOPs \geq c_i.MFLOPs) \Rightarrow \\ & \nexists c_k \in (\omega(c_k, c_j) = -1 \wedge c_k.MFLOPs \geq c_i.MFLOPs) \end{aligned} \quad (21)$$

Neben *Ersetzungsregeln* können Instanzen geteilt, vereint und gelöscht werden, ohne dass funktionale Einschränkungen der PAM-Instanz entstehen. Für die Instanziierung der *Teilungs-*, *Vereinigungs-* und *Löschregeln* sind weitere Funktionsanforderungen zu erfüllen. Diese Funktionsanforderungen beziehen sich auf topologische Eigenschaften z. B. den Verknüpfungsgrad einer Netzwerkkomponente. Der Katalog besteht aus insgesamt 15 *Ersetzungs-*, *Teilungs-*, *Vereinigungs-* und *Löschregeln*. In Abbildung 5 ist eine instanziierte visualisierte *Teilungsregel* dargestellt. Aufgrund der Optimierungsanforderung ist die Menge der anwendbaren Regeln endlich.

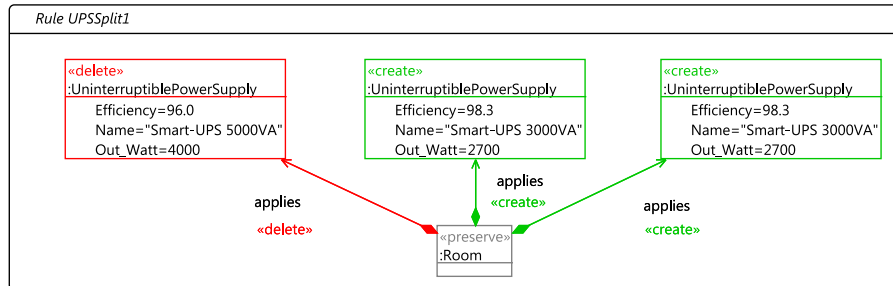


Abb. 5. Visualisierte Teilungsregel für eine Stromversorgungskomponente

6 Evaluation

Die Evaluation der Methode für die Analyse und Optimierung der Energieeffizienz wird mit dem Ziel geführt, die Wirksamkeit der iterativen Optimierungsschritte zu ermitteln. Zusätzlich zeigt die Simulation grundsätzlich die Anwendbarkeit der Methode durch den Prototypen und den Nutzen für IT-Entscheider. Bei der Datenerhebung werden Gerätetypen heutiger IKT ausgewählt. Die Datenerhebung erzeugt eine Mindestmenge von Instanzen der Modellentitäten. Dadurch wird definiert *was* sich in einer PAM-Testinstanz befindet. Anschließend muss beschrieben werden *wie* sich die Komponenten zusammenfügen (Topologie). Hierfür werden Testtopologien definiert, da Topologieerhebungen aufgrund der mangelnden Datenquellen zu nicht repräsentativen Testinstanzen führen. Die generierten Testinstanzen fügen bereits erhobene realweltliche Instanzen in einer parametrisierten Topologie zusammen. Ergebnis dieser beiden Schritte sind valide PAM-Instanzen. Für diese können Optimierungsschritte durchgeführt werden. Die bereits beschriebene Modellanalyse liefert Testergebnisse in Form von Differenzauzeichnungen zwischen den Optimierungsschritten.

6.1 Datenerhebung

Für die Entitäten *Uninterruptible Power Supply* (U), *Cooling* (T), *Client Node* (C), *Server Node* (S), *Network Node* (N) wurden bei der Datenerhebung die entsprechenden Kennzahlen erhoben. Die Daten stammen aus verschiedenen Datenquellen und Datenbanken. Für die Entitäten *Network Node* (N), *Uninterruptible Power Supply* (U) und *Cooling* (T) wurden Herstellerangaben ausgewertet. Bei der Erhebung der Instanzen des Typs *Server Node* (S) wurden unabhängige Messungen der Standard Performance Evaluation Corporation (SPEC) ausgewertet. Zur vollständigen Wertbelegung der Attribute war die Auswertung des Energieperformancetests SPECpower 2008 [18] sowie des Performancetests SPEC CPU 2006 [19] notwendig. Dabei konnten nur Gerätekonfigurationen berücksichtigt werden, die in beiden Testdatenbanken geführt werden. Bei der Datenakquisition der Instanzen des Typs *Client Node* (C) wurde ebenfalls SPEC CPU 2006 für die Performanceinformationen und die Messmethode sowie Ergebnisse der EU Energy Star Kommission [20] für die Energiebedarfsinfor-

mation herangezogen. Dabei können ebenfalls nur Geräte berücksichtigt werden, die in beiden Testdatenbanken verzeichnet sind. Insgesamt wurden 10 Instanzen vom Typ *Client Node* (C), 20 Instanzen vom Typ *Server Node* (S), 15 Instanzen vom Typ *Network Node* (N), 10 Instanzen vom Typ *Uninterruptible Power Supply* (U) und 10 Instanzen vom Typ *Cooling* (T) für die Evaluation ausgewählt.

6.2 Testinstanzen

Die generierten PAM-Testinstanzen verwenden die erhobenen Instanzen der einzelnen Modellentitäten bzw. fügen diese in einer Topologie zusammen. Die PAM-Testinstanzen werden auf Grund des Umfangs von einem Generator erzeugt, der die zuvor erhobenen Einzelinstanzen sowie einen Satz von Argumenten erhält. Die Argumente erlauben die Variation der Modelle in drei definierten Dimensionen. Diese Dimensionen lauten *Heterogenität*, *Quantität* und *Assoziativität*.

Die Dimension *Heterogenität* variiert indirekt die Größe eines O/F Graphen (Abbildung 4) und bestimmt dadurch den Umfang der Anwendung von Ersetzungsregeln. Verfügt die abgebildete IKT nicht über ausreichend heterogene Elemente so können der Modellinstanz weitere beliebige (potenziell einsetzbare) Elemente hinzugefügt werden. Die Dimension *Quantität* erlaubt eine Variation der Anzahl von Instanzduplikaten. Skalierungseffekte und –muster werden dadurch erkennbar. Die Dimension der *Assoziativität* bestimmt den Grad der Kopplung und variiert damit die Anzahl der topologisch abhängigen Regelinstanzierungen.

Diese Dimensionen wurden gewählt, um die notwendige Repräsentanz der Testinstanzen sicherzustellen, da eine vollständige Erhebung von Anwenderinfrastrukturen insbesondere der Topologie nur bedingt durchführbar ist.

6.3 Testergebnisse

Anhand der generierten Testinstanzen wurden die in Kapitel 5 beschriebenen Optimierungen durchgeführt. Die Ergebnisaufzeichnung schließt die Energiebedarfsveränderung für den primären Energiebedarf einer IKT (d. h. *Client Node*, *Server Node*, *Network Node*) sowie den sekundären Energiebedarf (d. h. *Uninterruptible Power Supply* und *Cooling*) ein. Ferner wird die Anzahl der instanziierten Regeln und die klassifizierten Metriken in aggregierte Form aufgezeichnet. Die Abbildung 6 zeigt eine Aufzeichnung der Optimierung für eine PAM-Instanz mit 150 Clients, 60 Servern und 30 Netzwerkknoten. Die Optimierung terminiert nach 185 Optimierungsschritten. Die Menge der Regelinstanzierungen fällt streng monoton, aufgrund der Optimierungsanforderung. Zum Zeitpunkt t_0 beträgt der PUE $\sim 1,6684$ d. h. 35.410 Watt primärer und 23.540 Watt sekundärer Energiebedarf. Die Optimierung kann beide Werte reduzieren. Nach 185 Optimierungsschritten (t_{185}) hat die PAM-Instanz einen primären Energiebedarf von 28.657 Watt (-19,07 % / - 6.753 Watt) und einen sekundären Energiebedarf von 14.477 Watt (-38,50 % / -9.063 Watt). Der Wert des PUEs kann auf $\sim 1,5052$ verbessert werden.

Nach einer konservativen Schätzung (220 Betriebstage á 8 Stunden) ergibt sich für den Anwendungsfall eine Reduktion des Energiebedarfs von 103.752 kWh auf 75.916

kWh. Unter den gemittelten Emissionszahlen (2008-2010) von 577 CO₂ (g/kWh) ergibt sich eine Emissionsvermeidung von 9,13 Tonnen CO₂ pro Betriebsjahr.

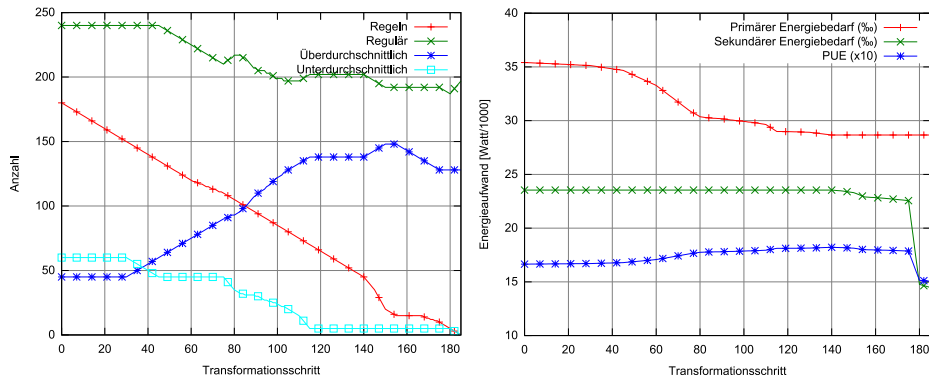


Abb. 6. Aufzeichnung der aggregierten Metriken (links) und des Energieaufwands (rechts)

7 Schlussfolgerung

Es wurde eine Methode für die modellgestützte Energieeffizienzanalyse und -optimierung von betrieblich eingesetzter IKT entwickelt. Diese Methode ist in einem prototypischen Softwarewerkzeug implementiert. Die Evaluation anhand eines Simulationsexperiments zeigt die Anwendbarkeit der Methode, die Validität der Analyse- und Optimierungsergebnisse sowie die Entscheidungsunterstützung für IKT-zentrierte Green-IT-Maßnahmen. Die Methode kann im Rahmen von Migrationsentscheidungen angewendet werden. Konkurrierende IKT werden einer einheitlichen Analyse unterzogen, die so vergleichbare Kennzahlen zur Energieeffizienz liefert. Das von der Methode verwendete EAM verfügt über vergleichsweise wenige Attribute und bildet keine Informationen über den tatsächlichen, d. h. gemessenen Energieaufwand ab. Der Zusammenhang zwischen geschätzten und gemessenen Einsparungen muss noch hergestellt werden. Weiterhin beschränkt sich die Betrachtung auf betriebliche genutzte IKT. Höherliegende IS-Schichten fehlen. Aus Sicht der operationalen Prozesse (z. B. Beschaffung) des Informationsmanagements muss die energetische Optimierung in existierende Prozesse integriert werden, da der Aufwand andernfalls die Einsparungseffekte übersteigt. Die Auswahl der Optimierungsschritte unter der Menge der instanziierten Transformationsregeln erfolgt willkürlich. Dadurch ist unklar ob optimale Lösungen durch die willkürliche Auswahl eventuell nicht mehr erreicht werden können. In diesem Fall wäre eine Regelpräzedenz oder alternativ der Nachweis von Konfluenz der Transformationsschritte notwendig.

Danksagung. Diese Arbeit ist im Rahmen des Projektes "MIGRATE! Modelle, Verfahren und Werkzeuge für die Migration in Cloud-basierte energieoptimierte Anwen-derinfrastrukturen und deren Management" entstanden. Das Projekt wird mit Mitteln

des Bundesministeriums für Wirtschaft und Technologie gefördert (Technologieprogramm IT2Green, Förderkennzeichen 01ME11052).

Literatur

1. Dao, V., Langella, I., Carbo, J.: From green to sustainability: Information Technology and an integrated sustainability framework. *J. Strategic Inf. Syst.* 20, 63–79 (2011)
2. Corbett, J.: Unearthing the value of green IT. In: ICIS 2010 Proceedings, Paper 198 (2010)
3. Jenkin, T.A., Webster, J., McShane, L.: An agenda for 'Green' information technology and systems research. *Inf. Organ.* 21, 17–40 (2011)
4. Matthes, D.: *Enterprise Architecture Frameworks Kompendium*. Springer, Berlin (2011)
5. Zachman, J.A.: A framework for information systems architecture. *IBM Syst. J.* 26, 276–292 (1987)
6. Office of Management and Budget: FEA Consolidated Reference Model Document Version 2.3 (2007)
7. US Department of Treasury: TEAF - Treasury Enterprise Architecture Framework Version 1.0 (2000)
8. The Open Group: TOGAF 9 - The Open Group Architecture Framework Version 9 (2009)
9. Scheer, A.-W.: *Architektur Integrierter Informationssysteme: Grundlagen der Unternehmensmodellierung*. Springer, Berlin (1992)
10. Frank, Ulrich: Multi-perspective Enterprise Modeling (MEMO) – Conceptual Framework and Modeling Languages. In: HICSS-35 Proceedings (2002)
11. The Open Group: ArchiMate 1.0 Specification. Van Haren Series (2009)
12. Ferstl, K.O., Sinz, E.J.: *Handbook on Architectures of Information Systems - Modeling of Business Systems Using SOM* (2006)
13. Aier, S., Kurpjuweit, S., Saat, J., Winter, R.: *Coherency Management: Architecting the Enterprise for Alignment, Agility and Assurance - Business Engineering Navigator: A 'Business to IT' Approach to Enterprise Architecture Management*. AuthorHouse (2008)
14. Bernus, P., Nemes, L.: A framework to define a generic enterprise reference architecture and methodology. *Comput. Integr. Manuf.* 9, 179–191 (1996)
15. Aier, S., Riege, C., Winter, R.: *Unternehmensarchitektur Literaturüberblick und Stand der Praxis*. *Wirtschaftsinformatik* 50, 292–304 (2008)
16. Frank, U.: The MEMO Meta-Metamodel, *Arbeitsberichte des Institut für Wirtschaftsinformatik der Universität Koblenz-Landau*, Nr. 9 (1998)
17. Guerra, E., Diaz, P., de Lara, J.: Visual Specification of Metrics for Domain Specific Visual Languages. *Electron. Notes Theor. Comp. Sci* 211, 99–110 (2006)
18. Standard Performance Evaluation Corporation: SPEC – Power and Performance : Run and Reporting Rules SPECpower_ssj2008 (2008)
19. Henning, J.L.: SPEC CPU2006 Benchmark Descriptions. *ACM SIGARCH Newsletter, Computer Architecture News* 34 (4) (2006)
20. Official Journal of the European Union. Commission Decision 2009/489/EC (2009)

Energy-aware Service Allocation for Cloud Computing

Tobias Widmer¹, Marc Premm¹, and Paul Karaenke^{1,2}

¹University of Hohenheim, Department of Information Systems 2, Stuttgart, Germany
{tobias.widmer, marc.premm, paul.karaenke}@uni-hohenheim.de

²FZI Forschungszentrum Informatik, Karlsruhe, Germany

Abstract. Energy efficiency has become an important managerial variable of IT management. Whereas cloud computing promises significantly higher levels of energy efficiency, it is still not known, if and to what extent outsourcing of software applications to cloud service providers affects the overall energy efficiency. This research is concerned with the allocation of cloud services from providers to customers and addresses the problem of energy-aware service allocation. The distributed nature of the problem, i.e., the multiple loci of control, entails the failure of centralised solutions. Hence, we approach this problem from a multiagent system perspective, which preserves the distributed setting of multiple service providers and customers. The contribution of our research is a game-theoretic framework for analysing service provider and customer interactions and a novel distributed allocation mechanism based on this framework to approximate energy-efficient, optimal allocations. We demonstrate the usefulness and efficacy of the proposed artifact in several simulation experiments.

Keywords: Game Theory, Green IT, Multiagent Systems, Negotiation, Resource Allocation

1 Introduction

Cloud-based service providers (SPs) such as Amazon and Google offer a portfolio of cloud services. Customers rent internet-accessible computing resources including CPU time and data storage services. Customers reserve such resources on demand without having to face particular operational setup costs [1] and are commonly charged according to the pricing-schemes “pay-per-use” or subscriptions [2]. Cloud computing, however, does not only promise significant cost reduction by lower IT system maintenance and operational costs, but also provides potential for increased energy efficiency, since the resources of large data centres can be managed more efficiently [3].

We consider a setting where customers intend to outsource software applications to cloud-based SPs to reduce costs and energy consumption of their IT systems. Therefore, we address the problem of *service allocation* from SPs to customers, i.e., the allocation of individual customer requests to matching available services. When allo-

cating scarce resources, a typical objective is to find allocations which are optimal regarding a metric that depends on the preferences of the individual actors. Individual preferences can be aggregated using the notion of *social welfare* from welfare economics [4]. We assume that individual actors model their preferences by utility functions and thus apply the concept of *utilitarian social welfare*, denoting the sum of all individuals' utility for a given allocation. However, the problem of finding optimal allocations is often computationally infeasible in network settings such as in cloud computing with multiple SPs and customers, since computing optimal allocations is *NP*-complete [5].

Resource allocation is a well established field in multiagent systems research [6]. Thus, each actor (i.e., SPs and customers) is represented as a software agent in the proposed approach. These agents are able to perform autonomous actions in order to pursue its individual objectives [7]. Participating agents reason about the *processes of coordination* among themselves and negotiate bilateral agreements [8]. Existing approaches, however, do not consider energy efficiency, do not address utilitarian social welfare maximisation, or require a central coordinating agent which does not exist in the setting considered.

We develop a distributed heuristic for energy-aware cloud service allocation and evaluate this artifact in a set of experiments to demonstrate its usefulness and efficacy. This research makes two specific contributions: (i) a formal framework for modelling energy-aware cloud service allocation and (ii) a novel distributed allocation mechanism that integrates energy efficiency into the allocation rationale. The formal framework is based on game theory which provides formal means to analyse the most rational choice of actions for the interacting agents. We analyse optimal allocations based on the proposed framework and show that the utilitarian social welfare maximisation problem is *NP*-complete. While applying the *second-score auction* proposed by Che [9] to our setting, we design a distributed heuristic for this problem as an auction-based allocation mechanism, where rational SP agents are incentivised to truthfully reveal private information such as marginal costs.

The allocation mechanism is evaluated by means of multiagent simulation. This evaluation provides evidence of the mechanism's efficacy. Since the proposed approach constitutes a heuristic, it does not guarantee optimal solutions but provides approximations. Another limitation of the mechanism results from the distributed nature of the problem. There is no central coordinating agent in the setting considered. Therefore, the communication complexity is relatively high as compared to a centralised heuristic.

The remainder of this paper is structured as follows. In section 2, we describe the theoretical background of our research and discuss approaches in the extant literature. In section 3, we present the formal framework. In section 4, we describe the proposed allocation mechanism. Section 5 provides the experimental evaluation. Section 6 concludes the paper.

2 Related Work

Cloud service allocation is subject of inquiry in both multiagent systems and cloud computing research. We first introduce constructs of multiagent systems. Then we employ these constructs as a theoretical lens for reviewing the extant literature.

2.1 Multiagent Systems

Agent Definition. We adopt the definition of agents by Jennings: “*An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives*” [10]. Apart from mere objects, agents have control over their internal state and their own behaviour, i.e., they possess *autonomy* over their choice of action. Thus, in line with [7], a software agent is specifically designed with the capability to act independently on behalf of its user or owner. That is, an agent is equipped with different goals of its user or owner and can discover by itself what needs to be done for the achievement of these objectives. In the following we investigate agent properties and clarify the implications for our work. These properties include (i) embeddedness in agent environments, (ii) autonomy, (iii) social ability, and (iv) deliberation and reactivity.

Agent Environment. From a technical perspective, the environment consists of anything an agent can percept through its sensors and act on through its effectors [11]. The emerging organizational context between agents defines the agents’ relationship with each other [12]. Agent organizations provide a framework of constraints and expectations about the agents’ behaviour with focus on decision making and action of specific agents [8]. The agents in our work are designed to take the *task-specific* roles of service providers and customers. The underlying *organizational structure* of this multiagent system describes the set of the agents’ long-term responsibilities and interaction patterns [13].

Agent Autonomy. Agent autonomy implies that agents have their own goals, i.e., the agents’ actions are driven by their own interests. Autonomy is always a relational concept [14]. An agent is autonomous in relation to the influence of other agents and the agent’s environment, respectively. In our work, autonomous customer agents and service provider agents are equipped with utilitarian preferences reflecting their own delegated design objectives. Thus, our agents are *self-interested*, i.e., they are acting so to maximise their own utility. Hence, our game-theoretic approach describes the agents’ autonomy as a relational concept where each individual agent can be viewed as a player in a negotiation game.

Agent Social Ability. The social ability of agents comprises the ability to communicate and cooperate with other agents. Agents own the ability to interact with other agents aiming at the satisfaction of their design objectives [7]. However, the agent’s ability to merely exchange messages with other agents is insufficient from a goal

autonomy perspective. Since other agents are autonomous themselves, they pursue their own (potentially conflicting) delegated objectives. Thus, as we cannot assume agents to share common goals, they must therefore *negotiate* and *cooperate* with other agents to achieve their goals [15]. Customer and service provider agents communicate by message passing. Messages are defined in the agents' interaction protocol and allow them to negotiate for the achievement of individual delegated objectives. In detail, our agents negotiate about price and energy consumption for the provision of cloud services.

Agent Deliberation and Reactivity. Deliberation is determining which state of the world is desirable to be achieved (goals) and how this state can be reached by performing appropriate actions. Reactivity denotes the agents' capability to sense their environments and to react upon recognised changes. In environments that involve dynamic changes (e.g., by other agents), agents must be reactive. However, this does not imply that the agent is capable of purely reactive behaviour, only. In contrast, agents reason about appropriate actions to respond to changed in the environment [7].

The environmental changes *indirectly* affect the agents' goals. Customer agents may react to changes in their environment by considering market competition and demand/supply ratios when requesting for services. Similarly, SP agents may react to service requests by strategically composing their bids based on market changes.

Multiagent Systems. Considering many real-world scenarios such as the market environment presented in this work, the design of problem solving processes requires more than a single agent, since there are *multiple loci of control*, i.e., there is a *decentralised nature of the problem* [10]. Hence, the agents' autonomy is of major importance for the applicability in scenarios with multiple, conflicting objectives. Thus it is necessary to move from the *micro level* of individual agents to the *macro level* of multiagent systems. Following Jennings [10], we use the multiagent-based software paradigm in order to (i) represent the distributed nature of the problem, (ii) enable multiple loci of control, and (iii) support competing interests of entities.

Both customer and SP agents act on behalf of independent enterprises that may be organisationally or geographically distributed. Centralized approaches employing some kind of global control are not applicable in this setting, as knowledge about data and resources is private information of the different stakeholders. Distributing control to multiple entities reduces the system's control complexity as a whole and causes individual components to be less coupled [10]. Further, both customer and SP agents are self-interested agents and pursue their individual business objectives as each agent tries to maximize its designated utility.

2.2 Energy-aware Service Allocation for Cloud Computing

Energy-efficient Cloud Computing. Most existing approaches addressing energy efficiency in cloud computing aim at the reduction of energy consumption in a single data centre. This goal is achieved by scheduling techniques and resource management for distributing thermal load or powering down server in times of low demand. Most

of these approaches also fit for conventional data centres which are run and used by the same enterprise. [16] gives an overview of potential strategies to improve energy-efficiency in cloud environments. A theoretical evaluation of scenarios in which cloud computing has an advantage above local computing resources in terms of energy-efficiency is done by [17]. The authors evaluate the balance between computing, storage and network resources on a formal basis. [18] propose a low-energy scheduling model for the optimisation of energy-efficiency and use a welfare-maximisation approach. However, these approaches are limited to the optimisation of energy efficiency of a single enterprise. A global view on the energy efficiency of cloud computing and its impact in terms of carbon dioxide emissions is examined in [19].

Agent-based Cloud Service Allocation. Multiagent resource allocation is concerned with the way *resources* are distributed among multiple intelligent agents [6]. Though the notion of a “resource” can be used in its most general way, the resources we consider in this work are computational *services*. These services are *discrete*, i.e., they cannot be divided into smaller units, and *non-shareable*, i.e., a service cannot be used by multiple agents at the same time.

Bo and Lesser [5] investigate multiagent resource allocation across computational networks, where a set of selfish agents route traffic for individual users. Before user agents can route traffic through node entities, contracts between user agents and the participating nodes need to be established. The negotiation approach proposed is of distributed nature, i.e., agents act on behalf of themselves, and the corresponding resource allocation emerges from sequences of distributed negotiations. In addition to mutual contracting, nodes are allowed to decommit from existing contracts at a penalty cost. The authors investigate the relationship between stability and optimality of the network resource allocation game. Parkes et al. [20] present guidelines for the development of distributed allocation mechanism implementations based on the Vickrey-Clarke-Groves [21] mechanism. The aim of the approach is to distribute as much computation load as possible onto network nodes and thus help to determine a suitable allocation result. However, the proposed approach requires a ‘center’ entity which communicates with network nodes through a trusted channel and selects and enforces allocation outcomes.

In [22], a market-based approach is employed to efficiently allocate computing resources by means of an automated negotiation mechanism. Agents make bilateral contracts for resource leases and are given the ability to decommit from a contract by paying a negotiated penalty to the other contract party. However, the work above does not consider energy efficient allocation of computing resources.

3 Formal Framework

This section introduces a formal framework allowing us to calculate the optimal service allocation from a social welfare perspective. By assumption, switching off hosting servers results in the reduction of power consumption. Thus, each customer agent seeks to migrate all applications (henceforth services) running on the same physical

machine to cloud-based SP agents. Since these services often depend on each other, they can only be migrated as *bundles* of services. However, customer agents will only move service bundles to cloud-based SP agents (and thus switch off physical servers) if cloud-based SP agents are able to offer a more energy-efficient way to execute the designated services. Thus, we assume each SP agent is aware of its energy efficiency. With regards to the standardized measurement of IT energy efficiency, we employ the widely accepted metric “performance-per-watt” (see [23] for a justification).

3.1 Agents and Services

To induce a migration decision, agents have to consider services executed on hosts, computational capacities, as well as energy efficiency specifications.

Agents. The set of customer agents is denoted by A_C and the set of SP agents is A_{SP} . The set of all agents is given by $A = A_C \cup A_{SP}$.

Services. The tuple of all services is denoted by $S = (s_1, s_2, \dots, s_n)$. Each service has a required computational capacity given by $w(s_k)$ with $k \in \{1, \dots, n\}$ measured in performance, i.e., computing operations per second. For example, a common performance metric, server side java operations per second (ssj_ops), is defined in SPECpower_ssj2008 [24]. We assume that different machines support the same performance metric.

Hosts. Each customer agent $i \in A_C$ owns a set of hosts (e.g., servers) denoted by $H_i = \{h_1^i, h_2^i, \dots, h_{m_i}^i\}$ with $m_i \in \mathbb{N}$. The average energy consumption rate (in watt) of each host is given by $E_i(h_l^i)$ with $l \in \{1, \dots, m_i\}$. Each host is defined by the binary tuple $h_l^i = (r_1, r_2, \dots, r_n)$, where $r_k = 1$ if service s_k is hosted on h_l^i and $r_k = 0$ if not. Due to the fact that i can only switch off its hosting hardware (and thus enhance its energy efficiency) once all services hosted by that particular hardware are migrated to SP agents, a customer agent i derives a valuation $V_i(h_l^i)$ for host h_l^i if all services hosted by h_l^i are allocated to SP agents.

Capacity. Each SP agent $j \in A_{SP}$ owns an infrastructure which provides a limited computational capacity denoted by W_j . Further, the energy efficiency of j is given by E_j (measured in performance/watt).

3.2 Agent Utility

The agents’ individual utilities are influenced by their costs, payments for service provisioning, migration valuation and energy savings of specific hosts.

Cost. The cost of SP agent j for providing computational capacity for service s_k at energy consumption rate e_k is given by $c_j(w(s_k), e_k)$, where $c_j(w(s_k), e_k) = 0$ if no service is provided.

Payment. The payment from i to j for providing service s_k is represented by p_{ijk} . Since i seeks to migrate *all* services hosted by h_l^i , the total payment for host h_l^i is defined as the sum of payments of all services hosted by h_l^i , i.e. $P_i(h_l^i) = \sum_{j \in A_{SP}} \sum_{k=1}^n p_{ijk}^l$ where p_{ijk}^l is the price for s_k hosted by h_l^i .

Allocation. An allocation is given by the tuple $X_{ijk} = (i, j, s_k, e_{ijk}, p_{ijk})$, where e_{ijk} denotes the contracted energy consumption rate of s_k . Let X denote the set of final allocations (“winning” allocations). The binary variable x_{ijk} is defined as $x_{ijk} = 1$ if $X_{ijk} \in X$, and $x_{ijk} = 0$ otherwise.

Utilities. The utilities of each customer agent i and SP agent j are given by

$$U_i = \sum_{j \in A_{SP}} \sum_{l=1}^{m_i} (V_i(h_l^i) - P_i(h_l^i)) x_{ijk}, \quad U_j = \sum_{i \in A_C} \sum_{k=1}^n (p_{ijk} - c_j(w(s_k), e_k)) x_{ijk}.$$

3.3 Optimality of Service Allocations

We consider an allocation being optimal if the sum of all agents’ individual utilities is maximised. Literature denotes this criterion as utilitarian social welfare, the standard performance metric in multiagent resource allocation. Such a metric is often used to measure the quality of an allocation with regards to the system as a whole [6]. The utilitarian social welfare is calculated as

$$\begin{aligned} W &= \sum_{i \in A_C} U_i + \sum_{j \in A_{SP}} U_j \\ &= \sum_{i, j \in A} \left(\sum_{l=1}^{m_i} (V_i(h_l^i) - P_i(h_l^i)) x_{ijk} + \sum_{k=1}^n (p_{ijk} - c_j(w(s_k), e_k)) x_{ijk} \right) \quad (1) \\ &= \sum_{i, j \in A} \left(\sum_{l=1}^{m_i} V_i(h_l^i) x_{ijk} + \sum_{k=1}^n c_j(w(s_k), e_k) x_{ijk} \right). \end{aligned}$$

Note that the payments do not appear in this equation since these simply redistribute the wealth between the agents. In the following we are interested in finding the optimal set of services that needs to be allocated to SP agents such that the social welfare is being maximized, i.e.,

$$\begin{aligned} W^* &= \max_x \sum_{i, j \in A} \left(\sum_{l=1}^{m_i} V_i(h_l^i) x_{ijk} - \sum_{k=1}^n c_j(w(s_k), e_k) x_{ijk} \right) \quad (2) \\ & \text{s.t.} \end{aligned}$$

$$\forall j \in A_{SP} : \sum_{i \in A_C} \sum_{k=1}^n w(s_k) x_{ijk} \leq W_j. \quad (3)$$

Theorem 1. The computation problem for the socially optimal allocation is *NP*-complete.

Proof. *NP* membership is easy, since for any given solution to our allocation problem the calculation of the resulting social welfare and the verification of the capacity constraints defined in (3) can obviously be done in polynomial time. For hardness, we define a reduction from the 0-1 Knapsack problem which is known to be *NP*-complete [25]. The 0-1 Knapsack problem is defined as follow. Let U be a finite set and let $u \in U$. Further, $g(u)$ denotes the weight and $b(u)$ the valuation of element u , and K is a positive integer. Further, $a(u)$ is a non-negative integer for each $u \in U$. Now maximize $\sum_{u \in U} a(u)b(u)$ s.t. $\sum_{u \in U} a(u)g(u) \leq K$ (cf. [26]). Any instance of the 0-1 Knapsack problem can be reduced to our optimization problem as follows. Let $m_i = n = 1$ in equation (2), i.e., each customer agent in A_C has exactly one service bundle ($l = 1$) containing exactly one service ($k = 1$), and let $l = 1$, i.e., one SP agent only, then we can rewrite equation (2) as

$$W^* = \max_x \sum_{i \in A_C} (V_i(h_i) - c_1(w(s_1), e_1)) x_{i11} \quad (4)$$

with constraint $\sum_{i \in A_C} w(s_1) x_{i11} \leq W_1$. Further, set $A_C = U$, $x_{i11} = a(u)$, $V_i(h_i) - c_1(w(s_1), e_1) = b_i$, $w(s_1) = g(u)$ for all $u \in U$ and $W_1 = K$. Obviously, this reduction can be done in polynomial time. Hence, each SP agent's optimization problem and therewith the utilitarian social welfare maximization problem is *NP*-complete.

3.4 Illustrative Example

We provide an example to illustrate the formal framework. This examples considers two customers c1 and c2, three service providers sp1, sp2, and sp3, and a set of services $S = \{s_1, s_2, s_3\}$ as depicted in figure 1. Each SP has an energy efficiency E_i in performance per watt, a capacity W_j in the performance metric `ssj_ops`, and a cost in € to provide its services c_j with $j = \{1, 2, 3\}$. On the demand side, both c1 and c2 own one host h_i and derive a valuation $V_i(h_i)$ with $i = \{1, 2\}$, if that particular host can be switched off. Hence, c1 requests for computing capacities $w(s_1)$ and $w(s_2)$ and c2 for $w(s_3)$. Table 1 shows all possible service allocations denoted by a1 to a16 and presents the social welfare achieved by each allocation. For the calculation, customer valuation and SP cost are used as *per computing unit*, e.g., the valuation per computing unit of c1 equals $85.5/450 = 0.19$. Obviously, allocation a1 with $85.5 + 41.4 - 26.0 - 27.5 - 27.6 = 45.8$ maximizes the social welfare for the given setting.

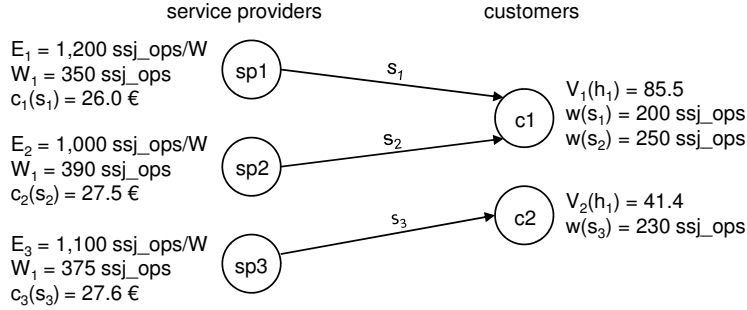


Fig. 1. Example 1 with two customers and three service providers

Table 1. Possible allocations and social welfare for example 1

Service request	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	
c1	s1	sp1	sp2	sp3	sp1	sp2	sp3	sp1	sp1	sp2	sp2	sp3	sp3	-	-	-	-
	s2	sp2	sp1	sp2	sp3	sp3	sp1	sp2	sp3	sp1	sp3	sp1	sp2	-	-	-	-
c2	s3	sp3	sp3	sp1	sp2	sp1	sp2	-	-	-	-	-	sp1	sp2	sp3	-	-
Social welfare	45.8	44.8	45.5	45.6	45.0	45.1	32.0	29.5	31.0	33.5	29.0	34.0	11.5	16.1	13.8	0.0	-

4 Allocation Mechanism

Since finding the optimal service allocation is a computationally infeasible problem, we propose a distributed heuristic for energy-aware cloud service allocation in order to approximate agents' decision-making. We employ a multi-attribute combinatorial procurement auction where multiple customer agents request for bundles of energy-efficient cloud services from strategic service provider agents. At first, we introduce a formal auction model and then describe an allocation protocol that is *incentive-compatible* for SP agents. An auction protocol is called *incentive-compatible* if it is each SP agents' dominant strategy to declare its true preference over the requested services regardless of other SP agents' actions [27].

4.1 Auction Model

The agents' participation in the auction requires service requests and bids as well as a score combining price with energy consumption.

Request. Each customer agent $i \in A_C$ requests for service bundles defined by $R_i = (R_1^i, R_2^i, \dots, R_{m_i}^i)$, where each service bundle R_l^i is represented by the binary request bundle of the form $R_l^i = (r_1, r_2, \dots, r_n)$, i.e., $r_k = 1$ if service s_k belongs to the requested bundle and $r_k = 0$ if not.

Score. Let e_k denote the energy consumption rate of service s_k and p_k the price for service provisioning. For each tuple (e_k, p_k) we adopt the notion of [9] and define a quasi-linear scoring function by

$$\sigma_i(s_k, e_k, p_k) = \alpha_i w(s_k)/e_k - \beta_i p_k, \quad (5)$$

where $\alpha_i, \beta_i \in \mathbb{R}^+$ are used for scaling purposes. The scoring function is assumed to be publicly known to all SP agents at the start of bidding. As in [9], we consider the second-score auction where the winning SP agent is obligated to match the second highest score. However, in meeting this score, the winner may choose any tuple (e_k, p_k) . As an example, let SP agent A's bid for service s_1 be (5, 0.2) and B's bid (6, 0.15), and the scoring function is such that $\sigma_i(s_1, 5, 0.2) = 10$ and $\sigma_i(s_1, 6, 0.15) = 9$. Here, SP agent A wins the auction and is free to choose any tuple (e', p') such that $\sigma_i(s_1, e', p') = 9$ is fulfilled.

Bids. Each SP agent j 's bid is given by the n -tuple $B_j = (B_1^j, B_2^j, \dots, B_n^j)$, where each B_k^j is a tuple of the form $B_k^j = (\sigma_i(s_k, e_k, p_k), e_k)$ with the score $\sigma_i(s_k, e_k, p_k)$, and the promised energy consumption rate e_k for providing service s_k . If j does not bid for service s_k , we set $B_k^j = (0, 0)$.

4.2 Auction Protocol

We consider the widely used one-shot protocol described in [28]. Further, our protocol is assumed to be *individually rational*, i.e., no agent will commit to a contract for service provision if its cost exceeds the payment received. This is natural as no agent is willing to participate in an auction where it spends more than what it earns. In the following we describe the individual steps of the protocol.

- Each customer agent i asks for bids from SP agents by announcing its request $R_i = (R_1^i, R_2^i, \dots, R_{m_i}^i)$ to all SP agents.
- For each service s_k in i 's request bundles R^i , SP agents send their binding bids B_i to customer agents.
- Customer agent i evaluates the received bids and informs SP agents about the acceptance of their bids by transferring the second highest score.
- The winning SP agents calculate the price based on the second highest score. A contract is established and second score payments are transferred. For each rejected bid, no contract is established.

Since customer agent i only derives non-zero valuations for complete service bundles, it is individually rational for i to establish contracts if and only if it has received acceptable bids for *each* service in the current bundle. Once this is the case, i selects the SP agent with the highest score for each requested service. The winning SP agent is then informed about the acceptance of its bid and the value of the auction's second highest score denoted by $\sigma_{i,2}$. Based on the scoring function defined in (5), the winning SP agent chooses an energy-price tuple (e', p') such that $\sigma_i(s_k, e_k, p_k) = \sigma_{i,2}$. A contract is established for each service s_k within the bundle specifying the price p' and the energy consumption e' for service s_k .

Note that SP agents may choose to lie when bidding, i.e., they may systematically overvalue their offered energy efficiency and base the transmitted score defined in (5) on costs that are higher than its true valuation. In the following we apply [9] and show

that this protocol is incentive-compatible for SP agents, i.e., it is each SP agent's dominant strategy to declare its true preferences for service allocations. Thus, each individually rational SP agent will always calculate its bidding score on its true marginal cost.

Theorem 2. The proposed auction protocol is incentive-compatible for SP agents, where each SP agent's bid for providing service s_k is calculated using $p_{ijk} = c_j(w(s_k), e_{ijk})$.

Proof. Applying lemma 1 from [9], we set $q = e_k, s(q) = a_i w(s_k)/e_k$, $c(q, \theta) = \beta_j c_j(w(s_k), e_k)$ and $q_s(\theta) = e_{ijk}$. Then $e_{ijk} = \arg \max(a_i w(s_k)/e_k - \beta_j c_j(w(s_k), e_k))$ which reduces the two-dimensional auction to a one-dimensional problem. Hence, proposition 1 (ii) in [9] yields $p_{ijk} = \beta_j c_j(w(s_k), e_{ijk})$. Consequently, SP agents' dominant strategy consists in calculating their bidding scores based on their marginal costs $c_j(w(s_k), e_{ijk})$.

From this result it becomes clear that SP agents will always calculate their score based on their true marginal cost, no matter on how other SP agents choose their score. With this in mind, we can deduce the following corollary.

Corollary 1. For a single customer agent request, the proposed protocol results in an optimal service allocation.

Proof. By theorem 2, it is individually rational for each SP agent to calculate its bid for providing service s_k by using its marginal cost $c_j(w(s_k), e_k)$. That is, only SP agents that offer minimal cost will be contracted by the customer agent. As customer agent's valuation for each service bundle h_l^i is given by $V_i(h_l^i)$ and costs $c_j(w(s_k), e_k)$ are minimized, the resulting service allocation is optimal.

5 Evaluation

This section provides the experimental evaluation of the proposed allocation mechanism. We describe the experimental setup, report the results, and discuss the findings.

5.1 Experimental Setup

The parameters used in our experiments are based on pricing models of Amazon in 2012 and latest power and performance information found in the SPECpower standard. According to the Amazon EC2 User Guide and SPECpower, the performance of a Dell PowerEdge R720 approximately corresponds to the computing capabilities of 88 EC2 small instances [24]. Hence, on average, one million `ssj_ops` cost 6.75 € and consume 186 Watt power. Thus, in our setup, costs are normally distributed with $\mu_c = 6.75$ and standard deviation $\sigma_c = 0.1$. Based on Amazon and SPECpower data, the energy efficiency of SPs and customer agents' hosts are assumed to follow a normal distribution with expectation $\mu_E = 5365$ `ssj_ops/W` and $\sigma_E = 268.2$. We assume that SP agents' costs correlate positively with its energy efficiency. The following random numbers were generated using normal distribution:

- $X_{c,j} \sim N(\mu_c; \sigma_c^2)$ refers to SP agent j 's fixed cost component incurred for service provisioning,
- $X_{E,j} \sim N(\mu_E; \sigma_E^2)$ refers to SP agent j 's energy efficiency,
- $c_j = X_{c,j} (X_{E,j} / \mu_E)$ refers to SP agent j 's cost for one million ssj_ops,
- $X_{c,j} \sim N(\mu_c; \sigma_c^2)$ refers to customer agent i 's fixed valuation component,
- $V_i(h_i^i) = X_{c,j} + \mu_E / X_{E,j}$ refers to customer agent i 's valuation per computing unit for host h_i^i .

The number of hosts per customer agent is Poisson distributed with an expectation $\lambda_k = 10$. The number of services per host is also Poisson distributed with an expectation $\lambda_s = s$ and each service is taken by uniform distribution from a set of 10 services with performance ranging from 0.3 to 2.1 million ssj_ops. The number of SP agents is set to 5 as we assume there exists a relatively small number of large cloud SPs on the market, e.g., Amazon and IBM. SP agents' computational capacities are generated by normal distribution such that demand and supply are balanced at 50 customer agents. Each experiment was performed 100 times. The optimal values of the allocations' social welfare are computed using CPLEX (a commercial optimization software by IBM).

5.2 Results

Both figures 2 and 3 show the average behaviour of the utility ratio (agents' social welfare relative to optimal utility) as a function of the number of customer agents where energy efficiency and price are weighted using three different ratios. Different weights can be set by means of α_i and β_i as defined in the scoring function (5) and are equal for all customer agents. In the following, energy-price weighting ratios are simply written in the form energy:price (e.g., 1:1 means energy and price are weighted equally).

Figure 2 shows a setting where higher weights are placed on energy efficiency. The utility ratio at 1:1 monotonically increases in the interval [2,44] until the maximum close to 1.0 is reached at 44 customer agents. The utility ratio then decreases again and seemingly converges to 0.9. Similarly, the utility ratio for 3:2 also increases in the interval [2,44], but starts at a lower utility ratio as compared to the function with 1:1, showing a higher slope. It then decreases and indicates a convergence to the utility ratio of 0.9. Further, the utility ratio for 4:1 is significantly low for a small number of customer agents while it increases at a high slope until the maximum close to 1.0 is reached at 44 customer agents. Hence, it finally decreases and reaches a utility ratio of 0.9.

Figure 3 shows a setting where higher weights are placed on price rather than on energy efficiency. The utility ratio with 1:1 is the same as the one displayed in figure 2. At 2:3 the utility ratio dominates the one for 1:1 and increases monotonically in the interval [2,44] until the maximum close to 1.0 is reached at customer agent number 44. Similarly, the weights 1:4 cause the utility ratio to exceed all other ratios while constantly remaining close to 1.0 in the interval [2,44]. It then decreases for a growing number of customer agents and reaches a utility ratio of 0.9.

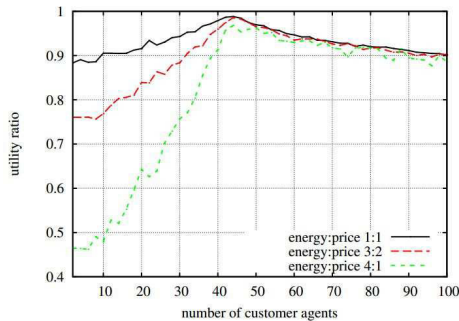


Fig. 2. Utility ratio as function of customer number with higher weights on energy.

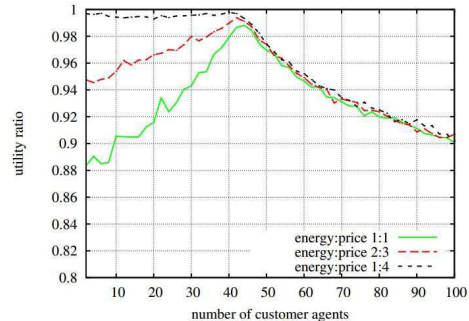


Fig. 3. Utility ratio as function of customer number with higher weights on price.

5.3 Discussion

Our experiments demonstrate the influence of energy-aware customer agents on the agents' social welfare and provide evidence for the usefulness and efficacy of the proposed auction mechanism. In the interval [2,44] all "energy-aware" utility ratios displayed in figure 2 increase monotonically. This result demonstrates the impact of the changing demand/supply-ratio with an increasing number of customer agents: Within the interval [2,48], on average, service supply is greater than the requested service demand. Hence, on a market with excess supply, energy-aware customer agents easily satisfy their demand and choose energy-efficient services. The incurring costs of SP agents, however, are high when providing services of high energy efficiency. Therefore, customer agents that overvalue energy efficiency as compared to price will purchase services from expensive SP agents. Hence, capacities of cheap SP agents remain unused.

Since the social welfare decreases at higher SP agent costs and unchanging customer agent valuation, both utility ratios for 4:1 and 3:2 are low for a small number of customer agents but increase until the market is saturated (on average, this happens at 50 customer agents). That is, increasing weights for energy efficiency result in higher costs for service provisioning and hence in lower utility ratios. When the market is balanced, the utility ratios reach its maximum for all energy-price weights. The fact that the utility ratios then decrease within the interval [46,100] is intuitive: Once services for sale are beginning to be scarce, resource competition is high and the mechanism can only achieve a high social welfare if services of high valuation can be migrated to SP agents. In such situations, customer agents are willing to forgo their energy-price preferences in order to obtain their valuations. Further, with a growing number of customer agents, the success rate of the mechanism decreases as it becomes more difficult for competing customer agents to migrate their services. Hence, the utility ratios for all energy-price preferences reach 0.9.

Experiments for the setting with "price-aware" customer agents as displayed in figure 3 confirm a rather trivial intuition: When placing higher weights on the price

and lower weights on energy efficiency (e.g., 1:4), the mechanism achieves a constant maximal utility ratio close to 1.0. Obviously, in a market with low competition, customer agents acquire services from SP agents with lowest cost while paying only little attention to energy efficiency. Once resources are being scarce (i.e., beyond 50 customer agents), the high competition among customer agents causes the utility ratio to finally reach 0.9 for all considered energy-price preferences.

6 Conclusion

This work presents a formal framework for modelling energy-aware cloud service allocation and proposes a distributed allocation mechanism that integrates energy efficiency into the allocation rationale. Customers and cloud SPs are represented as software agents that autonomously negotiate service agreements based on a set of different energy-price preference ratios. We employ game theory to analyse optimal service allocation and show the *NP*-completeness of the underlying utilitarian social welfare maximisation problem. We develop a distributed heuristic for the allocation problem. The proposed allocation mechanism, a second-score auction protocol, is shown to be incentive-compatible for SP agents. We evaluate this mechanism by means of multiagent simulation. The current formal framework is limited to two non-functional properties of services, i.e., energy efficiency and performance. In our future work, we plan to include additional properties such as response time and availability.

Acknowledgements. This work has been supported by (1) the project MIGRATE!, funded by the German Federal Ministry of Economics and Technology (BMW, FKZ 01ME11052), and (2) the eHealthMonitor project (<http://www.ehealthmonitor.eu>), funded by the European Commission under contract FP7-287509.

References

1. Weiss, A.: Computing in the clouds. *netWorker* 11 (4), 16–25 (2007)
2. Weinhardt, C., Anandasivam, A., Blau, B., Borissov, N., Meinel, T., Michalk, W., Ster, J.: Cloud-computing. *Wirtschaftsinformatik* 51, 453–462 (2009)
3. The green grid consortium, <http://www.thegreengrid.org> (2011)
4. Sen, A.K.: *Collective Choice and Social Welfare*. Holden-Day, Michigan (1970)
5. Bo, A., Lesser, V.: Characterizing contract-based multi-agent resource allocation in networks. *IEEE Sys. Man. Cybern.* 40, 575–586 (2010)
6. Chevaleyre, Y., Dunne, P.E., Endriss, U., Lang, J., Lemaitre, M., Maudet, N., Padget, J., Phelps, S., Rodriguez-Aguilar, J.A., Sousa, P.: Issues in multiagent resource allocation. *Informatica* 30, 3–31 (2006)
7. Wooldridge, M.: *An Introduction to MultiAgent Systems*. 2nd edn. John Wiley & Sons, Chichester (2009)
8. Bond, A.H., Gasser, L. (eds.): *Readings in Distributed Artificial Intelligence*. Morgan Kaufmann Publishers, San Mateo (1988)

9. Che, Y.K.: Design competition through multidimensional auctions. *Rand J. Econ.* 24, 668–680 (1993)
10. Jennings, N.: On agent-based software engineering. *Artif. Intell.* 117, 277–296 (2000)
11. Russell, S.J., Norvig, P.: *Artificial Intelligence: A Modern Approach*. 2nd edn. Prentice Hall, New Jersey (2003)
12. Ferber, J., Gutknecht, O.: A meta-model for the analysis and design of organizations in multiagent systems. In: *Proceedings on the Third International Conference on Multi-Agent Systems (ICMAS-98)* (1998)
13. Durfee, E.H., Lesser, V.R., Corkill, D.D.: Coherent cooperation among communicating problem solvers. *IEEE Transactions on Computers* C-36, 1275–1291 (1987)
14. Castelfranchi, C., Falcone, R.: *Founding Autonomy: The Dialectics Between (Social) Environment and Agent's Architecture and Powers*. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.): *Agents and Computational Autonomy*. LNCS, Vol. 2969, pp. 40–54. Springer, Heidelberg (2004)
15. Castelfranchi, C.: Modelling social action for AI agents. *Artif. Intell.* 103, 157–182 (1998)
16. Berl, A., Gelenbe, E., Di Girolamo, M., Giuliani, G., De Meer, H., Dang, M.Q., Pentikousis, K.: Energy-efficient cloud computing. *Comput. J.* 53, 1045–1051 (2009)
17. Baliga, B.J., Ayre, R.W.A., Hinton, K., Tucker, R.S.: Green cloud computing: Balancing energy in processing, storage and transport. In: *Proceedings of the IEEE 99*, pp. 149–167 (2010)
18. Bodenstein, C., Hedwig, M., Neumann, D.: Low-energy automated scheduling of computing resources. In: *1st IEEE/ACM Workshop on Autonomic Computing for Economics* (2011)
19. Garg, S.K., Yeo, C.S., Buyya, R.: Green cloud framework for improving carbon efficiency of clouds. In: Jeannot, E., Namyst, R., Roman, J. (eds.): *Euro-Par 2011*. LNCS, Vol. 6852, pp. 491–502. Springer, Heidelberg (2011)
20. Parkes, D., Shneidman, J.: Distributed implementations of Vickrey-Clarke-Groves mechanisms. In: *Proceedings of the 3rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS'04)*, pp. 261–268. ACM, New York (2004)
21. Vickrey, W.: Counterspeculation, auctions, and competitive sealed tenders. *J. Financ.* 16, 8–37 (1961)
22. Bo, A., Lesser, V., Irwin, D., Zink, M.: Automated negotiation with decommitment for dynamic resource allocation in cloud computing. In: *Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'10)*, pp. 981–988 (2010)
23. Sharma, S., Hsu, C., Feng, W.: Making a case for a green500 list. In: *Proceedings of the Workshop on High-Performance, Power-Aware Computing* (2006)
24. SPEC: Standard performance evaluation corporation, <http://www.spec.org> (2012)
25. Karp, R.: Reducibility among combinatorial problems. In: Miller, R., Thatcher, J. (eds.): *Complexity of Computer Computations*. Plenum Press, New York (1972)
26. Garey, M.R., Johnson, D.S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York (1979)
27. Mas-Colell, A., Whinston, M.D., Green, J.R.: *Microeconomic Theory*. Oxford University Press (1995)
28. Rubinstein, A.: Perfect equilibrium in a bargaining model. *Econometrica* 50, 97–109 (2006)

Modeling and Development Methods

Stefan Strecker¹ und Klaus Turowski²

¹ FernUniversität Hagen, Hagen, Germany

² Otto-von-Guericke-Universität, Magdeburg, Germany

Zielgerichtete Abstraktion ist eine wesentliche Voraussetzung für die differenzierte Analyse, Planung und Gestaltung komplexer Systeme. Rechnergestützte Informationssysteme sowie die organisatorischen Handlungssysteme, in denen sie eingesetzt werden, weisen nicht nur eine erhebliche Komplexität auf, an ihrer Entwicklung und Nutzung sind zudem Akteure mit divergierenden professionellen Prägungen und Interessen beteiligt. Die konzeptuelle Modellierung und konzeptuelle Modelle zählen vor diesem Hintergrund zu den Kernforschungsgebieten der Wirtschaftsinformatik. Sie sind darauf gerichtet, die Analyse, den Entwurf und die Gestaltung rechnergestützter Informationssysteme im Kontext der sie umgebenden Handlungssysteme durch zweckadäquate Abstraktionen zu unterstützen und dabei die Perspektiven übergreifender Zusammenarbeit der beteiligten Akteure zu befördern. Die Modellierungsforschung in der Wirtschaftsinformatik zielt darauf, Sprachen und Methoden zur Modellierung betrieblicher Informationssysteme zu entwickeln und kritisch zu evaluieren. Weitere Themen betreffen u. a. die Gestaltung von Referenzmodellen für ausgewählte Domänen, den Entwurf domänenspezifischer Modellierungssprachen und die Untersuchung des Nutzungskontextes von konzeptuellen Modellen. Konzeptuelle Modelle bilden darüber hinaus die Grundlage von Entwicklungsmethoden, die auf den Entwurf, die Implementierung und Einführung betrieblicher Informationssysteme zielen. Die mit diesem Tagungsband vorliegenden Beiträge berichten über aktuelle Forschungsarbeiten in diesem thematischen Kontext und zeigen neue Erkenntnisangebote auf.

Von 38 eingereichten Manuskripten wurden nach einem doppelt-blinden Begutachtungsverfahren elf Beiträge zur Veröffentlichung angenommen, was einer Annahmequote von rund 29 Prozent entspricht. Die überaus positive Resonanz auf den Aufruf zur Einreichung von Beiträgen in Verbindung mit einer erfreulichen Vielfalt an untersuchten Forschungsfragen und methodischen Herangehensweisen unterstreicht die anhaltende Dynamik und Bedeutung des Forschungsgebiets. Im Rahmen des Begutachtungsverfahrens wurde jede Einreichung von mindestens drei Gutachtern begutachtet, einem Mitglied des Programmkomitees zusammenfassend beurteilt und unter Berücksichtigung der Überarbeitungen seitens der Autoren und des „Rebuttals“ der Autoren im Kreise der Mitglieder des Programmkomitees diskutiert.

Für ihre sehr kompetente Unterstützung bedanken wir uns bei den Mitgliedern des Programmkomitees. Unser Dank gilt ebenso allen Gutachterinnen und Gutachtern, die mit Umsicht und Sorgfalt die Begutachtung der Einreichungen vorgenommen haben. Allen Autorinnen und Autoren danken wir für ihre Einreichungen.

Programmkomitee

Jörg Becker, Universität Münster
Jörg Desel, FernUniversität Hagen
Stefan Eicker, Universität Duisburg-Essen
Gregor Engels, Universität Paderborn
Werner Esswein, Technische Universität Dresden
Ulrich Frank, Universität Duisburg-Essen
Reinhard Jung, Universität St. Gallen
Dimitris Karagiannis, Universität Wien
Peter Loos, Universität des Saarlands und DFKI
Heinrich C. Mayr, Universität Klagenfurt
Jan Mendling, Wirtschaftsuniversität Wien
Andreas Oberweis, Karlsruher Institut für Technologie
Elmar Sinz, Universität Bamberg.

A Formal Specification of the Horus Modeling Language Using FDMM

Hans-Georg Fill¹, Susan Hickl², Dimitris Karagiannis¹, Andreas Oberweis²,
and Andreas Schoknecht²

¹ University of Vienna, Research Group Knowledge Engineering, Vienna, Austria
{hans-georg.fill, dimitris.karagiannis}@dke.univie.ac.at

² Karlsruhe Institute of Technology, Institute of Applied Informatics and
Formal Description Methods, Karlsruhe, Germany
{susan.hickl, andreas.oberweis, andreas.schoknecht}@kit.edu

Abstract. In this paper we show how a modeling language from the area of business process engineering can be formally specified using meta modeling concepts. This serves as a basis for the implementation on an industry-scale meta modeling platform. For this purpose we revert to the Horus modeling method and the FDMM formalism that has recently been introduced to formally describe meta models and models. Subsequently we report on the implementation of the modeling language on the ADOxx meta modeling platform and discuss the lessons learned by the application of this approach.

Keywords: Modeling, Meta Modeling, Business Process Engineering

1 Introduction

For many years the field of business informatics has developed a multitude of modeling methods and tools to support the representation and analysis of complex business-IT relationships [1]. When implementing such modeling methods in the form of modeling tools, meta modeling concepts and platforms today greatly facilitate the implementation of the contained modeling languages [2-5]. Based on formal specifications of the model, object and data types together with their respective attributes, a modeling language can thus be realized very efficiently with no or little programming effort. In the paper at hand we describe how the modeling language of the Horus modeling method is implemented using such meta modeling concepts [3]. For this purpose we created a formal specification of the Horus modeling language for exactly describing meta models and models [6]. This formal specification is then used for the implementation on an industry-scale meta modeling platform. Subsequently, we discuss our experience of using this approach for the realization of modeling methods.

The remainder of the paper is structured as follows: In section 2 the foundations for our approach will be outlined. These will encompass a characterization of the Horus modeling method for business process engineering, the FDMM formalism and the ADOxx meta modeling platform. In section 3 it will be shown how the FDMM for-

malism has been applied to Horus to establish the basis for the implementation on the ADOxx meta modeling platform. The lessons learned from this application will then be discussed in section 4. In section 5 we will discuss related work concerning similar modeling methods and related meta modeling approaches. The paper is concluded with an outlook on the future work in section 6.

2 Foundations

In this section we will briefly describe the foundations used for our approach. This includes at first the foundations of the Horus modeling method, then the FDMM formalism and finally the ADOxx meta modeling platform.

2.1 The Horus Modeling Method

The Horus modeling method for business process engineering comprises steps for an integrated modeling of business processes and for the improvement and further use of the created models [8]. The application of this method always considers a business process in terms of its organizational environment. This is realized by using a set of interrelated models describing different aspects of the business process which are part of the Horus modeling method.

The Horus modeling method classifies the business process engineering into four phases. In phase 0 the engineering project has to be prepared. During the phases 1 and 2 the integrated Horus process model is created. In phase 1 the strategic aspects and the description of the enterprise and system architecture are represented by the model types: *goal model*, *context model*, *supply- & services model*, *SWOT model*, *strategy model*, *risk model*, *key figure model*, *object model*, *rule model*, *business unit model*, *business process architecture model* and *system architecture model*. Based on this set of models, the business process analysis has to be done in phase 2. Besides improved models of phase 1, the results of this phase are a *resource model*, an *organization model* and a *procedure model*. The models created in phase 2 describe the business process in more detail and from a technical point of view. In phase 3 the use of the integrated process model for implementation and further activities follows.

A special feature of the Horus modeling method is the use of XML nets for process descriptions. XML nets [9] are high-level Petri nets where tokens represent identifiable objects. The places are typed by an XML schema whereby places can be interpreted as containers for XML documents describing relevant process objects. The edges are labeled with filter schemas that can be expressed by an XQuery expression, which describes the relevant process objects for the following transitions. The flow of an XML document is defined by the occurrences of transitions. A transition is activated if every place in the pre-set of the transition contains a valid XML document, which observes both conditions specified in the filter-schema of the adjacent edge and in the logical expression of the transition. Then the transition may fire and the respective XML documents will be assigned to the post-set of the transition. Using XML nets

enables the integrated modeling of structured business objects and object flows as well as process simulation.

2.2 The FDMM Formalism

According to a framework introduced in [3], modeling methods are comprised of a modeling technique and mechanisms and algorithms. The modeling technique can then be further detailed by a modeling language and a modeling procedure that defines the way how to apply the modeling language together with the mechanisms and algorithms to achieve results. In the following we will focus on the aspects of the modeling language and in particular the description of its syntax by using meta models and models.

In order to exactly describe meta models and models independently of their actual implementation, the FDMM formalism has been developed [6]. It can be used to describe meta models and models in a mathematical way. In this way, the resulting formal descriptions can serve as input for the implementation of the meta models and models in an IT environment. In FDMM a meta model MM contains the following parts:

$$\text{MM} = \langle \text{MT}, \preceq, \text{domain}, \text{range}, \text{card} \rangle \quad (1)$$

Thereby the set MT comprises the set of model types specified for this meta model:

$$\text{MT} = \{\text{MT}_1, \text{MT}_2, \dots, \text{MT}_m\} \quad (2)$$

Each model type MT_i is a tuple of a set of object types O_i^T , a set of data types D_i^T and a set of attributes A_i :

$$\text{MT}_i = \langle O_i^T, D_i^T, A_i \rangle \quad (3)$$

All object types, data types, and attributes of the model types are parts of the sets O^T , D^T and A whereby object types may also exist independently of model types:

$$O^T = \cup_j O_j^T, D^T = \cup D_i^T, A = \cup A_i \quad (4)$$

\preceq is an ordering on the set of object types, O^T , i.e. if $o_1^t \preceq o_2^t$ then we denote object type o_1^t as a ‘subtype’ of object type o_2^t .

For assigning attributes to object types the domain function maps attributes to the power set of object types:

$$\text{domain} : A \rightarrow \mathcal{P}(\cup_j O_j^T) \quad (5)$$

Similarly, the range function maps attributes to the power set of all pairs of object types and model types, to data types, and to model types. It thus constrains what values an attribute can take in the model instances. Apart from the assignment of data types, e.g. for strings and integers, this mapping also permits to link to object types of the same or other model types and to model types:

$$\text{range} : A \rightarrow \mathcal{P}(\cup_j (O_j^T \times \{\text{MT}_j\}) \cup D^T \cup \text{MT}) \quad (6)$$

The card function constrains how many attribute values an object may have:

$$\text{card}: O^T \times A \rightarrow \mathcal{P}(\mathbb{N}_0 \times (\mathbb{N}_0 \cup \{\infty\})) \quad (7)$$

The sets O^T , D^T , A are pairwise disjoint. For any attribute it is defined that a corresponding domain function must point to an object type of the same model type:

$$a \in A_i \Rightarrow \text{domain}(a) \subseteq O_i^T \quad (8)$$

The instantiation of a meta model MM is a tuple:

$$\langle \mu_{mt}, \mu_O, \mu_D, \tau, \beta \rangle \quad (9)$$

where μ_{mt} is a mapping from model types MT to the power set of model instances mt

$$\mu_{mt}: \text{MT} \rightarrow \mathcal{P}(\text{mt}) \quad (10)$$

with the set mt being the union of all mappings of model types to model instances so that every element of the set of model instances mt has to be derived from a model type:

$$\text{mt} = \bigcup \mu_{mt}(\text{MT}_j) \quad (11)$$

The function μ_O maps object types of a particular model type to the power set of object instances O :

$$\mu_O: \bigcup_j (O_j^T \times \{\text{MT}_j\}) \rightarrow \mathcal{P}(O) \quad (12)$$

where O is the union of all object instances so that there is no object instance without a mapping to an object type and a model type :

$$O = \bigcup_j \mu_O(O_j^T \times \{\text{MT}_j\}) \quad (13)$$

An object type $o_t \in O^T$ can be defined as an abstract type. This means that for all model types MT_i which contain the object type o^t ($o^t \in O_i^T$) the object type can only be instantiated through one of its subtypes:

$$\mu_O(o^t, \text{MT}_i) = \bigcup_{o_1^t \neq o^t, o_1^t \leq o^t} \mu_O(o_1^t, \text{MT}_i) \quad (14)$$

The function μ_D maps the data types to the power set of data objects. The data objects themselves are not further defined or constrained. The FDMM formalism thus leaves it to the user to further specify the nature and valid content of a data type:

$$\mu_D: D^T \rightarrow \mathcal{P}(D) \quad (15)$$

For describing the model instances FDMM uses triple statements τ defined as:

$$\tau \subseteq O \times A \times (D \cup O \cup \text{mt}) \quad (16)$$

The function β is then used to map the set of model instances mt to the power set of these triple statements and thus assigns triple statements to the model instances:

$$\beta: \text{mt} \rightarrow \mathcal{P}(\tau) \quad (17)$$

In this context, a correctness constraint is defined by FDMM so that τ is the disjoint union of $\beta(\text{mt}_i)$ with $\text{mt}_i \in \text{mt}$, meaning that every triple is contained in exactly one model instance.

Additionally, FDMM defines a number of disjointness and partitioning constraints for the instantiation of meta models. We will briefly describe these in the following in textual form, for the formal definitions we refer to [6]: in FDMM the instances of object types and data types must be disjoint; instantiations from multiple object types, from multiple model types or from multiple data types are not permitted.

For the correct application of the inheritance, domain, range, and cardinality constructs the following constraints must be satisfied [6]: the instantiation of an object type that is a subtype of another object type via the function μ_O is a subset of the instantiation of that other object type; instantiations of sibling object types are disjoint; the value part in the triple statements τ has to match the corresponding range definition for the used attribute and the object type – e.g. if the range for an attribute is to be defined as a ‘string’, then only a ‘string’ value can be used in a triple statement for that attribute; for triple statements consisting of an object instance, an attribute, and a value it must hold that the attribute is part of a domain function for the object type used for the instance; for data objects it must hold that their corresponding data type must be defined for the same model type as the attribute; and finally, the cardinality definitions constrain the minimum and maximum number of triple statements for a pair of object types and attributes.

2.3 The ADOxx Meta Modeling Platform

The ADOxx meta modeling platform is currently being used for several projects within the Open Models Initiative¹ and has been successfully deployed in industry for a wide range of business-IT scenarios [7]. In contrast to other meta modeling approaches, the ADOxx platform does not require programming skills for the implementation of meta models and model editors and thus supports a fast and efficient definition of modeling languages.

The ADOxx meta modeling approach builds upon three layers: On the top layer stands the ADOxx meta-meta model that defines the constructs that can be used for the definition of meta models. The basic constructs are classes and relationclasses that can be grouped by using model types. Both classes and relationclasses may contain attributes. On the second layer, meta models can be specified to define the abstract syntax of a modeling language. This abstract syntax is then instantiated on the third and bottom layer to multiple models. For the representation of the concrete syntax, a dynamic graphical notation is assigned to classes and relationclasses. This representation can be automatically adapted during run-time based on the current state of attributes.

At its core the ADOxx meta modeling platform is a client-server based software application that implements the ADOxx meta modeling approach. It provides a graphical user interface for the definition of classes, relationclasses, and attributes and stores both meta models and models in a relational database. From the specifications in the meta models, the platform automatically generates graphical model editors,

¹ Karagiannis, D., Grossmann, W., Hoefferer, P.: Open Model Initiative - A Feasibility Study: http://cms.dke.univie.ac.at/uploads/media/Open_Models_Feasibility_Study_SEPT_2008.pdf

provides generic XML import and export formats and handles the persistency of the models and meta models. Based on its client-server architecture it also supports multi-user environments with fine-grained access controls. Furthermore, it offers additional components such as the AdoScript language for implementing algorithms on top of meta models and models, a simulation component for applying process-based simulation algorithms, an analysis component for querying models using the AQL query language and a documentation component for the automatic generation of model documentations in various export formats.

3 Formal Specification of Meta Models for Horus

Based on the foundations presented in the previous section, we can now apply the FDMM formalism to the Horus modeling language. This will provide us with a specification for the implementation on the ADOxx platform. The formal specification presented in the following does, however, not comprise all aspects of the language but only a selected set of constructs. This will illustrate some of the core aspects of applying meta modeling concepts to the Horus modeling language with FDMM. Subsequently we will discuss the implementation on the ADOxx platform.

3.1 FDMM Specification of the Meta Model

As outlined in section 2.1 Horus is composed of a large number of model types that are all tightly interconnected. For illustrating the application of the FDMM formalism we selected four core model types that are used in Horus to describe XML nets. The model types we will discuss in the following are: the procedure model MT_{PM} , the employee pool model MT_{EM} , the role pool model MT_{RM} , and the object model MT_{OM} :

$$MT_{PM} = \langle O_{PM}^T, D_{PM}^T, A_{PM} \rangle \quad (18)$$

$$MT_{EM} = \langle O_{EM}^T, D_{EM}^T, A_{EM} \rangle \quad (19)$$

$$MT_{RM} = \langle O_{RM}^T, D_{RM}^T, A_{RM} \rangle \quad (20)$$

$$MT_{OM} = \langle O_{OM}^T, D_{OM}^T, A_{OM} \rangle \quad (21)$$

To illustrate how a model type is detailed by its object types, data types and attributes, we show this for the procedure model type. The object types of the procedure model are defined as:

$$O_{PM}^T = \{ \text{Abstract-Procedure-Class, Object-store, Activity,} \\ \text{Connection, Human-resource-requirements} \} \quad (22)$$

We can then define inheritance relationships between the object types by:

$$\begin{aligned} \text{Object-Store} &\preceq \text{Abstract-Procedure-Class} \\ \text{Activity} &\preceq \text{Abstract-Procedure-Class} \end{aligned} \quad (23)$$

Thereby, the object types ‘Object-Store’ and ‘Activity’ are defined as subtypes of the object type ‘Abstract-Procedure-Class’ that is defined as abstract. This makes it easier

to assign attributes to the sub types and also simplifies the specification of relationships between all subtypes of an object type.

Next, we define the data types for the procedure model type by:

$$D_{PM}^T = \{String, Integer, Float, File, \mathbf{Enum}_{WF}, \mathbf{Enum}_{AT}\} \quad (24)$$

As FDMM does not further define the data types that can be used, we are free to use either common types such as *String*, *Integer* or *Float* or custom ones such as *File* that have to be specified during the implementation based on the used implementation platform. By using another set as a data type as it is shown by the \mathbf{Enum}_{WF} and \mathbf{Enum}_{AT} types for example, we can also express data types with pre-defined, fixed values:

$$\mathbf{Enum}_{WF} = \{\text{yes, no}\} \quad (25)$$

$$\mathbf{Enum}_{AT} = \{\text{executing, checking, responsible, informing}\} \quad (26)$$

Subsequently, we can define the attributes necessary for the object types. In the set A_{PM} we only present an excerpt of the attributes that were actually defined for this model type for reasons of brevity. For example, the ‘Activity’ object type also requires a number of additional attributes for specifying time properties and simulation parameters. The attribute set also comprises elements that will later be used to specify the start- and endpoint of relation-classes, e.g. the connection-from and the connection-to attributes:

$$A_{PM} = \{\text{Name, Object-type, Object-number, Documents, connection-to, connection-from, sub-diagram, HR-req, Role-ref, Assignment-type, Quantity, Percentage, XQuery-transition-condition}\} \quad (27)$$

In the current formalization of the Horus modeling language, the employee pool model and the role pool model only contain a small number of object types, data types and attributes, i.e. for the employee pool model:

$$O_{EM}^T = \{\text{Employee}\} \quad (28)$$

$$D_{EM}^T = \{String, File, Calendar\} \quad (29)$$

$$A_{EM} = \{\text{Name, Availability, Documents}\} \quad (30)$$

And similarly the definitions for the role pool model:

$$O_{RM}^T = \{\text{Role, Employees}\} \quad (31)$$

$$D_{RM}^T = \{String, Float, File\} \quad (32)$$

$$A_{RM} = \{\text{Name, Quality, Assigned-employees, Documents, Employee, Intensity}\} \quad (33)$$

In contrast to these simple model types, the definition of the object model type requires more constructs as it constitutes a way of representing actual XML schemas. The object types of this model type are therefore defined as follows:

$$O_{OM}^T = \{\text{Object, Object-copy, Object-Aggregation, Collective-Constraint, Relationship, Inheritance, Constraint-connection, Keys, Attributes, Constraints}\} \quad (34)$$

For the data types of this model type we again use the possibility of referring to pre-defined sets of attribute values:

$$D_{OM}^T = \{String, Float, File, \mathbf{Enum}_{DT}, \mathbf{Enum}_{opt}, \mathbf{Enum}_{CT}, \mathbf{Enum}_{CARD}\} \quad (35)$$

$$\mathbf{Enum}_{DT} = \{\text{Unspecified, String, Integer, Float, Date, Enumeration}\} \quad (36)$$

$$\mathbf{Enum}_{opt} = \{\text{Yes, No}\} \quad (37)$$

$$\mathbf{Enum}_{CT} = \{\text{XOR, OR, SIM}\} \quad (38)$$

$$\mathbf{Enum}_{CARD} = \{\langle 0..n \rangle, \langle 1..1 \rangle, \langle 1..n \rangle\} \quad (39)$$

The set of attributes for the object model type are then defined as follows:

$$A_{OM} = \{\text{Name, Is-Root, XML-Schema, Key-attributes, Attributes, Constraints, Constraint-type, Relationship-from, Relationship-to, Inheritance-from, Inheritance-to, Constraint-connection-from, Constraint-connection-to, is-inside}\} \quad (40)$$

Finally, we can conclude the formal specification by adding domain, range, and cardinality definitions for the attributes. Again we selected some of the attributes and object types defined above to illustrate this. By assigning an attribute to a super-type, all subtypes automatically inherit the attribute definition as e.g. shown for the Abstract-procedure-class object type and the name attribute:

$$\begin{aligned} \text{domain}(\text{Name}) &= \{\text{Abstract-procedure-class}\} \\ \text{range}(\text{Name}) &= \{String\} \\ \text{card}(\text{Name}) &= \langle 1,1 \rangle \end{aligned} \quad (41)$$

For the specification of references between object instances and object instances and model instances it can be chosen from two directions in FDMM: the first is to directly reference another object type or model type and the second is to use an intermediary object type that has references to the object and/or model types that shall be connected. The references are in all cases expressed by attributes whose range contains other object types or model types. At first we illustrate a direct reference to other object types by the example of assigning an object type to an attribute of 'Object-store':

$$\begin{aligned} \text{domain}(\text{Object-type}) &= \{\text{Object-store}\} \\ \text{range}(\text{Object-type}) &= \{\text{Object, Object-Aggregation}\} \\ \text{card}(\text{Object-type}) &= \langle 0,1 \rangle \end{aligned} \quad (42)$$

In this way a core feature of Horus to represent XML nets is specified. The attribute 'Object-type' that is attached to the 'Object-store' object type can thus be used to reference object instances of the types 'Object' and 'Object-Aggregation' that are part of the object model type. As this reference points directly to another object type no further information can be assigned to the reference itself. We show in the following how to resolve this.

To enable more complex references where the reference itself can be further specified, an intermediary object type has to be used. We illustrate this in the following for the specification of edges in the procedure model type, which are denoted as 'Connec-

tions' in the Horus modeling language. For this purpose the two attributes 'connection-to' and 'connection-from' are assigned to the 'Connection' object type with the range definition pointing to the 'Abstract-procedure-class'. The cardinalities of the from and to attributes are set to <1,1> as edges can only occur with exactly two object instances attached to them:

$$\begin{aligned} \text{domain}(\text{connection-to}) &= \{\text{Connection}\} \\ \text{range}(\text{connection-to}) &= \{\text{Abstract-procedure-class}\} \\ \text{card}(\text{connection-to}) &= \langle 1,1 \rangle \end{aligned} \quad (43)$$

$$\begin{aligned} \text{domain}(\text{connection-from}) &= \{\text{Connection}\} \\ \text{range}(\text{connection-from}) &= \{\text{Abstract-procedure-class}\} \\ \text{card}(\text{connection-from}) &= \langle 1,1 \rangle \end{aligned} \quad (44)$$

When instantiating the object type 'Connection' it becomes possible to connect these instances to instances of the type 'Abstract-procedure-class' and to treat this relation separately from the objects it connects to. At the same time, further attributes may be assigned to 'Connection'. This is for example necessary to define transition conditions in XML nets that can be specified via XQuery strings:

$$\begin{aligned} \text{domain}(\text{XQuery-transition-condition}) &= \{\text{Connection}\} \\ \text{range}(\text{XQuery-transition-condition}) &= \{\text{String}\} \\ \text{card}(\text{XQuery-transition-condition}) &= \langle 1,1 \rangle \end{aligned} \quad (45)$$

When a separate treatment of the relationship between object types is not required but the relation should still be detailed by additional attributes we can express this in FDMM in the following way. The 'Activity' objects in Horus have to be detailed by their requirements in terms of human resources. Therefore they are linked to 'Role' objects in the role pool model. However, it should be possible to detail for each role, if the role is just executing or checking the activity or is responsible for it or has to inform someone. At first we specify the attribute 'HR-req' that can point to any number of 'Human-resource-requirements' objects.

$$\begin{aligned} \text{domain}(\text{HR-req}) &= \{\text{Activity}\} \\ \text{range}(\text{HR-req}) &= \{\text{Human-resource-requirements}\} \\ \text{card}(\text{HR-req}) &= \langle 0, \infty \rangle \end{aligned} \quad (46)$$

Then we specify the attribute 'Role-ref' for the object type 'Human-resource-requirements':

$$\begin{aligned} \text{domain}(\text{Role-ref}) &= \{\text{Human-resource-requirements}\} \\ \text{range}(\text{Role-ref}) &= \{\text{Role}\} \\ \text{card}(\text{Role-ref}) &= \langle 1,1 \rangle \end{aligned} \quad (47)$$

As the target reference is now also an object type, we can add further attributes to detail it - for example by reverting to the previously defined **Enum_{AT}** data type:

$$\begin{aligned}
\text{domain}(\text{Assignment-type}) &= \{\text{Human-resource-requirements}\} \\
\text{range}(\text{Assignment-type}) &= \{\mathbf{Enum}_{AT}\} \\
\text{card}(\text{Assignment-type}) &= \langle 1,1 \rangle
\end{aligned} \tag{48}$$

Another important feature in Horus procedure models is the use of refinements for individual activities [8]. Thereby it is controlled how many details should be displayed on each level. In FDMM these refinements can be expressed by references to other model instances of the same type. We show this for the refinement attribute ‘sub-diagram’ whose range encompasses model instances of the type MT_{PM} :

$$\begin{aligned}
\text{domain}(\text{sub-diagram}) &= \{\text{Activity}\} \\
\text{range}(\text{sub-diagram}) &= \{MT_{PM}\} \\
\text{card}(\text{sub-diagram}) &= \langle 0,1 \rangle
\end{aligned} \tag{49}$$

3.2 Examples for Model Instances in FDMM

As FDMM not only permits to formally specify meta models but also the instantiation of meta models, we will illustrate in the following how the definitions of the previous section can be applied. We show the instantiation by using a procedure model and an object model. In Figure 1 on the next page a sample procedure model for an order process and in Figure 2 a sample object model are depicted as they have been later implemented. The instantiation of these model types is specified by:

$$\begin{aligned}
\mu_{MT}(MT_{PM}) &= \{\text{mt}_{pm1}\} \\
\mu_{MT}(MT_{OM}) &= \{\text{mt}_{om1}\}
\end{aligned} \tag{50}$$

In the next step we illustrate the instantiation of object types for these two model types. We show this for the object stores that are represented by circles and the activities that are represented by rectangles in figure 1:

$$\begin{aligned}
\mu_O(\text{Object-store}, MT_{PM}) &= \{\text{os}_1, \text{os}_2, \dots, \text{os}_7\} \\
\mu_O(\text{Activity}, MT_{PM}) &= \{\text{a}_1, \text{a}_2, \dots, \text{a}_5\}
\end{aligned} \tag{51}$$

For the assignment of textual information – as shown in the form of labels for the elements in figure 1 – we first have to instantiate attribute values of the corresponding data types, for example by:

$$\mu_D(\text{String}) = \{\text{'Website opened'}, \text{'Buy products'}, \dots\} \tag{52}$$

These data objects can then be used in triple statements to assign them to the object instances via their attributes, e.g.:

$$\begin{aligned}
(\text{os}_1 \text{ Name 'Website opened'}) &\in \beta(\text{mt}_{pm1}) \\
(\text{a}_1 \text{ Name 'Buy products'}) &\in \beta(\text{mt}_{pm1})
\end{aligned} \tag{53}$$

For the specification of the edges in the procedure model we first have to instantiate an object of type ‘Connection’ in (54) and can then use two triple statements for defining the start and endpoint in (55):

$$\mu_O(\text{Connection}, \text{MT}_{PM}) = \{c_1, c_2, \dots, c_{11}\} \quad (54)$$

$$\begin{aligned} (c_1 \text{ connection-from } os_1) &\in \beta(mt_{pm1}) \\ (c_1 \text{ connection-to } a_1) &\in \beta(mt_{pm1}) \end{aligned} \quad (55)$$

To illustrate how references to other object instances are specified we also take into account the object model shown in Figure 2. For XML nets the object stores in the procedure model are typed by referencing object elements in an object model. Therefore an instance of an object type has to be made available by: $\mu_O(\text{Object}) = \{\text{obj}_1\}$. Then we can reference this object in a triple statement:

$$(os_1 \text{ Object-type } \text{obj}_1) \quad (56)$$

3.3 Horus on ADOxx: A Prototypical Implementation

For the implementation of the formal specification in ADOxx some differences between the FDMM formalism and ADOxx had to be taken into account [6]. An important characteristic is that in ADOxx it is distinguished between classes and relationclasses that are used to connect classes. In contrast, FDMM does not make such a distinction but only refers to object types. Therefore, it had to be decided for which object types in FDMM classes should be defined and for which object types relationclasses. In addition, also the way how references between object types and object types and model types are used in FDMM differs from ADOxx. ADOxx provides the attribute type ‘interref’ that is used to specify references from class instances to either other class instances or model instances. However, when such a reference should possess additional attributes, so-called ‘record classes’ have to be used in ADOxx that permit the combination of several attribute types in the form of a record, i.e. a table-based structure. This directly corresponds to the use of an additional object type in FDMM.

For the realization of the Horus modeling language also additional constraints for the application of constructs on the model level had to be taken into account, apart from cardinality constraints. As FDMM currently does not provide constructs to express such constraints, these had to be added in ADOxx by using the AdoScript language. It is envisaged to add constructs for such constraint specifications to FDMM in the future. This concerned for example the requirement in procedure models that ‘connections’ can only connect instances of different classes, i.e. that an object store can only be followed by an activity and vice-versa.

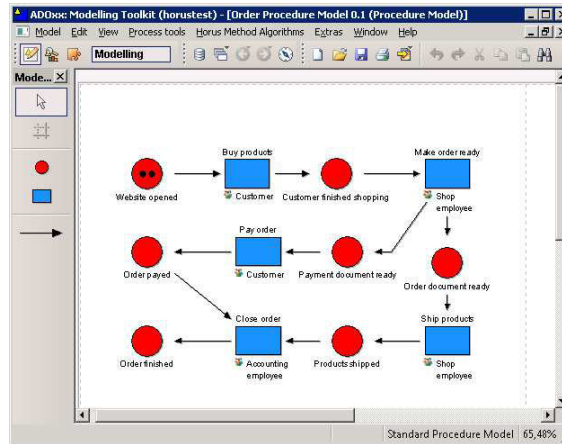


Fig. 1. Example for a Procedure Model in ADOxx

For the implementation of the Horus modeling language on ADOxx it was finally decided to create a total number of 17 model types that contain 44 classes, 9 relationclasses, and 11 record classes. From these classes 18 were defined as abstract classes. In addition, 161 attributes were created for the classes and 26 attributes for the relationclasses. Furthermore, for each non-abstract class and each relationclass a graphical representation was specified in the proprietary ADOxx-GraphRep grammar.

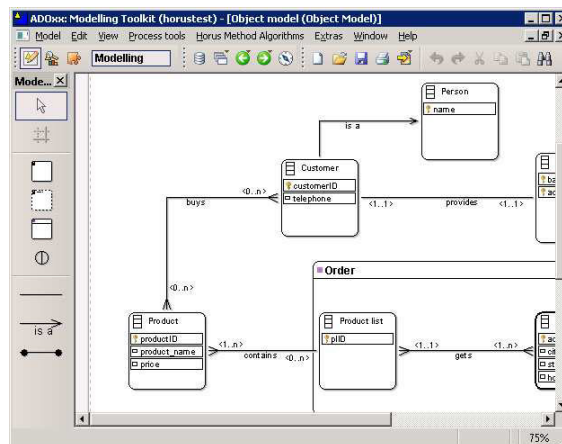


Fig. 2. Example for a Horus Object Model in ADOxx

All these tasks could be done using the graphical user interface of ADOxx for defining meta models and thus did not involve any programming effort. Overall, the implementation required about 60 person hours, not including testing, which is still be-

ing conducted at the time of writing this paper. The implementation will be made freely available in the context of the Open Models Initiative².

4 Lessons Learned

As modeling methods constitute an important means in the domain of business informatics, their exact specification is an important step to ensure a common understanding of the contained concepts between the designers of a modeling language and the developers of the modeling tool. In particular, the implementation of a modeling language on a meta modeling platform greatly benefits from a meta modeling oriented formal specification as it can be determined exactly how the language needs to be implemented. In addition, comparisons between the actual implementation and the specification can be made to ensure the quality of the implementation.

With the formal specification of the Horus modeling language by using the FDMM formalism, we could show how such a specification can be done in practice. Thereby, the FDMM formalism was directly applicable to the concepts of the Horus modeling language. However, the specification using FDMM also has some limitations. For example, in its current form it is not possible to formally specify the modeling procedure and the mechanisms and algorithms of a modeling method, i.e. the dynamic aspects. Also, the graphical representation of the elements and relations that constitute an indispensable part of many modeling methods in business informatics is currently not supported by FDMM. In regard to the implementation on the ADOxx platform, the differences between FDMM and the constructs in ADOxx had to be observed, e.g. the absence of explicit constructs for relations in FDMM that are available in ADOxx. Therefore, the use of FDMM specifications for the later implementation on ADOxx requires good knowledge of the platform functionalities. This also concerns the decisions on how to graphically represent the modeling language so that it is both user friendly and at the same time encompasses all required information for effectively creating models.

For the development of modeling tools, a formal description of the underlying meta model leads in particular to two main advantages: firstly, it has to be made explicit in an unambiguous representation, which concepts of a modeling language, including its elements, their relationships and the according attributes are provided to the users of a modeling language. Secondly, these descriptions can directly support the implementation of a modeling tool by providing a reference for the requirements of the implementation that can be exactly verified against the actual software implementation.

As a result of the application of FDMM to the Horus modeling method, we can derive the following requirements for a further extension of FDMM. To close the gap between the formalization in FDMM and the implementation on a meta modeling platform, a formal mapping between FDMM and the programming constructs of meta modeling platforms should be established. This mapping can then also be realized as a

² <http://www.openmodels.at/web/adoxx-horus-method/download>

software component for the automatic generation of meta modeling platform specific code from the formal specifications. Furthermore, the representation of dynamic aspects of modeling methods, such as the modeling procedure and mechanisms and algorithms that take into account the behavior of the modeling method, should be added to FDMM. Also these extensions can then be incorporated in the platform specific mapping to eliminate the step of manually translating the specifications into concrete code.

5 Related Work

In conclusion, two relevant areas can be distinguished to which the paper at hand can be related. First, the approach is compared to approaches in software development. Second, ADOxx and FDMM are compared to other meta modeling approaches.

The formal specification of meta models and models can be compared to formal specifications of software systems, e.g. by using graph transformations [15-16]. Similarly to our approach formal specifications in this domain are also used for making the consistency of the dimensions of a system explicit.

The FDMM formalism and ADOxx can be compared to other meta modeling approaches and formalizations of meta models and models. According to [14] FDMM and ADOxx can be directly compared to domain-specific modeling approaches which view meta models as *language specifications*. In this context, meta-meta models enable the automatic creation of graphical model editors from modeling language specifications. Kern et al. [4] provide an overview of such meta-meta models, which include for instance the GOPRRR/MetaEdit+ or the ARIS framework. Another group of meta modeling approaches uses meta models as *software structure specifications* (e.g. EMOF [13] or EMF [12]) and their formalisms consequently concentrate on the specification of software structures (see e.g. [11] for EMOF). In comparison, ADOxx does not focus on the generation of source code as e.g. intended by MetaEdit+ and neither requires programming skills. In comparison to the approach in [16], FDMM currently does not permit to express the dynamic behavior of modeling methods but rather focuses on the formalization of the abstract and the concrete syntax.

6 Outlook

Based on the insights gained by the application of FDMM to Horus and the implementation on ADOxx we will continue to work on facilitating the implementation of modeling methods. Possible next steps therefore concern the combination of the FDMM formalism with existing techniques for the automatic assignment of graphical representations [10] as well as the specification of model algorithms in a similar, non-programming oriented way. This will also be of further benefit for the implementation of the Horus method in order to specify additional graphical representations, e.g. to use domain-specific visualizations of Petri-nets, as well as for the realization of simulation functionalities for selected sets of model types.

References

1. Koch, S., Strecker, S., Frank, U.: Conceptual Modelling as a New Entry in the Bazaar: The Open Model Approach. Open Source Systems, Vol. 203, pp. 9-20. IFIP (2006)
2. Karagiannis, D., Fill, H.-G., Hoeffler, P., Nemetz, M.: Metamodeling: Some Application Areas in Information Systems. In: Kaschek, R., et al. (eds.): UNISCON 2008. LNBI, Vol. 5, pp. 175-188. Springer (2008)
3. Karagiannis, D., Kuehn, H.: Metamodeling Platforms. In: Bauknecht, K., Min Tjoa, A., Quirchmayr, G. (eds.): EC-Web 2002 – Dexa 2002. LNCS, Vol. 2455, p. 182. Springer (2002)
4. Kern, H., Hummel, A., Kuehn, S.: Towards a Comparative Analysis of Meta-Metamodels. In: 11th Workshop on Domain-Specific Modeling, Portland (2011)
5. Gulden, J., Frank, U.: MEMOCenterNG - A full-featured modeling environment for organization modeling and model-driven software development. In: Proper, E., Soffer, P.: Proceedings of the CAiSE Forum 2010, Hammamet (2010)
6. Fill, H.-G., Redmond, T., Karagiannis, D.: FDMM: A Formalism for Describing ADOxx Meta Models and Models. In: Maciaszek, L., et al. (eds.): ICEIS 2012, Wroclaw, Poland, Vol. 3, pp. 133-144 (2012)
7. BPTrends: The 2005 EA, Process Modeling and Simulation Tools Report - Adonis Version 3.81. Business Process Trends (2005)
8. Schönthaler, F., Vossen, G., Oberweis, A., Karle, T.: Business Processes for Business Communities - Modeling Languages, Methods, Tools. Springer (2012)
9. Lenz, K., Oberweis, A.: Inter-organizational Business Process Management with XML Nets. In: Petri Net Technology for Communication-Based Systems, Advances in Petri Nets. LNCS, Vol. 2472, pp. 243-263. Springer (2003)
10. Fill, H.-G.: Visualisation for Semantic Information Systems. Gabler (2009)
11. Favre, L.: A Formal Foundation for Metamodeling. In: Kordon, F. and Kermerrec, Y. (eds.): Reliable Software Technologies – Ada-Europe 2009. LNCS, Vol. 5570, pp. 177-191. Springer (2009)
12. McNeill, K.: Metamodeling with EMF: Generating concrete, reusable Java snippets, <http://www.ibm.com/developerworks/library/os-eclipse-emfmetamodel/index.html>
13. Object Management Group: Omg meta object facility (mof) core specification version 2.4.1, <http://www.omg.org/spec/MOF/2.4.1/PDF/>
14. Sprinkle, J., Rumpel, B., Vangheluwe, H., Karsai, G.: Metamodelling: state of the art and research challenges. In: Proceedings of the 2007 International Dagstuhl conference on Model-based engineering of embedded real-time systems. LNCS, Vol. 6100, pp. 57-76. Springer (2010)
15. Engels, G., Heckel, R.: Graph Transformation as a Conceptual and Formal Framework for System Modeling and Model Evolution. In: Montanari, U. et al. (eds.): ICALP 2000. LNCS, Vol. 1853, pp. 127-150. Springer (2000)
16. Engels, G., Soltenborn, C., Wehrheim, H.: Analysis of UML Activities Using Dynamic Meta Modeling. In: Proceedings of FMOODS 2007. LNCS, Vol. 4468, pp. 76-90. Springer (2007)

Applicability of Business Process Model Analysis Approaches – A Case Study in Financial Services Consulting

Jörg Becker¹, Mathias Eggert¹, Dennis Bruning¹, and Jan Saat²

¹ University of Münster, European Research Center for Information Systems (ERCIS),
Münster, Germany
{joerg.becker, mathias.eggert, dennis.bruning}@ercis.uni-muenster.de
² zeb/information.technology GmbH & Co. Kg, Münster, Germany
jsaat@zeb.de

Abstract. The analysis of business process models gains more and more attention in IS research. Several analysis approaches have been developed. All of them provide different features, such as syntax checking or pattern recognition. This paper investigates the applicability and relevance of business process model analysis approaches using a case study from financial services consulting. Two research contributions are provided. First, an overview about common model analysis features and its relevance for consulting processes are provided. Second, the applicability of the automatic business process model analysis approaches is investigated. Results show that the majority of features can raise efficiency of analyses in business process reengineering projects.

Keywords: business process analysis, model analysis, financial services, case study research, applicability check

1 Introduction

The financial services industry experiences a high degree of change which comes along with several challenges. Kohlmann [1] identifies new competition structures, regulatory and fiscal requirements (such as MaRisk, IFRS 9, or Basel III), changing customer demands and structures, product complexity, information technology, and competitive awareness as the central drivers of change initiatives. Consequently, the financial services industry is faced with a high competitive pressure, leading to specialization and consolidation [1-3].

In order to cope with an ever changing environment and the pressure to increase efficiency and effectiveness of business processes, financial institutions attempt to run large business process reengineering (BPR) projects [4] aiming to increase transparency and optimize the process management within the organization. Therefore companies use business process model repositories, i.e. databases for process models, as basis for analyses and optimization. Due to the high number of processes stored in

repositories and required industry knowledge, model analysis is often performed by consultants as they usually provide a cross-company perspective, best practices, industry standards, and benchmarks [5-6]. As more and more financial institutions engage in professional business process management, increasing requirements towards efficiency and effectiveness of process model analysis can be observed. IS research reacts to this demand by developing approaches that analyze business processes (semi-)automatically e.g.[7-10]. These approaches can assist in detecting syntax errors, compliance frauds or common optimization patterns. They have received a lot of attention in academia but or however, they lack empirical evidence in terms of applicability and relevance in practice.

The goal of this paper is to contribute to closing this gap and to provide insights into the perceived applicability of model analysis approaches (MAA) in practice. Considering the identified research gap, we aim to answer the following research question:

How relevant are automatic MAA capabilities for business process model analysis tasks in financial services industry consulting?

To answer this question, we conduct a case study in one of Germany's biggest financial services consulting companies and investigate the potential benefits of MAA capabilities for its consultants and BPR experts.

The remainder of this paper is structured as follows. In the next section, we briefly describe common features of business process MAA. Section 3 describes our research design and case setting and Section 4 presents our findings. Finally, in Section 5, we discuss our results and provide an outlook on further research in automated MAA.

2 Business Process Model Analysis Capabilities

Model analysis is often considered with model checking. The term "model checking" originates from computer science and is defined as "an automatic technique for verifying finite-state reactive systems, such as sequential circuit designs and communication protocols" [11]. We perceive business process model analysis in a less technical sense and define it as an automatic technique for verifying graph-based models in order to detect predefined patterns and execute search queries.

Common business process model analysis approaches have several different features which we briefly introduce in this section. Therefore we performed a structured literature search according to the procedure introduced by vom Brocke et al. [12]. In an initial step we searched all A and B journals from the Journal Ranking VHB Jourqual 2.0 (<http://vhbonline.org/service/jourqual/jq2/>), subsection information systems and information management and queried the database Ebscohost with the following search terms: model analysis approaches, process analysis, automatic process analysis, business process analysis, model analysis, automatic model analysis, automatic analysis. The search set was restricted to publications between 2006 and 2012 to ensure that only recent MAA were included. In addition to querying Ebscohost, forward and backward searches, starting from the initial result set, were conducted.

We analysed the results by first reading the title and abstract and, based on the relevance, we read the whole article and derive the features of the introduced approach. All features and relevant approaches are depicted in Table 1.

Structure Analysis is one of the fundamental analyses features for business process models. It is used to verify the correctness of the model in terms of completeness, consistency, and feasibility [13-14]. The structural analysis further reveals structural similarity between models, e.g. measured by distance metrics [15]. *Semantic Analysis* detects “the degree of similarity, based on equivalence between words” [16], such as labels, annotations or other semantic elements within the model. Semantic analyses may, for example, detect how similar two process models are based on the analysis of its event and activity labels.

The *Text / Label Analysis* covers a narrower model analysis, compared to the Semantic Analysis. Such analyses focus on analyzing and comparing element labels, model annotations and other text elements within the process model. In addition, MAA include word topologies, such as sentence structures or synonyms. The outcomes of such analyses are mainly distance metrics between two models, e.g. from a sample text, a company model, or process elements [13], [17-18].

Data / External Information Inclusion is of growing practical relevance. This model analysis feature includes elements which are not necessarily part of the analyzed model itself. External information sources or data can consist of linked models (such as organizational structure diagrams), data fragments from data-aware compliance rules, and annotations containing any forms of external information necessary for the process model [15], [19]. Thereby, the focus is set on inclusion and flow of external information, data or annotations. The included data is then used to create performance indicators, such as process costs or processing times [20].

The *Execution / Behavior Analysis* focuses on the process model and the transition states. The analysis often uses control-flow information to determine possible process states. Reachability or deadlock checks can be executed based on the information delivered by the execution and behavior analysis [16], [21].

Across Modeling Language Comparison, meaning the comparison of process models, developed with different modeling languages, becomes a more and more regarded feature. A typical application domain includes comparisons with reference models, as in the publications by Dongen et al. or Ghose and Kolliadis [15-16]. In addition, as a direct comparison is rather complicated, the MAA transforms the process models into a comparable modeling language before analyzing them [15].

Complex / Loop Construct Analysis enables the detection of complex process model structures, such as control flow loops and process execution rules. Dongen, Dijkman and Mendling use causal footprints for complex model analysis [16]. Awad and Sakr use a graph based representation to match graphs within repositories and therefore use the Business Process Modeling Notation Query language (BPMN-Q) within their approaches to handle and analyze models regardless their complexity [22].

Pattern / Model Construct / Sub graph Analysis is a feature that enables the analysis of a particular combination of model elements, such as Activity A must be executed before Activity B begins. Famous workflow patterns are, for example, described in

[23]. Thereby, patterns and constructs represent a specific rule that can express a structure, semantic, object, behavior or any other kind of requirement for a process model. MAA try to match the defined structure with the process model [24]. On the one hand, a sub graph can be a specific type of construct or pattern that can be searched for. On the other hand, pattern and construct analysis may use sub graph isomorphism [25] to reduce the analysis time and limit the analysis to defined sub graphs [26-27].

Table 1. Business Process Model Analysis Capabilities

	Structure Analysis	Semantic Analysis	Text / Label Analysis	Data / External Information Inclusion	Execution / Behavior Analysis	Comparison Across Modeling Languages	Complex / Loop Construct Analysis	Pattern / Model Construct / Sub-Graph Analysis	Tool Based Analysis / Visualization	Modeling Language Independence
Graph Matching Algorithms for Business Process Model Similarity Search (Dijkman et al. 2009) [12]	X	X	X		X		X			X
Measuring Similarity between Business Process Models (Dongen et al. 2008) [15]	X	X	X		X	X	X			X
On Enabling Data-Aware Compliance Checking of Business Process Models (Knuplesch et al. 2010) [18]				X			X		X	
Auditing Business Process Compliance (Ghose and Koliadis 2007) [14]	X			X	X			X	X	
Querying Graph-Based Repositories of Business Process Models (Awad and Sakr 2010) [21]	X	X		X	X		X	X	X	
Structural Patterns for Soundness of Business Process Models (Dongen et al. 2006) [25]							X	X	X	X
Pattern Identification and Classification in the Translation from BPMN to BPEL (García-Banuelos 2008) [20]	X				X		X	X	X	
Querying business processes with BP-QL (Beeri et al. 2008) [19]	X			X			X	X	X	
An Efficient Algorithm for Workflow Graph Structural Verification (Touré et al. 2008) [23]	X						X	X	X	
Clone Detection in Repositories of Business Process Models (Dumas et al. 2010) [16]	X		X					X	X	
Fast Business Process Similarity Search with Feature-Based Similarity Estimation (Yan et al. 2010) [17]	X		X	X				X		X
Fast Detection of Exact Clones in Business Process Model Repositories (la Rosa et al. 2012) [10]	X		X		X	X		X		X

The *Tool Based Analysis* feature expresses the support for model analysis by a software application. Therefore, it is not sufficient that the approach's analysis algorithm has been provided. Rather the feature is fulfilled whenever the approach has been implemented within a tool solution, providing visualization etc. Most approaches implemented the algorithms in a tool. If they do not provide tool support, the analyzed paper only provides the description of the analysis algorithm.

The last identified feature is *Modeling Language Independence*. Approaches that support this feature enable the analysis of process models developed with different languages [28]. Compared to the feature *Across Modeling Language Comparison*, modeling language independence does not focus on comparing process models. These approaches may rather analyze arbitrary process models individually. As Becker et al.

[28] argue that banks are not willing to change their common business process modeling technique for implementing MAA, this feature gets more and more attention.

3 Research Design

Given the high pressure to redesign business processes in financial institutes and the features of business process analysis approaches, we aim to explore the practical relevance of such analysis approaches. In particular, business process consultants are frequently requested to analyze and improve business processes [29-30]. Therefore, they can be seen as one key user group of business process MAA. In this paper, we investigate whether business process MAA are capable of supporting or even replacing manual model analysis tasks in financial services consulting.

Based on the findings of [4] who report on massive model repositories containing about 1,800 business process models, we assume enormous effort for consultants to analyze such a high number of business processes manually. In addition, the failure rate, meaning that not all process improvement capabilities or compliance frauds will be detected, is much higher compared to scenarios in which consultants make use of automation support.

The study at hand aims to investigate the relevance of MAA. Therefore we follow Rosemann and Vessey [31] and apply an applicability check prior to a typical research process. Applicability checks allow “[...] practitioners to provide feedback to the academic community on the research objects it produces or uses in theory-focused research” [31]. We thereby focus on separated semi-structured interviews with professionals of the financial services consulting domain.

The selected participants have been identified by applying the “snowball sampling approach” [32] which suggests accessing the best interviewees by asking interviewees for other potential interviewee suggestions. This approach is commonly used “to contact groups of people for whom there is no sampling frame” [32]. For this study, the interview topics have been specified at first, and then a snowball sampling took place. The investigated company is one of Germany’s biggest consulting companies for the financial services industry with customers in every segment of the industry. Altogether, five interviewees were selected, based on their expertise in BPR projects. One more participated in a pre-test which was conducted in order to ensure that all questions are understandable and unambiguous. The respondent details and its expertise are summarized in Table 2. Altogether, the face to face interview sessions led to the collection of about 37.700 words of transcribed data.

The interview was divided into nine question categories along the time slot of one hour. Question categories comprise of the following topics:

1. Short topic introduction, presentation of research subject and objectives, explanation of interview setting (audio recording, timeline), collection of demographics.
2. Presentation and evaluation of currently used model analysis approaches.
3. Exploration of model analysis tasks and model analysis requirements within the consulting process.

4. Collection of typical process flows and modeling problems (such as reoccurring weaknesses and difficulties and challenges).
5. Assessment of relevance and desired features of automatic business process MAA. Each interviewee rated the importance of each feature on a one-to-five scale.
6. Effort estimation (current, manual approaches vs. automated approaches).
7. Introduction of an automatic model analysis approach for previously discussed features. Presentation of sample use cases (interviewee is mostly passive here).
8. Mapping of model analysis features to the consulting tasks and process steps which resulted from categories 1 - 6.
9. Evaluation of the manual approach in comparison to the automatic approach. Collection of additional use cases in consulting and additional features desired.

Table 2. Respondent details

	Age	Position within the Consultancy	Unit	Department	Working Experience as Consultant	Employee of the Consultancy	Experience with Process-Consulting	Other
Participant I	31	Manager	IT	IT Strategy	6 Years	6 Years	6 Years	-
Participant II	29	Senior Consultant	Organization and Transformation	Process- and Qualitygmt.	4 Years	3 Years	2 Years	-
Participant III	35	Senior Manager	Organization and Transformation	Process- and Qualitygmt.	4 Years	4 Years	4 Years	Has been employed in Processmgmt in banking before
Participant IV	45	Senior Manager	Organization and Transformation	Banking Systems	12 Years	4 Years	8 Years	-
Participant V	46	Senior Manager	IT	IT Strategy	17 Years	5 Years	12 Years	EAM-Generalist
Average	37	Manager+	-	-	9	4	6	-

4 Study Findings

4.1 Business Process Reengineering Steps and Tasks

During the pre-test, we identified all the typical BPR stages at the consulting company. These stages are depicted in Figure 1 and defer from the well-known BPR stages, introduced by Hammer and Champy [33]. During the interviews we justified these stages and use them to classify the features of MAA in typical BPR projects.



Fig. 1. Business Process Reengineering Stages

Clarify Reengineering Objectives. In the beginning of a BPR project at the consulting company, the objectives have to be determined. The interviewees describe the business process analysis project as business and management driven. The objectives are identified by “document review of strategy papers, [conducting] top-management workshops, in order to find out in which direction you should proceed [...]” (interviewee four). According to interviewee number five, the central goals are to “struc-

ture the processes” and to analyze problems regarding “economic necessity” of business processes.

Create a Transparent As-is World. Reengineering projects require a deep and thorough understanding of the current state. Therefore, it is necessary to understand the as-is processes as well as respective strengths and pain points. This is usually reached by means of conducting workshops and reviews of existing documentation. According to interviewee five, process models are “only modeled deeply in areas where a positive effect can be expected”, meaning that solely for documentation purposes, they do not develop process models. Further, interviewee one said that for cost reduction projects, he “begins with the analysis of the current state to determine the cost drivers [...]”. When necessary, the elicited information is transferred into a process modeling tool of the consulting company and enriched by connecting it with other processes, organizational units, IT systems, and process performance indicators. These indicators might be “[...] key data, for example for project costs or capacities of the organizational units” (interviewee one) which are annotated to the organizational units and process steps by “[...] using tools such as ADONIS” (interviewee three). The process models are used to provide an overview on business areas and create a high level of transparency, interviewee one expressed.

Identify Requirements. In the subsequent step, drivers for change have to be identified which specify the requirements for the reengineering objectives and help to “[...] develop the target picture [...]” (interviewee one). The current state is analyzed to detect typical weaknesses in the business processes, interviewee three and four stated. When perceived as beneficial, best practice patterns will be applied. Manual business process analysis is supposed to avoid complex analysis (“[...] you do not need to create solutions with a complex analysis method, which are so complex that you actually can’t even use it”, interviewee one). Interviewees four and five described this analysis stage as a gap analysis between the as-is and to-be situation described by the business strategy. “This means that, out of the complete view, we try to operationalize the business strategy by breaking down the necessary operational and technological changes” (Interviewee five). If the objectives require far-reaching process changes, the as-is analysis is rather small since the business processes will be redesigned anyway. The analysis level always alters according to the project scope (“It means the level of detail is certainly dependent on what should be the results at the end”, interviewee four).

Develop To-be Situation. After specifying the requirements, a precise to-be situation is developed. By changing the parameters for the key factors influencing the reengineering objectives, continuously improved business process models are created, interviewee one stated. Interviewee three mentioned that “[...] the new processes [are also developed] as target-state-processes”.

Evaluate and Roll-out. After the development of the to-be situation, the to-be processes are created and implemented in the company. During the so called roll-out phase, the quality cycles ensure that the reengineering objectives are met and that the business continuation is guaranteed (“And then we pay attention at different phases during the roll-out and ensure the procedural quality through appropriate quality cycles”, interviewee three).

4.2 Weaknesses of the Manual Business Process Model Analysis

During the interviews, several business process analysis weaknesses evolve out of unclear process definitions. Redundancies in existing processes are common and market, product or services boundaries are not clearly set, which hampers the process analysis, interviewee five stated. Furthermore, the granularity often differs, which was mentioned as a particular problem in distributed modeling and analysis processes (“[...] there are too many people with different levels for detailed-modeling”, interviewee three).

The maturity of skills in process management, process documentation and process optimization is described as varying. Some companies provide and expect digital business process models; others provide process documentations and in some projects, information about business processes has to be gathered through interviews. “Some customers want specific tools and other customers have nothing and there is just a written instruction order and you have to go through available data, but in case of doubt, you’ll just conduct interviews”, interviewee five said. Business process models, even if digitalized, may not be maintained and are therefore outdated. They may also provide wrong data as inexperienced users make modeling mistakes. Moreover, the models may not provide information on the necessary level of detail or they are not representing the way processes are really handled within the company. This leads to the analysis of processes that are not described by models (“[...] we basically try to structure the processes which are not documented at all”, interviewee five). Moreover, even if modeling methods are used correctly, there are still gaps between the modeled information and the actual process.

Additionally, interviewee four mentioned that the application of professional process management tools is rare and mostly limited to companies which have a high process management competence. Most companies in the market use MS Visio or PowerPoint for process description and process analysis. Interviewee two additionally state that “processes are not documented properly. Processes will not be executed like they are modeled“. This leads to a lack of transparency in the as-is situation.

4.3 Importance of MAA Features for the Manual Business Process Analysis

The interviewees were asked for an overall rating of the feature importance independent of the process reengineering stages as well as an in depth evaluation of the feature importance for each BPR stage. Therefore, each interviewee rated the importance of each feature for the different phases on a scale

from one (meaning unimportant) to five (meaning very important). Figure 2 depicts the cumulated results. On the left side of the figure, the consultants rated the overall importance of the features, independently from the business process analysis stage. The “Structure Analysis”, the “Semantic Analysis” and the “Tool Based Analysis” are important for the process model analysis tasks. The “Text / Label Analysis” and the “Modeling Language Independent Analysis” are just of little importance for the interviewees.

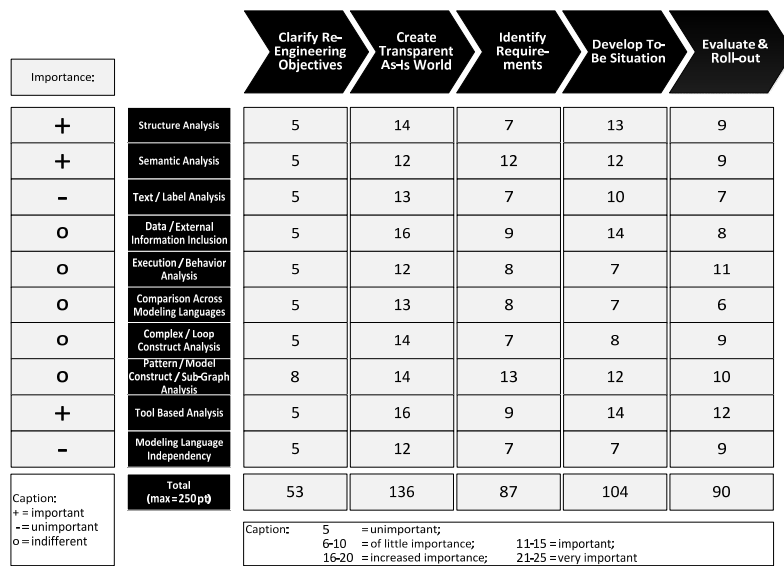


Fig. 2. Importance of Features for Manual Business Process Analysis

Regarding the assignment of analysis importance to process steps, three out of five interviewees expressed that the features of MAA are particularly relevant for the process steps “Create transparent as-is world” and “Develop To-Be Situation”. The stages “Identify requirements”, “Evaluation & Roll-out” are identified as less useful stages for the application of process model analysis features. The first BPR stage, “Clarify reengineering objectives”, is of minor to no importance for business process MAA.

4.4 Importance of Automatic Business Process Model Analysis

The rating of automatic model analysis features depends on the “level of automation of the analysis”, interviewee five said. A high level of automatic analysis is especially important if the analysis of “masses of models” is necessary.

Features Rated as Rather Important. It turned out that the analysis of Data / External Information Inclusion is of high importance for the business process reengineering tasks. In particular, “the analysis of attributes” (interviewee one), such as processing times, costs, locations, occurrence probabilities, and dependencies with other models

was described as important. The relationships between the process and further attributes and processes on other hierarchical levels are seen as important to understand the whole process model. Including external data and information into business process analysis ensures that all external information is included in the analysis (“I need the possibility to link to external data”, interviewee three).

Regarding the analysis approaches that provide Modeling Language Independency, interviewees stated that in contrast to the feature rating for manual process analysis, it is important to integrate models developed with different tools into the analysis tool and vice versa (“we used ARIS at Bank X [and] we found ADONIS at Bank Y”, interviewee five). We explain this rating difference with the brief discussion on that feature, e. g. by discussing a large merger between several independent banks using different modeling tools. The feature is especially relevant as the tool support that is currently available, is perceived to be insufficient for this task (“It is important for consulting [that] [...] in the end I can transfer all the processes into a tool which the customer already uses”, interviewee three).

Execution / Behavior is also been rated as important – especially for technical processes. Other processes are not in scope because they are perceived as “too inaccurate and fuzzy”. In addition, the combination of economic key factors, such as process costs, with the process simulation is considered to be highly relevant. The feature should be able to create a “what-would-be-if analysis regarding business ratios [...]” along the business process models.

The tool support is described as “important without a doubt”, interviewee five said. In particular, for large process model repositories, it is mandatory, mentioned interviewee three. The tools provide the possibilities for structuring as well as search and analysis functionalities across the models, which was also perceived as important.

Features Rated as Rather Neutral. The pattern, model-construct and sub-graph analysis is considered as semi important. Interviewee five perceives this functionality particularly important for “mass analysis”. The pattern analysis is, therefore, dependent on the usage scenario. Especially for routine tasks, where banks are confronted with different formal requirements which have to be checked in the process models, a pattern analysis “[...] is highly relevant”, interviewee five said. This analysis is seen as dependent on correct process and element labels (“The captions are highly relevant in order to detect [patterns] in large models”, interviewee three).

The interviewees have rated the Comparisons Across Modeling Languages as semi-important. Comparing different models, which can be created by using different languages, may be useful for understanding the model semantics. It (“[...] is more like a check of understandability”, interviewee five). The interviews have shown that this type of analysis has arbitrary application potentials, as very different modeling languages are used in companies (“I think it is important to analyze different syntax [...]”, interviewee four).

While three interviewees perceive Semantic Analyses as rather important, one interviewee considers this feature as less relevant (“Analysis approaches that consider the [semantic correctness] are rather irrelevant”, interviewee one). Interviewee three argues that “the bigger [the process models], the more important but also the more

difficult it becomes”. Semantic analyses become particularly important in distributed analysis settings. Furthermore, semantic checks are important if they can be used to identify the gaps between process models and real world process execution (“The focus of [semantic checks] is more on checking whether the modeled process represents the real-world process”, interviewee four).

The Structure Analysis has been evaluated as useful for decision tree analysis. Interviewees state the need to check each possible path of a process. The structure analysis can be used to identify the “catch-all branches” (interviewee five). The interviewees describe the task of creating consistency as important (“Checking inconsistency is rather important [...]”, interviewee four). The importance of structural correctness increases along with the increase of model complexity and model size (“Hence, the bigger the process quantity, the more complex the process landscape, the more important is this topic”, interviewee three).

Features Rated as Rather Unimportant. The Text / Label Analysis is recognized as supportive analysis. One major reason for the lack of support is that someone cannot “[...] trust these analyses”, interviewee five said. The correct naming of elements is considered important, but that has to be done during the modeling phase. This is in particular necessary in distributed modeling projects. Interviewee three stated that “due to the involvement of many different people with different modeling skills, the same process or process step will be described differently”. Furthermore, the Complex / Loop Construct Analysis is seen as an unimportant feature for the analysis. Interviewees state that they don’t have to search for such complex patterns.

4.5 Importance and Usage Scenarios along the Consulting Process Stages

After the tool- and presentation-based demonstration of the features of automatic MAA, the interviewees have been asked to state in which phase of the reengineering process they would apply automatic business process MAA and why. Therefore, the importance of MAA for each process reengineering stage should be rated by marking the relevant phases with a cross. The results are depicted in Figure 3, whereas the numbers indicate the cumulated crosses. All five interviewees rated the phase ‘creation of a transparent as-is world’ as an important application area for MAA.

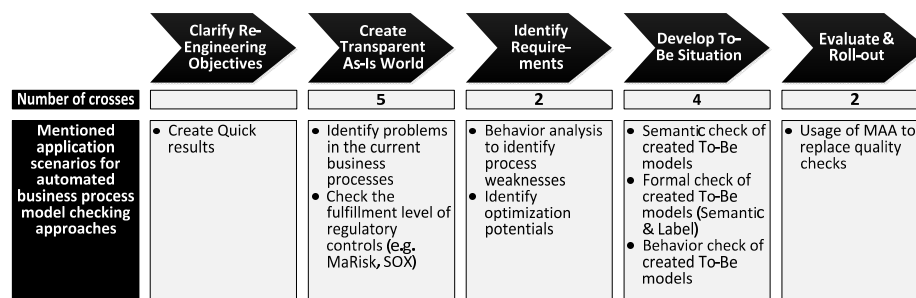


Fig. 3. Application scenarios of automated MAA for consulting process

Four interviewees see the development of the to-be situation as an important application area. The other phases show less importance. In general, the interviewees understand automated business process MAA as a chance to simplify manual analysis: “I think that we need more support in any case, the more automation the better”, interviewee five said. “In particular, in complex big projects, there exist meaningful application scenarios”, interviewee one said. Automated process model analysis can be used at different stages in the process reengineering project process. Although, not explicitly “checked” by the interviewees depicted in Figure 3, they mentioned that the approaches can support the project setup during the *Clarify reengineering goals* stage as well. The ability to use automatic model analysis for quick results as basis for clarifying the reengineering goals has been mentioned as relevant for this stage.

For *creating a transparent as-is world*, the automatic MAA seems to be useful to identify problems in the existing processes and to identify differences from, e. g., reference models. In particular, the semantic and syntax analysis are to be used for the task (“I can easily compare as-is processes with reference processes and detect differences”, interviewee five). In addition, tool support is important in this phase in order to “create transparent processes” and to “identify process weaknesses”, interviewee three said. The management of the regulatory requirements is another field of application. Therefore, the number of process models and necessary formal business process checks are seen as important application scenario. Within the as-is analysis, a completeness check is seen as one application scenario (“[...] in such environment [model analysis] is really relevant because you have a lot of regulatory requirements [...]”, interviewee five).

Within the *Identify Requirements* stage, comparisons across models and across modeling languages are identified as important for project success. In order to identify requirements, the interviews have shown that the process model comparison with “reference models and process landscapes” (interviewee four) is identified as one application area. Furthermore, behavior analysis to identify further requirements and optimization potential is perceived as a relevant application area (“You could make simulations based on expected case numbers in order to identify bottlenecks”, interviewee four).

During the stage *Develop To-Be Situation* stage, the automatic MAA is perceived as useful to analyze the to-be processes in order to ensure syntactical and behavioral correctness. Interviewee one identified the analysis of “completeness and clarity” and a “unique use of language” as an important application. Furthermore, the comparison of as-is and to-be models and checking “whether errors occurred during process modeling” (interviewee three) are seen as further use cases.

5 Discussion and Outlook

The study examines the potential of automated business process MAA within consulting projects in financial services industries. Interviews provide insights into the relevance of model analysis features and also pointed out potential use cases along typical project steps of business process reengineering. In general, business process MAA are

relevant for very large projects where large model repositories have to be analyzed. Lessons from consulting projects show that companies do not usually keep an accurately maintained repository ready for analysis, but that model translation efforts and alignment with “real world” processes is necessary prior to conducting model analysis.

The results indicate that business process MAA have the potential to support business process consulting in the financial services industry. Many features, such as the inclusion of external data and information, modeling language independent analysis, and execution / behavior analysis, are perceived as important for model analysis tasks. Secondly the results indicate that the applicability of MAA differ among the different consulting process steps. In particular, the creation of a transparent as-is world and the development of a to-be situation can benefit from automatic MAA. For the clarification of engineering objectives, no benefit could be identified.

Therewith, we provide the following two research contributions. First, we evaluate the practical relevance of MAA in a real-world scenario and provide an initial insight into its relevance. The results are in particular important for the ongoing development of research in model checking, as only little evidence for the practical relevance of MAA was given so far. Second, we provide an overview on the most important and less important features and their usefulness in different consulting stages. These insights can enable a focused ongoing development of MAA features. From a practical perspective, we provide some possible application scenarios for MAA which might encourage other consultancies to rethink their internal consulting processes in order to evaluate the usefulness of MAA for increasing BPR project efficiency.

Nevertheless, the presented study has some limitations. The case study comprises only a limited set of data. Thus, the results provide a first initial insight into in this area. Furthermore, the interviewees are employed at the same consulting company which might lead to some bias in their ratings. Interviewees might sometimes focus on one or two certain consulting phases and have only limited knowledge about the other phases which again biases the results.

As both intensity and frequency of BPR projects increase – especially in the financial services sector – one can predict increasing demand for analysis support in BPR projects. Research on automated business process MAA can contribute to these demands if developed in accordance with practical use cases. Therefore, we suggest two concrete enhancements of MAA. First, the *use of patterns and reference models should be extended*. Existing patterns and reference models (or model fragments) can be used to express standard queries for analysis. This could be, for example, deployed to analyze compliance issues in process models. Second, *unstructured data should be included in MAA*: According to the experience of the consultants in this study, few companies use one single modeling standard, language, or software solution (cf. results on feature inclusion of internal/external data). To overcome the syntax issues, tool-based converters and translators are promising to enhance the relevance of MAA. All in all, this study suggests that business process MAA are relevant for financial services industry consulting and encourage ongoing research in this area.

References

1. Kohlmann, F.: Geschäftsorientierte Referenz-Servicearchitektur für Banken. Logos, Berlin, Leipzig (2011)
2. Kaib, B.: Entwicklungslinien in Europa. In: Alt, R., Bernet, B., Zerdnt, T. (eds.): Transformation von Banken: Praxis des In- und Outsourcings auf dem Weg zur Bank 2015. Springer, Berlin Heidelberg (2009)
3. BaFin: Jahresbericht der Bundesanstalt für Finanzdienstleistungsaufsicht 2011, Bonn, Frankfurt am Main (2011)
4. Radulescu, C., Tan, H.M., Jayaganesh, M., Bandara, W., zur Muehlen, M., Lippe, S., Ljungberg, J., Andersson, M.: A Framework of Issues in Large Process Modeling Projects. In: Ljungberg, J., Andersson, M. (eds.): Fourteenth European Conference on Information Systems, pp. 1594-1605. Association for Information Systems (2006)
5. Alt, R., Bernet, B., Zerdnt, T.: Transformation von Banken: Praxis des In-und Outsourcings auf dem Weg zur Bank 2015 (Business Engineering). Springer, Berlin Heidelberg (2009)
6. Sadiq, S., Governatori, G., Naimiri, K.: Modeling control objectives for business process compliance. In: 5th international conference on Business process. LNCS, Vol. 4714, 149-164. Springer (2007)
7. Ekanayake, C., Dumas, M., Garcia-Banuelos, L., La Rosa, M., ter Hofstede, A.H.M.: Approximate Clone Detection in Repositories of Business Process Models. In: 10th International Conference on BPM. LNCS, Vol. 7481, pp. 302-318. Springer (2012)
8. Mendling, J.: Detection and prediction of errors in EPCs of the SAP reference model. Data & Knowledge Engineering 64, 312-329 (2008)
9. Sadiq, W., Orłowska, M.E.: Analyzing Process Models Using Graph Reduction Techniques. Information Systems 25, 117-134 (2000)
10. La Rosa, M., Dumas, M., García-Banuelos, L., Uba, R.: Fast Detection of Exact Clones in Business Process Model Repositories. Working Paper, accessible at http://eprints.qut.edu.au/49408/3/clones_main.pdf (2012)
11. Clake, E.M.: Model Checking. In: Goos, G., Hartmanis, J., van Leeuwen, J. (eds.): 17th Conference on Foundations of Software Technology and Theoretical Computer Science. Springer (1997)
12. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: Proceedings of the 17th European Conference on Information Systems, Verona, Italy, pp. 2206-2217 (2009)
13. Dijkman, R., Dumas, M., García-Banuelos: Graph matching algorithms for business process model similarity search. In: Dayal, U.e.a. (ed.): BPM 2009. LNCS, Vol. 5701, pp. 48-63. Springer (2009)
14. Soffer, P.: Goal-driven multi-process analysis. Journal of the Association for Information Systems 8, 175-202 (2007)
15. Ghose, A., Koliadis, G.: Auditing business process compliance. In: Krämer, B., Lin, K.-J., Narasimhan, P. (eds.): ICSOC 2007. LNCS, Vol. 4749, pp. 169-180. Springer (2007)
16. van Dongen, B., Dijkman, R., Mendling, J.: Measuring Similarity between Business Process Models. In: Bellahsène, Z., Léonard, M. (eds.): CAiSE 2008. LNCS, Vol. 5074, pp. 450-464. Springer (2008)
17. Dumas, M., Garcia-Banuelos, L., Rosa, M.L.R., Uba, R.: Clone detection in repositories of business process models. In: Rinderle-Ma, S., Toumani, F., Wolf, K. (eds.): 9th International Conference on BPM 2011. LNCS, Vol. 6896, pp. 248-264. Springer (2011)

18. Yan, Z., Dijkman, R., Grefen, P.: Fast business process similarity search with feature-based similarity estimation. In: Meersmann, R., Dillon, T., Herrero, P. (eds.): OTM 2010. LNCS, Vol. 6426, pp. 60-77. Springer (2010)
19. Knuplesch, D., Ly, L., Rinderle-Ma, S., Pfeifer, H., Dadam, P.: On enabling data-aware compliance checking of business process models. In: Parsons, J., Saeki, M., Shoval, P., Woo, C., Wand, Y. (eds.): Conceptual Modeling – ER 2010. LNCS, Vol. 6412, pp. 332-346. Springer (2010)
20. Beeri, C., Eyal, A., Kamenkovich, S., Milo, T.: Querying business processes with BP-QL. *Information Systems* 33, 477-507 (2008)
21. García-Banuelos, L.: Pattern Identification and Classification in the Translation from BPMN to BPEL. In: Meersman, R., Tari, Z. (eds.): On the Move to Meaningful Internet Systems: OTM 2008, pp. 436-444. Monterrey, Mexico (2008)
22. Awad, A., Sakr, S.: Querying graph-based repositories of business process models. In: Yoshikawa, M., Meng, X., Yumoto, T., Ma, Q., Sun, L., Watanabe, C. (eds.): 15th International Conference on Database Systems for Advanced Applications, Busan, South Korea. LNCS, Vol. 6193, pp. 33-44. Springer (2010)
23. Van der Aalst, W.M.P., ter Hofstede, A.H.M., Barros, A.P.: Workflow Patterns. *Distributed and Parallel Databases* 14, 5-51 (2003)
24. Touré, F., Baina, K., Benali, K.: An efficient algorithm for workflow graph structural verification. In: Meersman, R., Tari, Z. (eds.): On the Move to Meaningful Internet Systems: OTM 2008, pp. 392-408, Monterrey, Mexico (2008)
25. Ullmann, J.R.: An algorithm for subgraph isomorphism. *Journal of the ACM* 23, 31-42 (1976)
26. van Dongen, B.F., Mendling, J., van der Aalst, W.M.P.: Structural Patterns for Soundness of Business Process Models. In: 10th IEEE International Enterprise Distributed Object Computing Conference (EDOC'06), pp. 116-128, Hong Kong, China. IEEE (2006)
27. Jin, T., Wang, J., Wu, N., La Rosa, M., ter Hofstede, A.H.M.: Efficient and Accurate Retrieval of Business Process Models through Indexing. In: Meersman, R., Dillon, T., Herrero, P. (eds.): OTM 2010. LNCS, Vol. 6426, pp. 402-409. Springer (2010)
28. Becker, J., Bergener, P., Delfmann, P.: Modeling and Checking Business Process Compliance Rules in the Financial Sector. In: 32nd International Conference on Information Systems (ICIS 2011), pp. 1-19, Shanghai, China (2011)
29. Namiri, K.: Model-Driven Management of Internal Controls for Business Process Compliance. Dissertation. University of Karlsruhe (2008)
30. Pfeiffer, D.: Semantic Business Process Analysis: Building Block-based Construction of Automatically Analyzable Business Process Models. University of Muenster (2008)
31. Rosemann, M., Vessey, I.: Toward improving the relevance of information systems research to practice: the role of applicability checks. *MIS Quarterly* 32 (1), 1-22 (2008)
32. Bryman, A., Bell, E.: *Business Research Methods*. Vol. 3. Oxford University Press (2007)
33. Hammer, M., Champy, J.: *Reengineering the Corporation: A Manifesto for Business Revolution*. HarperBusiness, New York (1993)

An Empirical Investigation on the Design of Process Architectures

Monika Malinova¹, Henrik Leopold², and Jan Mendling¹

¹Vienna University of Economics and Business, Augasse 2-6, A-1090 Vienna, Austria
{monika.malinova, jan.mendling}@wu.ac.at

²Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
henrik.leopold@wiwi.hu-berlin.de

Abstract. Large-scale enterprises struggle with an effective alignment of business processes and IT services with business strategy. While process models play an important role for bridging between strategy and IT, there is a need to systematically organize the huge number of models. Process architecture defines an overarching structure for the organization of processes. However, there is a notable research gap on how process architectures are designed in practice. In this paper we address this problem by integrating insights and approaches from practice. We use Grounded Theory to analyze eleven in-depth interviews we conducted. Further, we present findings from studying documents provided by the interviewees. Our contribution is a conceptual framework about process architecture design, along with a classification of process architecture archetypes found in practice. Our results have strong implications since they demonstrate that process architecture design is more complex and context-dependent than assumed.

Keywords: Process Architectures, Process Modeling, Process Management

1 Introduction

Large-scale enterprises face an increasing challenge with aligning their business processes and IT services with business strategy. Traditionally, business process models play an important role for achieving the required level of transparency such that strategic decisions can be effectively operationalized. Corresponding modeling initiatives often yield several hundred, sometimes thousands of process models [1]. In this context, enterprises arrange their business process models in terms of a process architecture. In essence, a process architecture defines how the entire set of process models of a company can be systematically organized. In this way, process architecture can be regarded as a research area where business process modeling meets enterprise architecture.

At this stage, there are several studies available on success factors of business process modeling revealing different dimensions of organizational benefits [2-3]. In contrast, process architecture has hardly been subject to information systems research so

far, partially because it has been regarded as an art rather than an engineering discipline. Practitioners including Rob Davis [4] have summarized their design recommendations for process architecture. Also academic contributions tend more to the conceptual side [5-6]. However, we are not aware of any research exploring how companies design their process architectures and what factors influence the design.

In this paper we address this gap of research on process architecture. More specifically, our research objective is to explicate and integrate insights and approaches that have been implemented in industrial practice. To this end, we utilize a Grounded Theory approach, essentially because it has proven to be useful in settings where little explicit knowledge on phenomena and their relationships are documented. Our contribution is a conceptual framework about process architecture design, along with a classification of process architecture archetypes found in practice.

The remainder of this paper is structured as follows. We continue by revisiting essential concepts from business process modeling and process architecture. Then, we describe Grounded Theory as our research method. After that, we present the research findings in two parts. First, we discuss a conceptual framework for process architecture design which emerged from our analysis. Second, we define four process architecture archetypes resulting from our interviews and analysis of supplementary material provided by the interviewees. The subsequent section discusses the findings and emphasizes implications for research and practice. Finally, we conclude the paper with a summary and an outlook on future research.

2 Background

2.1 Business Process Management/Modeling

A business process is commonly defined as a sequence of activities, which is required for transforming input to some business-related output [7-8]. Accordingly, business process management can be understood as the set of all management activities that relate to business processes [9]. These activities are often structured in terms of a so-called business process management lifecycle, comprising the analysis, design, implementation, monitoring, and evaluation of business processes [9]. Business process models play an important role for many of these management activities. A model of a business process helps to systematically define the activities of a process along with relevant actors, inputs and outputs. Models are most often used in the analysis, design, and evaluation phase.

Many companies use process models not only for analyzing singular business processes, but also for systematically documenting their entire business operations. Such process modeling initiatives typically yield hundreds, sometimes thousands of process models. The usage and the benefits emerging from utilizing process modeling have been studied recently, but mainly with a focus on singular processes. Bandara and Gable find a set of success factors for process modeling [2]. These factors can be subdivided into project-related and modeling-specific factors. These factors have an organizational impact on different levels. Kock and Verville investigate the benefits

of mapping processes using flowchart-like notations [3]. They find that using a visual notation for processes increases process redesign success. While these insights are important for understanding benefits of modeling a singular process, they do not directly permit conclusions on the rationale for leading an entire process modeling initiative, and how the entire set of models shall be organized.

2.2 Process Architectures

The set of process models resulting from a modeling initiative typically requires a systematic classification. The design of such a classification is closely related to the concept of a process architecture. In an organizational context, an architecture is often understood as an abstraction of the enterprise, namely its elements of various types and their interrelations [4]. Consequently, a process architecture can be defined as a means for understanding the organization from a business process perspective [10]. It has been observed that organizations often focus on singular processes while failing to look at an integrated set of processes [11]. Therefore, one of the main concepts which characterize a process architecture is identifying the linkage between processes.

It is understood that enterprises can differ in various regards. From the perspective of research on process architecture, it is still not clear whether a range of architecture approaches is needed, or if one specific approach might be viewed as superior in all situations [12]. It is common practice in process modeling projects to employ a top-down approach of process development. This implies that processes are often decomposed into higher levels of granularity which provides greater detail on the constituent parts of the integrated process [13]. Therefore, a key element of a process architecture is the notion of model composition and decomposition in order to manage complexity at each architectural level [6]. Additionally, in some cases it may be required to structure the business processes relative to the process responsibilities and functional organizational structure for a given enterprise [14].

In literature, there are several suggestions on how companies should design their process architecture. Different reference models such as SCOR, eTom [15] or the Handels-H model [16] focus on specific domains and define architecture levels. Other works like ARIS [4], [17], the BPTrends BPM Pyramid [18] or the DoD Architecture Framework [19] take a more general, domain-neutral perspective. Both have in common that they provide a means of positioning the business processes on a predefined structure based on hierarchical decomposition. However, some authors have reported problems when applying these approaches in practice. For instance, Spanyi [20] emphasizes that reference models lack cross functionality and that this would be also the case for industry-specific models such as eTom. Interestingly, most authors suggest that a generic approach can be applied in any business setting. As a consequence, the decompositional process architecture proposed by zur Muehlen and Wisnosky solely focuses on the scope and the number of architecture levels [6]. However, the portability of this approach is explicitly emphasized. Hence, adapting the definition and the number of levels is considered to be an appropriate customization for any organization. A similar perspective with regard to the levels is taken by Frolov and Megel [5]. On the other hand, the potential impact of the specifics of a certain organization type

on the design of process architectures is indicated by the Riva Framework [21]. It suggests that, for instance, any university can adopt a similar process architecture because of belonging to the same type of business.

Considering the process architecture literature, we observe differing perspectives. Nevertheless, all approaches focused on the notion of decomposition and a universal process architecture design. In order to shed some light on the actual application of these concepts, we investigate the practice of process architecture design in an industrial setting.

3 Research Method

In order to gain a deep understanding of the applied process architecture approaches, we apply the systematic methodology *Grounded Theory*. The Grounded Theory method was developed in 1967 by Glaser and Strauss and supports inductive discovery of a theory that is grounded in data [22-23]. It is particularly suited and recommended for emerging topics where little research has been done before. Due to the lack of research on process architecture design we hence consider Grounded Theory to effectively support us in accomplishing our research goals.

In conformance with the method, we start with the data collection and then continue with coding of the material. Subsequently, we derive a model explaining how companies organize their business process models in a process architecture. Finally, we create a classification of the observed process architecture approaches.

3.1 Data Collection

We employed two different sources for our study. First, we conducted eleven semi-structured in-depth interviews with organizations from various industries documenting their business operations using process models. Our interviewees were BPM experts involved in all steps of the BPM implementation in the respective organization. In preparation for the interviews we designed an interview guideline containing specific questions about approaches used to organize the process models, and also general questions with regard to the business process modeling initiative. In total we classified the questions into the three categories *process documentation* (e.g. *How do you document your processes?*), *process identification* (e.g. *What are the process models used for?*) and *process architecture* (e.g. *Do you prioritize your documented processes?*). Table 1 gives an overview of the interview participants.

Table 1. Interview Participants

ID	Industry	Company Size	Years of BPM	Number of Processes	Documented Processes
1	Service/Retail	93	Not known	~1000	~20
2	Service/Retail	740	Not known	400	Not known
3	Service/Medical	~21000	Not known	Not known	Not known
4	Insurance	881	Not known	Not known	242
5	Service/Energy	313	1	Not known	Not known
6	Consulting	~4300	1	>150	~80
7	Service/Retail	~100	3	~100	~50
8	Service/Retail	~1000	3	Not known	120
9	Insurance	~5900	8	Not known	~350
10	Consulting	75	Not known	Not known	Not known
11	Consulting	~160	Not known	Not known	Not known

In order to maximize the opportunity to explore relevant insights, we included organizations varying in size, industry and their temporal experience with regard to process modeling. In addition to the interviews, we tried to collect complementary material from the interviewed organizations. In total seven companies provided us with internal documents visualizing and describing their process modeling practices.

3.2 Data Analysis

As in many qualitative studies, the data collection and the data analysis were conducted simultaneously. This enabled us to accordingly adapt questions, adjust the focus in the subsequent interviews and hence to effectively study the phenomena of process architecture. Following the Grounded Theory approach, the data analysis was done in three coding phases: open, axial and selective coding.

Open Coding. The *Open coding* phase is an analysis procedure with the goal of identifying concepts and categories in the data [23]. Thereby a concept is understood as the basic unit of analysis in the Grounded Theory method. Concepts are labels associated with happenings, events, and other instances of phenomena. A category is an abstract element grouping several concepts together. In line with the methodology, we started our analysis by going through the interview transcripts and tagging paragraphs or sentences with a discrete name, which properly reflected their content. We used the tool *ATLAS.ti* to keep track of all concepts, categories and conceptual relationships. In order to achieve representativeness and consistency, we iteratively went through the material and validated the concepts. For discovering categories from the identified concepts, we grouped concepts that are found to pertain to the same phenomenon.

Axial Coding. *Axial coding* is used to identify connections between the categories accordingly organizing them in a new way [23]. In order to accomplish this, we employed a coding paradigm with four columns: causal conditions, phenomenon, actions / strategies and consequences. We assigned each category to each of these columns by accordingly determining the role of the category with regard to the phenomenon. If, for instance, a category represents an event which leads to the occurrence of a certain process architecture type, we assigned this category to *causal conditions*. As a final

result, the initial categories (e.g. *Goal for Process Architecture Design*) turn into sub-categories of the main categories (columns) offered by the paradigm.

Selective Coding. The *Selective coding* phase is the process of selecting and focusing on a core category. Thereby, a core category is understood as the central phenomenon around all the other categories are organized [23]. The main goal of this phase is given by the refinement of the previously identified categories to a set of more abstract and focused categories. Accordingly, we iteratively went through the material and created more abstract categories related to the phenomenon. As a result, we derived a theoretical framework explaining the phenomenon of process architecture design.

Derivation of Process Architecture Approaches. In addition to the qualitative analysis of the interviews we also studied the complementary material. In particular, we complemented the information from the interviews with the graphical descriptions from the documentations. Focusing on the process architecture design, we were able to derive a set of different process architecture approaches. By abstracting from the specifics of the approaches and by focusing on differences and commonalities, we developed a classification of process architecture approaches.

4 Research Findings

As a result from the application of the Grounded Theory method, we derived a conceptual framework on process architecture design. Figure 2 illustrates the derived framework including its main categories, namely, the causal conditions, the phenomenon, the actions and strategies carried out for designing process architecture, and the according outcome of it.

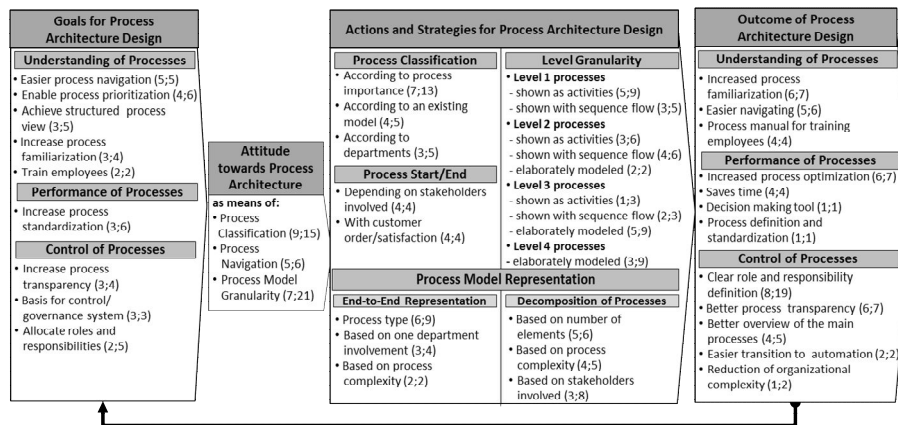


Fig. 1. Process Architecture Design Conceptual Framework

The values in the brackets next to each concept denote the number of companies which mentioned this concept and the total number of occurrences of this concept among all interviews. In the following subsections we explain the contents of the framework, accordingly using insights from our interview material.

4.1 Goals for Process Architecture Design

The goals for process architecture design can be classified into three categories: goals concerned with understanding, performance and controlling of processes. Most of our interviewees explicitly emphasized the role of understanding processes. They indicated the enablement of process prioritization as one of the major reasons for having a process architecture: *“We don’t just sit there and say ‘ok’, we have like a 1000 processes...”*(I9). In addition to aiming towards better understanding of processes, three of the companies pointed out the increased performance a process architecture would yield, especially in a case of a merger: *“...because of every merge the company changes and we had to consolidate our processes, so to avoid all the time moving around the processes we moved to ‘general terms’”*(I9). This is also closely connected with the goal of controlling processes by increasing the overall transparency of the company’s operations: *“It [the process architecture] will make it easier for us to achieve a complete transparency of our company”*(I5). As a result of the clear overview of their business processes, some organizations were also able to significantly increase the process familiarization on the different levels.

4.2 Attitudes towards Process Architecture

The phenomenon of our study is the attitude of companies towards the design of a process architecture. By this means, the attitude is based on the perceptions about future benefits that can be derived from introducing a process architecture. We observed that organizations particularly considered three specific use cases for process architecture design: process classification, process navigation and process model granularity.

Process Classification. Classification of processes is often considered as one of the most critical aspects. We observed that our interviewees emphasize the prioritization of their processes. *“...because these processes are used on a daily basis and contribute to the company’s strategy”*(I8). As a result, this focus often determines the design of the process architecture: *“All of the processes are aligned around our main process, ‘the execution of projects’ process”*(I10). Consequently, this could lead to managing organizational complexity: *“... to avoid the system getting more complex and burdensome over time”*(I3).

Process Navigation. Many of our interview partners emphasized the importance of transparency: *“The process architecture will make it easier for us to achieve complete transparency of our company”*(I5). In addition, the structured overview enabled easier navigation, which in turn could result in an increased organizational efficiency: *“The process architecture gives us an overview of the most important processes without going into depth, unless necessary”*(I8).

Process Model Granularity. More than half of the organizations wish to model their key processes at a higher level of granularity: *“Usually granularity is interrelated with the importance of the specific processes. The more important the process is for the company, the more detailed it is modeled”*(I11). A process architecture enables them to do so, while still having an overall view of the company’s operations.

4.3 Actions and Strategies for Process Architecture Design

Process Classification. The foremost issue when designing a process architecture is deciding on the manner of classifying the process models. Whereas some of our interviewees classify their processes based on an existing model: *“We use the Handels-H reference model to classify our processes”*K(18). Most of them do the same in accordance with the process relevance: *“Those processes that are executing the services our company offers and the orders from the customers are classified higher”*(I10). The manner organizations classify their processes may also reflect their initial goal for designing a process architecture. Thus, companies that claimed to classify their process models based on their departments, focused on roles and responsibilities allocation: *“... then we have process map for each department”*(I7).

Process Start/End. For defining the relationship between the processes, an organization needs to decide where one process ends and another begins. Our data shows that most companies used the stakeholders involved in the process execution as a separation criterion. If more than one stakeholder was involved in the execution of a single process, the process was decomposed accordingly: *“The process ends when one department did its job, and another department needs to continue”*(I9). While others based this on a customer order: *“The process starts when an order has been placed by a customer, and it ends when the customer has been served”*(I11).

Process Model Representation. Processes can be either represented in an end-to-end manner or as decomposed processes. The former is a whole process with a higher level of granularity. The later represents parts of a whole process at a lower level of granularity. Our findings show that some companies capture all details of certain process types, as for instance support processes: *“The support processes show the beginning and ending of the procedures, and what tasks are done in between”*(I3). Others do not practice process decomposition when only one stakeholder is involved: *“...so we don’t decompose the process if it is the same department that deals with it, but model it until the end no matter how big and complex the process gets”*(I9). Nevertheless, the complexity of a process is an important criterion for the representation decision: *“If the process gets too big, more than 15 elements, we try to decompose it”*(I8).

Level Granularity. The granularity of the process models belonging to the different levels of a process architecture varied among our interviewees. We found that the first level was mainly used to show processes represented as non-connected activities: *“The 1st level holds the major business processes shown as activities”*(I3). In the residual levels, we encountered several strategies. Whereas many companies used the second level for representing their process models forming a value chain, others already introduced detailed modeling. In the third level, most organizations captured elaborately modeled processes: *“Detailed modeling is done in the 3rd level”*(I4). In the case of this level still being used for holding non-detailed models, the total number of levels was accordingly higher.

4.4 Outcome of Process Architecture Design

As a result of the process architecture design, organizations experienced various outcomes. Congruent with the goals for designing a process architecture, these outcomes can be differentiated into three categories: outcomes concerned with the understanding, performance and control of processes.

Understanding Processes. We found that organizations used their process architecture as a manual for training employees: *“Also when there is a new employee, it is easier to show this person how the company works”*(I5). By showing the structured process models to new employees, they could gain knowledge not only of particular processes, but also of the overall organization.

Performance of Processes. More organizations than anticipated experienced process optimization as a result of the process architecture enabling process prioritization: *“...if a weakness in a process is discovered, the 4th level process ‘execute improvement’ is called to resolve this issue”*(I10). One organization even used their process architecture as a decision making tool: *“Setting up such an architecture helps in deciding if a certain request from a customer can be executed”*(I10).

Control of Processes. The findings show that an established process architecture significantly eased the allocation of roles and responsibilities to processes. It was stated that it is much easier to identify roles when the process models have been aligned on the different levels of an architecture. In many cases, the lower levels were used to allocate the roles or to define the collaboration between them: *“Levels 3 and 4 are used to divide the responsibilities among the employees”*(I3). In addition, a well-defined process architecture increases the process transparency among the different decision makers. Consequently, it contributed to a faster identification of the main processes: *“We have a better overview of the processes mostly used by our customers”*(I1). Moreover, the overall understanding of the business was improved: *“It makes it easier to understand the culture of the process, depending on the level it belongs to”*(I3).

5 Process Architecture Classification

As a result of the analysis of the complementary material, we derived a general classification of the observed process architectures. Although the majority of the companies designed their process architecture by hierarchically decomposing their processes, we also identified deviating approaches. At the center of our classification is the observation that process architectures vary in terms of the restrictions defined upon relationships between processes. Relationships between processes can be either horizontal (one process succeeded by another) or vertical (an activity of one process is decomposed into a whole process at more fine granular level). Restrictions relate to the cardinalities of these relationships. In the following we describe two major classes of process architectures: decompositional and service-oriented process architectures.

5.1 Decompositional Process Architectures

A Decompositional Process Architecture is a structure where activities of processes are decomposed into more fine-granular sub-processes. We observed three different ways how companies decomposed their processes. Hence, we identify three types of decompositional process architectures: the *Hierarchical Process Architecture*, the *Pipeline Process Architecture* and the *Divisional Process Architecture*. All have in common that a fine-granular process relates to exactly one activity in a more coarse-granular model.

Hierarchical Process Architecture. The Hierarchical Process Architecture is characterized by a hierarchy where each level captures processes having a particular granularity. Hence, a more detailed process model will be placed on one of the lower levels while more abstract models are assigned to higher levels.

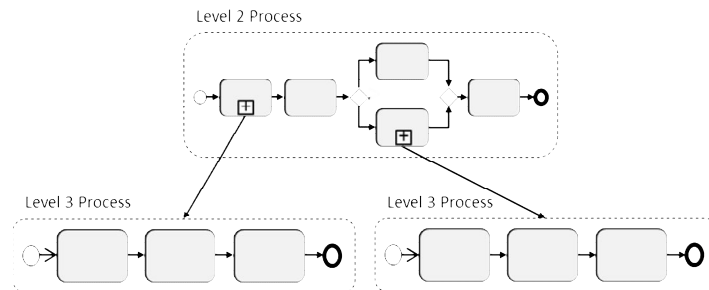


Fig. 2. Process Decomposition in Hierarchical Process Architecture

We observed that all interviewed organizations considered at least three levels, while further levels depended on the organization's complexity and their individual perception of it. Although many companies introduced terms such as main, core, or support processes, their definition of these terms varies significantly. Accordingly, we use numbers to refer to the decomposition levels. The upper level of the Hierarchical Process Architecture contains those processes which are considered to be most general and most relevant. Organizations select these processes depending on their contribution to the company's goals. Various relationships between these upper-level processes can be defined without any formal restrictions to the structure. If required, the first level processes are decomposed into second level processes. The decision whether a process is decomposed is based on its complexity, the stakeholders involved or the process type. As a result, a level 1 process might be decomposed in one or more level 2 processes. This principle is recursively applied down to the last level of the architecture. A restriction is here that a fine-granular process always has one single more coarse-granular process it relates to.

Figure 2 visualizes the decomposition relationship by showing a level 2 process and its decomposition into level 3 processes. In the depicted case, two activities of the level 2 process were considered to be excessively complex. Consequently, these activities were further specified on the underlying architecture level.

Pipeline Process Architecture. The Pipeline Process Architecture is a specialization of the Hierarchical Process Architecture. Similarly to the Hierarchical Architecture, it is characterized by a hierarchy of levels, which is capturing processes of different granularity. However, the relationship among the processes on each level and the relationship between two levels of the architecture is more restrictive. Essentially, a Pipeline Process Architecture only contains a single process. On the first level an all-encompassing process is subdivided into a number of processes. These processes relate to each other in terms of order such that they define a sequential chain of partial processes. In this way, the completion of the first level 1 process provides the trigger for continuing with the second process, and so forth. In this way, the level 1 processes form a sequential pipeline of processing. The relationships among the processes on the levels below follow the same logic. As a result, the formal structure of the relationships between the processes can be regarded as an ordered decomposition tree, which is processed in a depth-first way. On the right hand side of Figure 3 the Pipeline Process Architecture is visualized.

The Pipeline Process Architecture offers less flexibility due to the tight connection between the processes on the same level. We identified this process architecture type only in organizations facing a strict order of production. The most prominent example is given by an organization having a manufacturing process. In this case, each level 1 process was representing a manufacturing step of a product. The first level 1 process is initiated with the requirement of a new product, and the last level 1 process is finished when the product has been produced. Nevertheless, this architecture was also used by a company which solely focused on the customer order process.

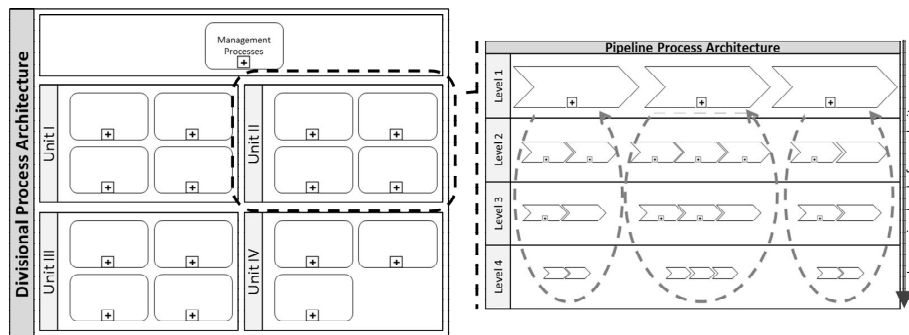


Fig. 3. Divisional Process Architecture – Pipeline Process Architecture

Divisional Process Architecture. The Divisional Process Architecture is an extension of the previously introduced approaches. Instead of just having one set of level 1 processes and their decomposition, the Divisional Process Architecture defines several process categories or so-called units. Thereby, each of these divisions contains a set of level 1 processes and their decomposition. The left hand side of Figure 3 visualizes this concept. The figure shows multiple units and their associated level 1 processes. Organizations applying this architecture first divide their processes into units. Thereby a unit can be derived via any suitable categorization criterion. We observed that

organizations commonly use business divisions as units as they often desire to have a clear cut between them. This indicates their necessity for a precise allocation of roles and responsibilities of their business processes. Accordingly, each unit contains the process models the associated department is concerned with. Each set of level 1 processes is further decomposed as shown in the previous architectures. Thus, in each unit the processes are either decomposed with the Hierarchical or the Pipeline Process Architecture. In addition to the units, the Divisional Process Architecture contains a set of management processes. These processes are visible to all units as they relate to the organization's strategy.

5.2 Service-Oriented Process Architecture

The second class of process architectures we identified is the Service-Oriented Process Architecture. It considerably differs from the perspective taken in the pure decompositional approaches. In a Service-Oriented Process Architecture, the processes are categorized in different groups. The main feature is the reuse of processes across the process categories. Hence, a single, more fine-granular process can relate to several more coarse-granular processes. This means that Service-Oriented Process Architecture does not include any strict level dependency as the strict top-down decomposition is replaced by a service-oriented perspective of reuse. Although each process in a process category may entail a hierarchical decomposition, the Service-Oriented Architecture provides more flexibility.

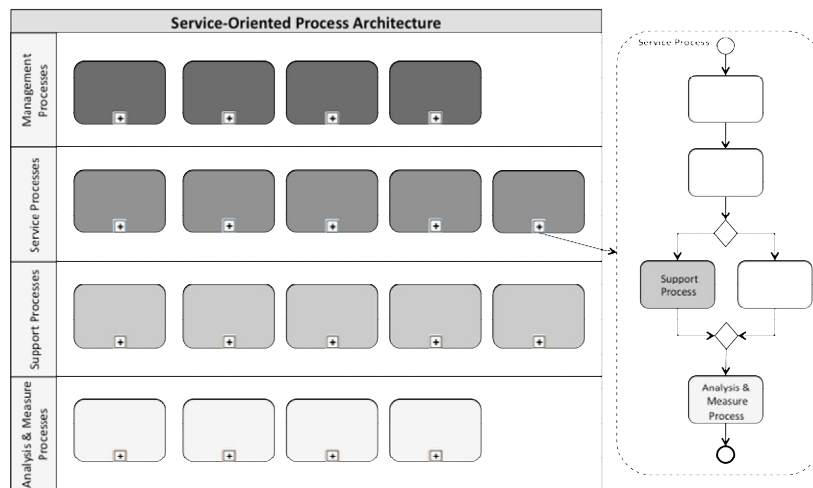


Fig. 4. Service-Oriented Process Architecture

Figure 4 illustrates an exemplary Service-Oriented Architecture including a small example process. It consists of four process categories: Management Processes, Service Processes, Support Processes and Analysis & Measure Processes. Among others, the management category could include processes such as risk and human resource

management. Processes from the service category could be concerned with the manufacturing of products. The support processes encapsulate functionality for the execution of the service processes. They consist of processes like invoicing, purchasing or IT support. The processes from the analysis & measure category are impartial processes that are called by all processes from the service or support category. They are used for handling errors that occur during process execution or to analyze costs. The model next to the architecture shows a process from the service category. For accomplishing its process goal it requires processes belonging to the analysis & measure and the support category. In the depicted example this is reflected by using the according tone of grey from the architecture visualization. As an example from the interviewed organizations, consider a sales process which is using a support process for issuing an invoice. Since tasks like issuing an invoice are required as service by various other processes, this example shows the benefits of a Service-Oriented Architecture. Instead of establishing a pure hierarchical decomposition, the redundant definition of processes is reduced.

6 Implications

There are two main implications for research and practice from our study, namely the impact of context on process architecture design and the hierarchical decomposition.

The interviews show that there are several commonalities in the way how process architectures are designed. However, we were able to identify different archetypes of process architectures. Apparently, there is a connection between the type of a company and the way a process architecture has to be designed. This observation has two implications. First, in terms of as-is modeling, there are companies that, according to their industry or business model, are closely associated with a particular architecture type. As an example, we have seen that a manufacturing company with a single product or service can be nicely described using a pipeline architecture, while a diversified enterprise would be better captured in terms of a divisional or service-oriented architecture. Second, in terms of to-be modeling, this observation suggests that certain types of process architectures are simply not appropriate for certain types of businesses. It would simply not be possible to design a pipeline process architecture for a diversified enterprise. This contextual perspective on process architecture design has mainly been ignored in practitioner-oriented contributions, which tend to work with a single approach, which is supposed to work in all circumstances [5-6].

We also observe commonalities in the organization of the different layers. Clearly, hierarchical decomposition plays a central role for organizing processes in an understandable way and for refining coarse-granular towards a fine-granular representation. This is in line with conceptual research on structuring process models in an easy to comprehend manner [24-25]. However, hierarchical decomposition may be broken from two directions: from a top-down perspective and from a bottom-up perspective. First, the top-down break of decomposition can be observed if enterprises have different product categories or different divisions. Those companies are likely to use a divisional process architecture. The corresponding top-level divisions can be defined

based on alternative criteria. For such an architecture, the relationship between a top level element and elements on the next level is not of a type decomposition, but rather that of belonging to a category. Second, the decomposition relationship is broken in the Service-Oriented Process Architecture. While in the first three architecture archetypes, at each level one element is further refined by a set of elements, we observe that here a fine-granular service process can be reused in several more coarse-grained processes on a more abstract level. This means that a service-oriented architecture is only in the upper levels of a decomposition tree, until a certain level of granularity is achieved where the reusable services are defined. The explicit representation of this complexity provides for a better traceability and alignment between processes and services.

7 Conclusion

In this paper we investigated the design of process architectures in the context of a qualitative study with Grounded Theory. Based on eleven in-depth interviews with different companies and seven sets of complementary material we derived a conceptual framework about process architecture design and a classification of process architecture approaches. Although our study showed that many companies follow decompositional approaches, we were also able to demonstrate that there are more specific and also deviating concepts. Particularly, we showed that the type and the structure of the company is an important factor for process architecture design. Thus, for instance the Pipeline Architecture should not be adopted by a diversified organization but rather by a business with a focused production line. Beyond that, we found that companies see the benefits of process architecture not only in terms of organizing and understanding processes, but also in terms of better performance analysis.

In future research we plan to extend our study with regard to several dimensions. At this stage, the insights presented in this paper are based on a limited set of interviews. Therefore, we aim to conduct additional interviews with partners from industries not yet covered. In this way, we want to find additional process architecture concepts and further relations between company characteristics and process architecture design. In addition, we plan to have a more detailed look into the modeling practice on the different levels of a process architecture. This may, for instance, include an investigation on the degree of granularity among processes on the different levels. As a result, a detailed understanding of process architecture can be acquired and a precise definition of their design could be provided.

References

1. Rosemann, M.: Potential pitfalls of process modeling: part A. *Business Process Management Journal* 12 (2), 249-254 (2006)
2. Bandara, W., Gable, G.G., Rosemann M.: Factors and measures of business process modelling: model building through a multiple case study. *European Journal of Information Systems* 14 (4), 347-360 (2005)

3. Kock, N., et al.: Communication flow orientation in business process modeling and its effect on redesign success: Results from a field study. *Decision Support Systems* 46 (2), 562-575 (2009)
4. Davis, R.: *ARIS Design Platform: Advanced Process Modelling and Administration*. Springer (2008)
5. Frolov, V., et al.: Building an ontology and process architecture for engineering asset management. In: *WCEAM 2009*. Springer (2009)
6. zur Muehlen, M., Wisnosky, D, Kindrick, J.: Primitives: Design Guidelines and Architecture for BPMN Models. In: *ACIS 2010 Proceedings*, paper 32. AIS (2010)
7. Becker, J., Kugeler M.: The Process in Focus. *International Journal of Operations & Production Management*, 1-12 (2003)
8. Davenport, T.H.: *Process Innovation: Reengineering Work Through Information Technology*. Harvard Business School Press (1993)
9. Mendling, J.: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction and Guidelines for Correctness. *LNBIP*, Vol. 6. Springer (2008)
10. Pritchard, J.-P., Armistead, C.: Business process management - lessons from European business. *BPMJ* 5 (1), 10-35 (1999)
11. Armistead, C.: Principles of business process management. *Managing Service Quality* 6 (6), 48-52 (1996)
12. Green, S., Ould, M.: The Primacy of Process Architecture. In: *3rd International Conference on Thinking*, pp. 1-6 (2004)
13. Armistead, C., Machin, S.: Implications of Business Process Management for Operations Management. *JoO & PM* 17 (9), 886-898 (1997)
14. Grady, J.O.: *System Management: Planning, Enterprise Identity, and Deployment*. Taylor and Francis Group (2010)
15. Cisco Systems: *Introduction to eTOM* (2009)
16. Becker, J., Hansmann, H., Rieke, T.: Architekturen von Informationssystemen. In: *Informationswirtschaft: Ein Sektor mit Zukunft*. LNI, P-33, pp. 103-118. GI, Bonn (2008)
17. Scheer, A.-W., Nüttgens, M.: ARIS Architecture and Reference Models for Business Process Management. In: *BPM*. LNCS, Vol. 1806, pp. 376-389. Springer (2000)
18. Harmon, P.: Business Process Trends. In *Handbook on Business Process Management*. Springer (2010)
19. DoD: *DoD Architecture Framework Version 1.5* (2007)
20. Spanyi, A.: Business Process Management Governance. In *Handbook on Business Process Management*. Springer (2010)
21. Green, S., Ould, M.: A framework for classifying and evaluating process architecture methods. *Software Process: Improvement and Practice* 10 (4), 415-425 (2005)
22. Glaser, B., Strauss, A.: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Publishing Company (1967)
23. Strauss, A., Corbin, J.: *Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory*. Sage Publications (1998)
24. Reijers, H.A., Mendling, J., Dijkman, R.M.: Human and Automatic Modularizations of Process Models to Enhance their Comprehension. *IS* 36 (5), 881-897 (2011)
25. La Rosa, M., et al.: Managing Process Model Complexity via Abstract Syntax Modifications. *IEEE Trans. on Industrial Inf.* 7 (4), 614-629 (2011)

Analyse des Qualitätsdiskurses zur Modellverständlichkeit in experimentellen Studien

Constantin Houy, Armella-Lucia Vella, Tom Thaler, Peter Fettke, und Peter Loos

Institut für Wirtschaftsinformatik (IWi)
im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI)
und Universität des Saarlandes, Campus, Gebäude D32, 66123 Saarbrücken, Germany
{Constantin.Houy, Armella-Lucia.Vella, Tom.Thaler, Peter.Fettke,
Peter.Loos}@iwi.dfki.de

Abstract. Die Erforschung der Qualität von Modellen gewinnt im Kontext der Unternehmensmodellierung zunehmend an Bedeutung. Die Verständlichkeit von Unternehmensmodellen stellt eines der bedeutendsten Kriterien der Modellqualität dar und es existiert mittlerweile eine erhebliche Menge experimenteller Arbeiten zu diesem Thema. Vor dem Hintergrund, dass dieses Qualitätskriterium aus verschiedenen Perspektiven sehr unterschiedlich aufgefasst werden kann, wird im Rahmen der vorliegenden Arbeit der auf Experimenten basierende wissenschaftliche Diskurs systematisch analysiert, um gängige Konzeptualisierungen und Operationalisierungen der Modellverständlichkeit zu untersuchen. Die Diskursanalyse zeigt, dass der Terminus Modellverständlichkeit in der experimentellen Forschung sehr unterschiedlich konzeptualisiert und operationalisiert wird, was erhebliche Implikationen für die Vergleichbarkeit der Forschungsergebnisse und die kumulative Forschung zur Unternehmensmodellierung hat.

Keywords: Unternehmensmodellierung, Modellverständlichkeit, Modellqualität, Diskursanalyse, experimentelle Forschung

1 Einleitung

Die Unternehmensmodellierung gilt in Theorie und Praxis als Erfolg versprechendes Instrument zur Gestaltung von Unternehmen. Unternehmensmodelle werden als methodische Hilfsmittel im Rahmen der initialen Informationssystementwicklung (*primary design*) und sämtlichen darauffolgenden Phasen der Weiterentwicklung (*secondary design*) [1] eingesetzt. Sie bieten u. a. erhebliche Potentiale für das Geschäftsprozessmanagement, die Softwareentwicklung sowie für die Auswahl, Einführung und Anpassung betriebswirtschaftlicher Standardsoftware. Gleichwohl ist davon auszugehen, dass Unternehmensmodelle ihre wichtige Funktion nur dann erfüllen, wenn sie eine angemessene Qualität besitzen. Somit sind Fragen zur Modellqualität von erheblicher Bedeutung für Forschung und Praxis.

Im Kontext der Modellqualität gilt vor allem die Verständlichkeit von Unternehmensmodellen als wichtiges Qualitätskriterium [2]. Eine hohe Modellverständlichkeit ist insbesondere dann von Bedeutung, wenn Unternehmensmodelle der menschlichen Kommunikation über die Funktionalitäten und die Struktur eines Informationssystems sowie der Schaffung eines kollektiven Verständnisses von diesen Systemen dienen sollen [3]. Aufgrund der Bedeutung der Verständlichkeit von Unternehmensmodellen existieren zahlreiche Forschungsarbeiten, die sich mit der Modellverständlichkeit auseinandersetzen. Aus verschiedenen vorliegenden Arbeiten wird indes ersichtlich, dass der Begriff *Modellverständlichkeit* dort sehr unterschiedlich interpretiert, konzeptualisiert und operationalisiert wurde. Unter Berücksichtigung dieser zahlreichen Perspektiven auf und Verständnisse von dem Begriff der Modellverständlichkeit kann dieser Begriff nicht nur als mehrdeutig bezeichnet werden, auch bleibt die Ausdeutung wichtiger Teilfacetten und Charakteristika bisher unklar.

Vor diesem Hintergrund setzt sich die vorliegende Arbeit das *Ziel*, zur Klärung des Begriffes Modellverständlichkeit beizutragen, indem im wissenschaftlichen Diskurs vorliegende Auffassungen von diesem Begriff detailliert herausgearbeitet, untersucht und transparent gemacht werden. Es wird dabei davon ausgegangen, dass die Identifikation, Rekonstruktion und Gegenüberstellung vorhandener Perspektiven auf dieses bedeutende Modellqualitätsmerkmal sowie seine gängigen Konzeptualisierungen und deren Operationalisierung die Entwicklung eines umfassenderen Verständnisses der Modellverständlichkeit fördert.

Zur Realisierung dieses Ziels wird im vorliegenden Beitrag eine *Methode* für die Analyse von Modellqualitätsdiskursen [4-5] im Kontext der Modellverständlichkeit angewendet und darüber hinaus die Potentiale dieser Methode für die Wirtschaftsinformatik weiter untersucht. Dabei wird der Diskurs zur Modellverständlichkeit analysiert, der auf experimentellen Arbeiten basiert. Die Konzentration auf Experimente liegt darin begründet, dass die Interpretation, Konzeptualisierung und Operationalisierung von Modellverständlichkeit in Experimenten stets direkt zugänglich sowie gut dokumentiert sind und sich somit eindeutig und transparent rekonstruieren lassen. Zwar liefern zuweilen auch andere empirische, bzw. konzeptionelle oder auch theoretische Arbeiten detaillierte Konzeptualisierungen des Begriffes Modellverständlichkeit. Allerdings weichen diese bezüglich ihrer Qualität stark voneinander ab und sind somit nur bedingt miteinander vergleichbar. Da für die Publikation experimenteller Arbeiten stets besonders hohe Mindestqualitätskriterien an die Konzeptualisierung und Operationalisierung von Konstrukten gestellt wird, werden im Rahmen der vorliegenden Arbeit ausschließlich Experimente betrachtet, um eine möglichst hohe Qualität der durchgeführten Diskursanalyse zu ermöglichen.

Der vorliegende Beitrag weist folgende *Struktur* auf: nach dieser Einleitung werden Grundlagen zu den Themen Unternehmensmodelle, Modellverständlichkeit und experimentelle Forschung in diesem Bereich eingeführt. Abschnitt drei erläutert den Forschungsansatz und beschreibt das verwendete Vorgehensmodell zur Diskursanalyse. Im vierten Abschnitt werden die Ergebnisse der Analyse des experimentellen Qualitätsdiskurses zur Modellverständlichkeit präsentiert, diskutiert und anschließend Implikationen für die experimentelle Forschung aufgezeigt. Abschnitt fünf gibt ein Resümee des Beitrags sowie einen Ausblick auf den weiteren Forschungsbedarf.

2 Zu Unternehmensmodellen, Verständlichkeit und experimenteller Forschung

Unternehmensmodelle sind Instrumente zur Klärung und Formulierung der Bedeutungen und Beziehungen von Begriffen innerhalb einer Domäne, um die Kommunikation zu unterstützen und Probleme aufgrund unterschiedlicher Interpretationen dieser Begriffe während der Entwicklung von Informationssystemen zu vermeiden [6-7]. Unternehmensmodelle sind bedeutende Artefakte für das Design und die Wartung von Informationssystemen, da sie auf eine übersichtliche Dokumentation und Kommunikation der Funktionalitäten und Strukturen abzielen [8]. Vor diesem Hintergrund ist es wichtig, dass Unternehmensmodelle qualitativ hochwertig sowie für alle Nutzer und Stakeholder leicht verständlich sind [9]. Versteht ein Nutzer oder Stakeholder ein Unternehmensmodell nicht, so ist die Kommunikation über den Modellinhalt, ebenso wie die Verifikation einer korrekten Umsetzung von Anforderungen kaum möglich [2]. Um den Erfolg der Unternehmensmodellierung zu verbessern, ist es deshalb von Bedeutung, die charakteristischen Faktoren zu identifizieren, welche die Verständlichkeit von Unternehmensmodellen beeinflussen. Aus diesem Grund haben die Forschung zur Modellverständlichkeit im Allgemeinen und die experimentelle Forschung im Speziellen, eine lange Tradition in der Forschungs-Community, die sich mit Unternehmensmodellierung beschäftigt.

Verständlichkeit kann als eine Art pragmatische Qualität von Modellen verstanden werden [10]. Sie bezieht sich auf die Einfachheit der Nutzung und somit auf den Aufwand, der für das Lesen, richtige Verstehen und Begreifen eines Modells erforderlich ist. Das Verstehen ist ein kognitiver Prozess, bei dem verschiedenen Teilen eines Unternehmensmodells Bedeutungen zugeordnet werden [11]. Im Gegensatz zu Aspekten, die einfach und objektiv zugänglich sind, z. B. die Anzahl der Elemente in einem Modell, ist es eine entscheidende Frage, *wie* Verstehensprozesse im Rahmen experimenteller Forschung beobachtet werden können, falls diese überhaupt objektiv zugänglich sind. Mentale Prozesse sind mit Emotionen, Intentionen oder Wünschen verbunden und deshalb nicht einfach zu greifen und zu erfassen. Solche Phänomene werden im Allgemeinen über *theoretische Konstrukte* erfasst, welche im Rahmen experimenteller Forschung untersucht werden und durch *beobachtbare Konstrukte* zugänglich zu machen sind [11]. Die Identifikation und Auswahl von beobachtbaren bzw. messbaren Konstrukten, die ein theoretisches Konstrukt valide abbilden können, ist ein nicht unproblematischer Vorgang. In der experimentellen Forschung wird in der Regel das Verhalten von Menschen beobachtet bzw. gemessen, um Rückschlüsse auf kognitive Prozesse ziehen zu können.

Vorliegende *experimentelle Arbeiten* in diesem Bereich zeigen, dass der zentrale Begriff *Modellverständlichkeit* bislang sehr unterschiedlich definiert, konzeptualisiert und operationalisiert wird. Somit bleiben der Begriff und seine Teilfacetten unklar. Ebenso ist nicht geklärt, inwiefern vorhandene Operationalisierungen bestimmten Anforderungen hinsichtlich der Validität gerecht werden und tatsächlich messen, was gemessen werden soll. Aus diesem Grund wird in der folgenden Untersuchung der experimentelle Diskurs zur Modellverständlichkeit systematisch analysiert und rekonstruiert, um einen Beitrag zur Klärung dieser offenen Punkte zu leisten.

3 Forschungsansatz der Diskursanalyse

Im Rahmen des vorliegenden Beitrags soll ein tiefergehendes Verständnis der Modellverständlichkeit auf Basis der Rekonstruktion und Analyse des auf Experimenten basierenden Qualitätsdiskurses zur Modellverständlichkeit erarbeitet werden. Unter einem *Diskurs* wird bildungssprachlich eine methodisch aufgebaute Abhandlung verstanden, die unterschiedliche Perspektiven auf und Meinungen zu einem Betrachtungsgegenstand zusammenfasst. Die Betrachtung von Diskursen im Bereich der Modellqualität ist notwendig, um der Eigenschaft des Begriffes Qualität gerecht zu werden, immer an die Perspektive unterschiedlicher Rollen (z. B. Modellentwickler oder Nutzer) gekoppelt zu sein [4-5].

Das Vorgehensmodell in Tabelle 1 gibt einen Überblick darüber, wie Diskurse zur Qualität von Unternehmensmodellen rekonstruiert werden können [4-5]. Ziel ist dabei die Identifikation und das Entwickeln eines Verständnisses von den relevanten Qualitätskriterien in einem Diskurs, die eine angemessene Bewertung von Unternehmensmodellen erlauben. Das Vorgehensmodell lässt sich in Wissenschaft und Praxis anwenden, um Qualitätsdiskurse zu Unternehmensmodellen zu analysieren und zu beurteilen. Nach der Darstellung des Vorgehensmodells in Tabelle 1 wird die detaillierte Ausgestaltung der einzelnen Phasen des Modells im Rahmen der hier durchgeführten Untersuchung beschrieben.

Tabelle 1. Vorgehensmodell zur Rekonstruktion und Evaluation von Modellqualitätsdiskursen

Vorgehensmodell für die Rekonstruktion und Evaluation von Modellqualitätsdiskursen
Schritt 1: Identifikation und Abgrenzung des Qualitätsdiskurses
Schritt 2: Rekonstruktion des Qualitätsdiskurses <ul style="list-style-type: none"> a. Definition von Modellqualität b. Konzeptualisierung von Modellqualität c. Operationalisierung / Messung von Modellqualität
Schritt 3: Validierung der Diskursrekonstruktion
Schritt 4: Analyse und Evaluation des Qualitätsdiskurses
Schritt 5: Gesamtbeurteilung

Im Rahmen der *Identifikation und Abgrenzung des Qualitätsdiskurses* werden systematisch alle relevanten Beiträge, die experimentelle Forschung zur Modellverständlichkeit präsentieren, recherchiert und analysiert. Bei der *Rekonstruktion des Qualitätsdiskurses* wird untersucht, welche Dimensionen von Modellverständlichkeit genannt werden und wie diese konzeptualisiert bzw. operationalisiert werden. Darüber hinaus wird ein Bezugsrahmen zur Strukturierung der identifizierten Verständnisse von Modellverständlichkeit entwickelt. Im Zuge der *Validierung der Rekonstruktion* sollen mithilfe der Rekonstruktion und Diskussion der Diskursperspektiven durch mehrere Wissenschaftler Unklarheiten und Inkonsistenzen aufgelöst werden. Vor dem Hintergrund der verwendeten diskursorientierten Konzepte und Methoden zur Be-

stimmung der Qualität von Diskursen zu Unternehmensmodellen wird im Rahmen der *Analyse und Evaluation des Qualitätsdiskurses* auch die Diskussion um die Qualität eines Modellierungsdiskurses selbst als Diskurs verstanden (*Modellierungsmetadiskurs*). Ziel und Zweck des Modellierungsmetadiskurses ist es, die Qualität des Modellierungsdiskurses zu untersuchen. Dadurch können Urteile über die Bewertung der Qualität der rekonstruierten Qualitätsdiskurse gefällt werden. Bei der *Gesamtbeurteilung* werden Hinweise zur abschließenden Evaluation des Modellqualitätsdiskurses unterbreitet, um das Verständnis des Betrachtungsgegenstandes aus den verschiedenen Perspektiven zusammenzufassen.

4 Analyse des experimentellen Qualitätsdiskurses zur Verständlichkeit von Unternehmensmodellen

4.1 Identifikation und Abgrenzung des Qualitätsdiskurses

Im Rahmen der hier vorgestellten Untersuchung wird der Qualitätsdiskurs zur Modellverständlichkeit insbesondere anhand experimenteller Beiträge analysiert. Es ist aufgrund der hohen Anforderungen an eine detaillierte Operationalisierung der untersuchten Variablen in Experimenten möglich, das zugrunde gelegte Verständnis von Modellverständlichkeit präzise zu rekonstruieren und zu untersuchen. Zur Identifikation und Abgrenzung des untersuchten Diskurses wurde ein systematisches Review [12] experimenteller Literatur zum Thema Verständlichkeit von Unternehmensmodellen unter Verwendung dreier international führender Literaturdatenbanken (*Science Citation Index*, *Scopus* und *EBSCO Business Source Premier*) durchgeführt. Dabei wurden insbesondere englischsprachige Beiträge recherchiert, da die einschlägige experimentelle Forschung vorwiegend in englischsprachigen Publikationsorganen veröffentlicht wurde. Relevante Beiträge wurden durch die Verwendung folgender Terme bei der Vorwärtssuche in Titel, Abstract und Keywords ermittelt: *understand**, *comprehen** kombiniert mit *conceptual model**, *process model** bzw. *data model** sowie *experiment**. Darüber hinaus wurden zusätzliche Arbeiten mithilfe einer Rückwärtssuche identifiziert sowie weitere, den Autoren bekannte Artikel zum Thema in die untersuchte Literaturmenge aufgenommen. Insgesamt wurden so 42 relevante Experimente identifiziert und untersucht.

4.2 Rekonstruktion des Qualitätsdiskurses

Zunächst wurden die Artikel hinsichtlich ihres Forschungsdesigns untersucht. Dabei wurden sämtliche Variablen, ihre Operationalisierung sowie insbesondere die Messinstrumente der abhängigen Variable („Verständlichkeit“) analysiert und dokumentiert. Tabelle 2 zeigt einen Ausschnitt aus der daraus resultierenden Dokumentation der identifizierten Konstrukte und ihrer Operationalisierung, die ebenso weitere Informationen zum Diskurs festhält, wie das Forschungsdesign eines Beitrags bzw. die unabhängigen Variablen. In der dritten Spalte beschreibt *N* die Teilnehmerzahl der jeweiligen Durchführung eines Experiments bzw. der Replikationen eines Experiments.

Tabelle 2. Übersicht der untersuchten Konstrukte und ihrer Operationalisierung (Ausschnitt)

Quelle	Forschungsdesign	N	Unabhängige Variablen	Abhängige Variablen	Messinstrument
1. Agarwal et al. 1999 [13]	Laborexperiment + Replikation, zufällig zugeordnete Teilnehmer	36 + 35	<i>Modellierungsansatz:</i> 1. Verwendung objektorientierter Modelle (<i>structure</i>) 2. Verwendung prozessorientierter Modelle (<i>behaviour</i>)	<i>Verständlichkeit:</i> 1. Genauigkeit des Modellverständnisses (<i>accuracy of model comprehension</i>)	1. Verständnistest: Bewertung von Teilnehmerantworten durch den Leiter des Experiments (7-Punkt-Likert-Skala) zu acht Verständnisfragen
2. Bavota et al. 2011 [14]	Quasi-experiment + zwei Replikationen	37 + 52 + 67	<i>Notation:</i> 1. UML-Klassendiagramme 2. ER-Diagramme <i>Modellierungserfahrung</i>	<i>Verständlichkeit:</i> 1. Grad des Modellverständnisses (<i>model comprehension level</i>) 2. Subjektiv wahrgenommenes Verständnis (<i>preference for notation</i>)	1. Verständnistest: Anzahl der korrekten Antworten zu 10 Multiple-Choice Fragen (eine oder mehrere richtige Antworten pro Frage) 2. Qualitative Einschätzung der Notationsvorlieben
3. Bodart et al. 2001 [15]	Drei Laborexperimente, zufällig zugeordnete Teilnehmer	52 + 52 + 96	<i>Komplexität der Modellrepräsentation:</i> 1. Ausschließliche Verwendung notwendiger Modellkomponenten (<i>Mandatory properties representation</i>) 2. Zusätzliche Verwendung optionaler Modellkomponenten (<i>Optional properties representation</i>)	<i>Verständlichkeit:</i> 1. Oberflächliches Verständnis (<i>surface-level understanding</i>) 2. Tiefes Verständnis (<i>deeper-level understanding</i>) (<i>response accuracy and problem-solving</i>)	1. Sieben Messinstrumente zur Korrektheit der Modellwiedergabe: Gesamtanzahl der korrekt wiedergegebenen Konstruktionseinheiten (Entitäten, Beziehungen, Attribute, Wiedergabe der Attribute und korrekte Darstellung, korrekte Wiedergabe der Beziehungen und Kardinalitäten etc.) 2. Antwortgenauigkeit: 10 Verständnisfragen, Antwortzeit (in Sekunden), normalisierte Genauigkeit (Verständnispunkte geteilt durch Schnelligkeitspunkte) und drei Messinstrumente zur Problemlösungsperformance hinsichtlich 9 Fragen (die Anzahl der korrekten Antworten basierend auf Informationen im Unternehmensmodell; (b) die Anzahl der korrekten Antworten der Teilnehmer basierend auf erweitertem Modellwissen; und (c) die Anzahl der falschen Antworten der Teilnehmer)
4. Burton-Jones et al. 1999 [16]	Laborexperiment, zufällig zugeordnete Teilnehmer	67	<i>Ontologische Klarheit des Modells (Ontological model clarity)</i> (Beziehungen in ERM mit/ohne Attribute) <i>Domänenwissen der Modellnutzer</i>	<i>Verständlichkeit:</i> 1. Güte der Modellbasierten Problemlösung (<i>problem-solving performance</i>) 2. Wahrgenommene Einfachheit des Modellverständnisses (<i>perceived ease of understanding</i>)	1. Messung der Problemlösung: Anzahl der akzeptablen Antworten zu sechs Problemlösungsaufgaben (PST), Anzahl der Antworten, die aus unterschiedlich repräsentierten Aspekten der beiden Gruppen resultieren (PSD) 2. PEU: Sechs Messinstrumente zur Messung der wahrgenommenen Einfachheit des Modellverständnisses
5. Burton-Jones et al. 2006 [17]	Laborexperiment, zufällig zugeordnete Teilnehmer + Replikation	57 + 66	<i>Modellzerlegung (model decomposition)</i>	<i>Verständlichkeit:</i> 1. Verständnis (<i>comprehension, problem-solving</i>) 2. Wahrgenommene Verständlichkeit (<i>perceived understanding</i>)	1. Problemlösungstest: Anzahl der akzeptablen Antworten zu Problemlösungsfragen und Lückentexten (Vermögen der Teilnehmer, eine Geschichte zur Domäne zu vervollständigen, Anzahl der ausgefüllten Lücken) 2. Vier Messinstrumente für das wahrgenommene Verständnis
6.

Nach der detaillierten Dokumentation der untersuchten Konstrukte im Diskurs erfolgten sowohl eine Analyse als auch ein Vergleich der identifizierten Verständnisse und der verwendeten Konzeptualisierung und Operationalisierung von Modellverständlichkeit. Im Rahmen dieser Untersuchung zeigte sich, dass der Begriff in den 42 Experimenten sehr unterschiedlich konzeptualisiert und gemessen wurde. Während beispielsweise Agrarwal et al. [13] lediglich eine Dimension von Modellverständlichkeit in ihrem Experiment untersuchen, die sie anhand inhaltlicher Verständnisfragen überprüfen, definieren Bodart et al. [15] unterschiedliche „Verständnistiefen“. Nach ihrer Auffassung entspricht einem *oberflächlichen Verständnis* von Modellen („surface-level understanding“) das korrekte Erinnern und „Reproduzieren“ von Modellteilen und einem tieferen Verständnis („deeper-level understanding“) das korrekte Beantworten inhaltlicher Fragen zu einem Modell sowie das Lösen von Problemstellungen mithilfe eines Modells.

Interessante Ergebnisse lieferte auch die Betrachtung der unterschiedlichen Instrumente für die Messung von Verständlichkeit. Während Agrarwal et al. [13] die Qualität der Beantwortung von Verständnisfragen durch die Leiter des Experiments anhand einer 7-Punkte-Likert-Skala subjektiv einschätzen und dokumentieren ließen, wurde bei Bodart et al. [15] die Anzahl korrekter Antworten bzw. Problemlösungen gezählt. Es zeigt sich folglich, dass nicht nur Verständlichkeit sehr unterschiedlich definiert, sondern auch die Messung von Verständlichkeit sehr unterschiedlich gestaltet wird.

Sämtliche im Rahmen der Rekonstruktion identifizierten Verständnisse des Begriffs *Modellverständlichkeit* wurden in Tabelle 3 in ein induktiv entwickeltes Kategoriensystem eingeordnet [5], [18], dass sowohl objektiv messbare bzw. subjektive als auch auf Effektivität bzw. Effizienz des Verstehens abzielende Dimensionen unterscheidet.

4.3 Validierung der Diskursrekonstruktion

Der im Rahmen dieses Beitrags analysierte Diskurs zur Modellverständlichkeit konnte aufgrund der experimentellen Grundlagenarbeiten präzise und auf transparente Art und Weise rekonstruiert und analysiert werden. Begründet liegt dieser Umstand vor allem in der hohen Qualität und der ausführlichen Dokumentation der Methodik der rekonstruierten experimentellen Arbeiten. An experimentelle Arbeiten werden i. d. R. hohe Anforderungen bezüglich der Konzeptualisierung und Operationalisierung der verwendeten Konstrukte erhoben. Die Tatsache, dass viele der untersuchten Experimente darüber hinaus in hochrangigen Publikationsorganen, wie z. B. dem *Journal Information Systems Research* (ISR), dem *Journal of the AIS* (JAIS) oder *Management Information Systems Quarterly* (MISQ) publiziert wurden, die ihrerseits hohe Ansprüche an die Operationalisierung der verwendeten Konstrukte stellen, trug deshalb erheblich zu einer hohen Qualität des Ausgangsmaterials der vorliegenden Diskursanalyse bei. Die Rekonstruktion konnte somit auf transparente Art und Weise durchgeführt werden. Traten im Rahmen der Rekonstruktion des Qualitätsdiskurses Unklarheiten oder unterschiedliche Meinungen bei den Autoren dieses Beitrags auf, z. B. bei der Einordnung von experimentellen Arbeiten in die Kategorien in Tabelle 3, so wurden diese Punkte diskutiert und ausgeräumt.

Tabelle 3. Untersuchte Dimensionen der Modellverständlichkeit

Quellen	Objektiv messbare Dimensionen					Subjektive Dimension
	Effektivität				Effizienz	Effektivität
	1. Wieder-gabe von Modell-inhalten	2. Korrektes Beantworten von Fragen zum Modell-inhalt	3. Prob-lem-lösung mit dem Modell	4. Verifi-kation von Modell-inhalten	5. Benö-tigte Zeit, um ein Modell zu verstehen	6. Wahrge-nommene Einfachheit des Modell-verständnisses
1. Juhn und Naumann 1985, [19]		•	•			
2. Palvia et al. 1992, [20]		•			•	
3. Shoval und Fruerman 1994, [21]		•				
4. Hardgrave und Dalal 1995, [22]		•			•	•
5. Kim und March 1995, [23]		•		•		•
6. Shanks 1997, [24]						•
7. Agarwal et al. 1999, [13]		•				
8. Burton-Jones und Weber 1999, [16]			•			•
9. Nordbotten und Crosby 1999, [25]		•				
10. Bodart et al. 2001, [15]	•	•	•			
11. Moody 2002, [26]		•		•	•	
12. Purchase et al. 2002, [27]		•			•	•
13. Parsons 2003, [28]		•			•	
14. Moody 2004, [29]		•		•	•	
15. Serrano et al. 2004, [30]					•	
16. Gemini und Wand 2005, [31]		•	•		•	•
17. Poels et al. 2005, [32]		•				
18. Sarshar und Loos 2005, [33]		•				•
19. Burton-Jones und Meso 2006, [17]		•	•			•
20. Khatri et al. 2006, [34]		•	•			
21. Cruz-Lemus et al. 2007, [35]	•	•	•			
22. Mendling et al. 2007, [36]		•				•
23. Recker und Dreiling 2007, [37]		•	•		•	
24. Serrano et al. 2007, [38]		•			•	
25. Burton-Jones und Meso 2008, [39]		•	•			•
26. De Lucia et al. 2008, [40]		•				
27. Genero et al. 2008, [41]		•			•	•
28. Mendling und Strembeck 2008, [42]		•				
29. Patig 2008, [11]		•			•	
30. Reijers und Mendling 2008, [43]		•				
31. Vanderfeesten et al. 2008, [44]		•				
32. Fuller et al. 2010, [45]		•				
33. Sánchez-González et al. 2010, [46]		•			•	
34. Bavota et al. 2011, [14]		•				
35. Figl und Laue 2011, [47]		•				•
36. Ottensooser et al. 2011, [48]		•	•			
37. Parsons 2011, [49]		•				•
38. Recker und Dreiling 2011, [50]		•	•		•	
39. Reijers und Mendling 2011, [51]		•				
40. Reijers et al. 2011, [52]		•				
41. Sánchez-González et al. 2011, [53]		•			•	
42. Schalles et al. 2011, [54]	•	•			•	

•: Verständlichkeitsdimension wurde im Beitrag betrachtet

4.4 Analyse und Evaluation des Qualitätsdiskurses

Im Rahmen des rekonstruierten Qualitätsdiskurses wird deutlich, dass die Qualitätsdimension *Modellverständlichkeit* sehr unterschiedlich interpretiert, konzeptualisiert und gemessen wurde. Dies führt zu einem ersten zentralen Ergebnis der Diskursanalyse: *Die Forschungsergebnisse hinsichtlich der Verständlichkeit von Unternehmensmodellen sind mehrdeutig und ohne eine klare Rekonstruktion der zugrunde liegenden Konzeptualisierungen des Verständnisses kaum vergleichbar.*

Verstehen zeichnet sich in den untersuchten Arbeiten durch unterschiedliche Dimensionen aus, u. a. durch die *Wiedergabe von Modellinhalten* oder *Problemlösung mit dem Modell*. Wie im Bezugsrahmen dargestellt, zeichnet sich Verständlichkeit im untersuchten Diskurs sowohl durch objektiv messbare und subjektive Dimensionen als auch durch auf Effektivität und Effizienz bezogene Dimensionen aus. Dabei fällt insbesondere auf, dass die Dimension *Korrektes Beantworten von Fragen zum Modellinhalt* in 39 von 42 der untersuchten Experimente betrachtet wird. Ebenso werden häufig die *Problemlösung mit dem Modell*, die *Benötigte Zeit, um ein Modell zu verstehen* sowie die *Wahrgenommene Einfachheit des Modellverständnisses* als wichtige Dimensionen von Modellverständlichkeit betrachtet, während die anderen identifizierten Dimensionen eine wesentlich geringere Rolle spielen. Dennoch wurden in allen untersuchten Arbeiten Forschungsergebnisse zum Thema Modellverständlichkeit präsentiert, welche jedoch ohne die explizite Berücksichtigung des Grundverständnisses im Detail unklar bleiben.

Wenn z. B. zwei Experimente die Verständlichkeit von UML-Klassendiagrammen und Entity-Relationship-Modellen (ERM) vergleichen, und eines davon die *Wahrgenommene Einfachheit des Modellverständnisses* untersucht, während das andere die *Benötigte Zeit, um ein Modell zu verstehen* misst, und beide zum Ergebnis kommen, dass UML-Klassendiagramme einfacher zu verstehen sind, sind diese Aussagen nicht direkt vergleichbar. Diese Details zur zugrunde gelegten Konzeptualisierung und Operationalisierung von Modellverständlichkeit ist unerlässlich für eine sinnvolle Interpretation der Ergebnisse experimenteller Forschung. Diese Sachlage mag vordergründig unproblematisch erscheinen. Gleichwohl ist sie insbesondere in denjenigen Fällen von Bedeutung, wenn für einen Leser bei der Lektüre experimenteller Forschungsergebnisse zunächst die Zusammenfassung der Ergebnisse, die typischerweise im Abstract oder am Ende eines Artikels präsentiert wird, im Vordergrund steht bzw. auch ohne genauere Prüfung beim Verfassen einer Forschungsarbeit zitiert würde, ohne dass die detaillierte Konzeptualisierung und Operationalisierung der Konstrukte im Methodenabschnitt einer zitierten Publikation berücksichtigt wird. Solche Fälle sind grundsätzlich eher als Ausnahme zu betrachten, allerdings liegen sie im Bereich des Möglichen.

Fragwürdig bleiben zuweilen auch die von den experimentellen Forschungsarbeiten präsentierten Messergebnisse und deren Vergleichbarkeit. Es wurde bereits das Beispiel der unterschiedlichen Messung der Modellverständlichkeitsdimension *Korrektes Beantworten von Fragen zum Modellinhalt* bei Agarwal et al. [13] (subjektive Einschätzung der Antwortqualität auf einer 7-Punkte-Likert-Skala durch den Leiter des Experimentes) und Bodart et al. [15] (Anzahl korrekter Antworten) angeführt.

Betrachtet man den gesamten rekonstruierten experimentellen Diskurs, so werden zwar gelegentlich vergleichbare Messinstrumente verwendet, z. B. offene Inhaltsfragen, Multiple-Choice-Inhaltsfragen, Lückentexte etc. Allerdings kommen nur sehr wenige *identische* Messinstrumente zum Einsatz. Ein Messinstrument, das beispielsweise von verschiedenen Untersuchungen gleichermaßen benutzt wurde und somit auch vergleichbare Ergebnisse produzieren kann, ist ein Instrument zur Erhebung der *Wahrgenommene Einfachheit des Modellverständnisses (perceived ease of understanding)*, das aus dem Kontext der Forschung zum *Technology Acceptance Model (TAM)* [55] stammt und für den Bereich der Modellverständlichkeitsforschung angepasst wurde. So bleibt es fragwürdig bzw. offen, inwiefern die Messergebnisse, die von den zahlreichen unterschiedlichen Instrumenten zur Messung der Modellverständlichkeit erhoben werden, als vergleichbar betrachtet werden können.

Eine weitere interessante Beobachtung ist die folgende: *Das Verständnis und die Konzeptualisierung der Verständlichkeit von Unternehmensmodellen in der experimentellen Forschung wurden während der letzten Jahre (1985-2011) nicht wesentlich weiter ausdifferenziert.*

Es könnte intuitiv angenommen werden, dass aus insgesamt 26 Jahren Forschung zur Modellverständlichkeit ein differenziertes Verständnis des Untersuchungsgegenstandes hervorgegangen wäre, was jedoch gemäß Tabelle 3 offensichtlich nicht der Fall ist [18]. Vorherige Arbeiten und die dort verwendeten Konzeptualisierungen und Operationalisierungen von Modellverständlichkeit werden nur bedingt in späteren Arbeiten wieder aufgegriffen, was aus Sicht einer kumulativen experimentellen Forschung als problematisch einzuschätzen ist. Die im Rahmen der vorliegenden Diskursanalyse verwendeten Dimensionen der Modellverständlichkeit, die in der publizierten experimentellen Forschung zugrunde gelegt werden, können zur weiteren Schärfung eines allgemeinen Verständnisses von Modellverständlichkeit und einer weiteren Ausdifferenzierung dieses Begriffs beitragen, was durch zukünftige Untersuchungen weiter zu klären ist. Die hier formulierten Beobachtungen haben bedeutende Implikationen, die im Rahmen der Gesamtbeurteilung des Diskurses weiter diskutiert werden sollen.

4.5 Gesamtbeurteilung

Im Rahmen der Analyse des Qualitätsdiskurses zur Modellverständlichkeit nehmen die Autoren des vorliegenden Beitrags die Perspektive eines einzelnen Teilnehmers am Modellierungs*metadiskurs* ein. An dieser Stelle beschränken sich die Ausführungen deshalb auf eine zusammenfassende Gesamtbeurteilung des hier analysierten Qualitätsdiskurses zur Modellverständlichkeit. Die Fülle von experimentellen Arbeiten und Publikationen im untersuchten Bereich zeigen interessante Unterschiede hinsichtlich ihres Verständnisses und der Messung von Modellverständlichkeit. Die Ergebnisse der präsentierten Diskursrekonstruktion bekräftigen die Annahme, dass die Berücksichtigung und Analyse sämtlicher verfügbarer Perspektiven einen interessanten Beitrag für ein tieferes Verständnis der Modellverständlichkeit leisten können. Während die Diskursrekonstruktion eine aufschlussreiche Übersicht über die verwendeten Konzeptualisierungen und Operationalisierungen von Modellverständlichkeit in

den 42 untersuchten Experimenten gibt, erscheint es umso interessanter, dass sich im Rahmen der langjährigen Forschung zur Modellverständlichkeit offensichtlich keine konsensfähige und scharfe Ausdifferenzierung des Modellverständlichkeitsbegriffs ergeben hat. In Kombination mit der potentiellen Mehrdeutigkeit und Unklarheit von Forschungsergebnissen aufgrund unterschiedlicher Grundannahmen, Konzeptualisierungen und Messinstrumente ist diese Entwicklung aus Sicht einer kumulativen Unternehmensmodellierungsforschung als ungünstig einzuschätzen, da im Rahmen des untersuchten Diskurses vorangegangene und verfügbare Forschungsergebnisse zur Modellverständlichkeit nicht immer berücksichtigt werden und die experimentelle Forschung in diesem Bereich in manchen Fällen wieder „von vorne“ zu beginnen scheint.

5 Resümee und Ausblick

Im vorliegenden Beitrag wurde der auf experimentellen Arbeiten basierende Qualitätsdiskurs zur Modellverständlichkeit systematisch mithilfe einer Methode zur Rekonstruktion und Evaluation von Modellqualitätsdiskursen analysiert. Vor dem Hintergrund der Prämissen, dass Modellqualitätskriterien im Allgemeinen und die Modellverständlichkeit im Speziellen von unterschiedlichen Diskursteilnehmern sehr unterschiedlich eingeschätzt werden können, zeigte die hier durchgeführte Diskursanalyse, dass eine multiperspektivische Betrachtung ein tiefergehendes Verständnis der Modellverständlichkeit fördern kann. In diesem Zusammenhang zeigte sich weiter, dass Modellverständlichkeit in der experimentellen Forschung sehr unterschiedlich konzeptualisiert und operationalisiert wird. Die Übersicht über die unterschiedlichen verwendeten Konzeptualisierungen und Operationalisierungen der Modellverständlichkeit kann insbesondere die zukünftige experimentelle Forschung unterstützen, indem sie einen Überblick über gängige Dimensionen der Modellverständlichkeit liefert und gleichzeitig zeigt, wie unterschiedlich diese im Rahmen von Experimenten angewendet werden. Neben diesen primär inhaltlichen Aspekten zeigten sich auch methodische Auffälligkeiten bezüglich des rekonstruierten Diskurses. So bleibt es bisher offen, inwiefern aufgrund der Heterogenität der verwendeten Messinstrumente eine Vergleichbarkeit der Ergebnisse gegeben ist. Es gilt, diesen Aspekt weiter und detaillierter zu untersuchen. Zur Unterstützung einer kumulativen Forschung, die breit gestützte Messergebnisse aus experimenteller Forschung benötigt, erscheint langfristig die Herausbildung „etablierter“ und möglichst konsistent eingesetzter Messinstrumente erstrebenswert.

Vor dem Hintergrund, dass die diskursbasierte Evaluation als Forschungsmethode in der Wirtschaftsinformatik im Allgemeinen und im Bereich der Modellqualität im Besonderen bisher eine untergeordnete Rolle spielt, kann die vorliegende Diskursanalyse zur Modellverständlichkeit auch interessante Potentiale und die Relevanz der Diskursanalyse für die Wirtschaftsinformatik aufzeigen. Diskursorientierte Betrachtungen des Qualitätsbegriffes – und hier insbesondere des Kriteriums der Modellverständlichkeit – fördern ein umfassenderes Verständnis bedeutender Qualitätsmerkmale von Unternehmensmodellen. Welche weiteren Potentiale und welche Grenzen der

verwendete diskursbasierte Ansatz für die Analyse von Qualitätsdiskursen im Allgemeinen im Rahmen der Wirtschaftsinformatikforschung anbieten kann bzw. aufweist, bleibt weiter zu untersuchen.

Danksagung: Die hier vorgestellten Forschungsergebnisse wurden im Rahmen des Forschungsprojektes „Pluralistische Beurteilung der Qualität von Unternehmensmodellen – Qualitätsdiskurse und Diskursqualität innerhalb der Wirtschaftsinformatik (*PluralistiQue*), gefördert durch die DFG (GZ LO 752/4-1), erarbeitet. Die Autoren möchten sich weiterhin bei den drei anonymen Gutachtern für die konstruktiven Anmerkungen bedanken, die zur Verbesserung dieser Artikel beigetragen haben.

Literatur

1. Germonprez, M., Hovorka, D., Gal, U.: Secondary Design: A Case of Behavioral Design Science Research. *Journal of the AIS* 12, 662-683 (2011)
2. Lindland, O.I., Sindre, G., Sølvberg, A.: Understanding Quality in Conceptual Modeling. *IEEE Software* 11, 42-49 (1994)
3. Krogstie, J.: Modelling of the People, by the People, for the People. In: Krogstie, J., Opdahl, A.L., Brinkkemper, S. (eds.): *Conceptual Modelling in Information Systems Engineering*. Springer, Berlin (2007)
4. Fettke, P., Vella, A.-L., Loos, P.: From Measuring the Quality of Labels in Process Models to a Discourse on Process Model Quality: A Case Study. In: Sprague, R. (ed.): *Hawaii International Conference on System Sciences (HICSS-45)*, pp. 197-206. Grand Wailea, Maui, HI (2012)
5. Fettke, P., Houy, C., Vella, A.-L., Loos, P.: Towards the Reconstruction and Evaluation of Conceptual Model Quality Discourses – Methodical Framework and Application in the Context of Model Understandability. In: Bider, I., Halpin, T., Krogstie, J., Nurcan, S., Proper, E., Schmidt, R., Soffer, P., Wrycza, S. (eds.): *BPMS 2012 and EMMSAD 2012. LNBP*, Vol. 113, pp. 406-421. Springer, Berlin (2012)
6. Wand, Y., Storey, V.C., Weber, R.: Analyzing the Meaning of a Relationship. *ACM Trans. Database Systems* 24, 494-528 (1999)
7. Maass, W., Storey, V.C., Kowatsch, T.: Effects of External Conceptual Models and Verbal Explanations on Shared Understanding in Small Groups. In: Jeusfeld, M., Delcambre, L., Ling, T.-W. (eds.): *Conceptual Modeling – ER 2011. LNCS*, Vol. 6998, pp. 92-103. Springer, Berlin (2011)
8. Wand, Y., Weber, R.: Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda. *Information Systems Research* 13, 363-377 (2002)
9. Moody, D.L.: Cognitive load effects on end user understanding of conceptual models: An experimental analysis. In: Benczúr, A., Demetrovics, J., Gottlob, G. (eds.): *Advances in Databases and Information Systems. LNCS*, Vol. 3255, pp. 129-143. Springer, Berlin (2004)
10. Moody, D.L., Sindre, G., Brasethvik, T., Sølvberg, A.: Evaluating the Quality of Process Models: Empirical Testing of a Conceptual Model Quality Framework. In: Spaccapietra, S., March, S.T., Kambayashi, Y. (eds.): *Conceptual Modeling - ER 2002. LNCS*, Vol. 2503, pp. 380-396. Springer, Berlin (2003)

11. Patig, S.: A practical guide to testing the understandability of notations. In: Proceedings of the Fifth Asia-Pacific Conference on Conceptual Modelling (APCCM '08). Wollongong, Australia (2008)
12. Fettke, P.: State-of-the-Art des State-of-the-Art - Eine Untersuchung der Forschungsmethode „Review“ innerhalb der Wirtschaftsinformatik. *Wirtschaftsinformatik* 48, 257-266 (2006)
13. Agarwal, R., De, P., Sinha, A.: Comprehending object and process models: An empirical study. *IEEE Trans. on Softw. Eng.* 25 (4), 541-556 (1999)
14. Bavota, G., Gravino, C., Oliveto, R., De Lucia, A., Tortora, G., Genero, M., Cruz-Lemus, J.A.: Identifying the weaknesses of UML class diagrams during data model comprehension. In: MODELS 2011, LNCS, Vol. 6981, pp. 168-182. Springer, Berlin (2011)
15. Bodart, F., Patel, A., Sim, M., Weber, R.: Should Optional Properties Be Used in Conceptual Modelling? A Theory and Three Empirical Tests. *Information Systems Research* 12, 384-405 (2001)
16. Burton-Jones, A., Weber, R.: Understanding Relationships with Attributes in Entity-Relationship Diagrams. In: De, P., DeGross, J.I. (eds.): Proceedings of the International Conference on Information Systems (ICIS 1999), pp. 214-228, Charlotte, North Carolina (1999)
17. Burton-Jones, A., Meso, P.N.: Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis. *Information System Research* 17, 38-60 (2006)
18. Houy, C., Fettke, P., Loos, P.: Understanding Understandability of Conceptual Models – What Are We Actually Talking about?. In: Atzeni, P., Cheung, D., Ram, S. (eds.): Conceptual Modeling - ER 2012. LNCS, Vol. 7532, pp. 64-77. Springer, Berlin (2012)
19. Juhn, S., Naumann, J.: The Effectiveness of Data Representation Characteristics on User Validation. In: Proceedings of the International Conference on Information Systems (ICIS 1985), pp. 212-226 (1985)
20. Palvia, T., Lio, C., To, P.: The Impact of Conceptual Data Models on End User Performance. *Journal of Database Management* 3, 4-14 (1992)
21. Shoval, P., Fruerman, I.: OO and EER Schemas: A Comparison of User Comprehension. *Journal of Database Management* 5, 28-38 (1994)
22. Hardgrave, B.C., Dalal, N.P.: Comparing Object Oriented and Extended Entity Relationship Models. *Journal of Database Management* 6, 15-22 (1995)
23. Kim, Y.-G., March, S.T.: Comparing Data Modelling Formalisms. *Communications of the ACM* 38, 103-115 (1995)
24. Shanks, G.: Conceptual Data Modelling: An Empirical Study of Expert and Novice Data Modellers. *Australasian Journal of Information Systems* 4, 63-73 (1997)
25. Nordbotten, J.C., Crosby, M.E.: The Effect of Graphic Style on Data Model Interpretation. *Information Systems Journal* 9, 139-155 (1999)
26. Moody, D.L.: Complexity Effects on End User Understanding of Data Models: An Experimental Comparison of Large Data Model Representation Methods. In: Wrycza, S. (ed.): Proceedings of the 10th European Conference on Information Systems (ECIS 2002), pp. 482-496, Gdansk, Poland (2002)
27. Purchase, H.C., Colpoys, L., McGill, M., Carrington, D.: UML collaboration diagram syntax: an empirical study of comprehension. In: First International Workshop on Visualizing Software for Understanding and Analysis, pp. 13-22, Paris, France (2002)
28. Parsons, J.: Effects of local versus global schema diagrams on verification and communication in conceptual data modeling. *Journal of Management Information Systems* 19, 155-184 (2003)

29. Moody, D.L.: Cognitive load effects on end user understanding of conceptual models: An experimental analysis. In: Benczúr, A., Demetrovics, J., Gottlob, G. (eds.): *Advances in Databases and Information Systems*. LNCS, Vol. 3255, pp. 129-143. Springer, Berlin (2004)
30. Serrano, M., Calero, C., Trujillo, J., Lujan, S., Piattini, M.: Empirical validation of metrics for conceptual models of data warehouse. In: 16th International Conference on Advanced Information Systems Engineering (CAISE'04), pp. 506-520, Riga, Latvia (2004)
31. Gemino, A., Wand, Y.: Complexity and Clarity in Conceptual Modeling: Comparison of Mandatory and Optional Properties. *Data & Knowledge Engineering* 55, 301-326 (2005)
32. Poels, G., Gailly, F., Maes, A., Paemeleire, R.: Object class or association class? Testing the user effect on cardinality interpretation. In: Akoka, J., Liddle, S.W., Song, I.-Y., Bertolotto, M., Comyn-Wattiau, I., Si-Said Cherfi, S., van den Heuvel, W.-J., Thalheim, B., Kolp, M., Bresciani, P. (eds.): *Perspectives in Conceptual Modeling*. LNCS, Vol. 3770, pp. 33-42. Springer, Berlin (2005)
33. Sarshar, K., Loos, P.: Comparing the Control-Flow of EPC and Petri Net from the End-User Perspective. In: van der Aalst, W.M.P. et al. (eds.): *Business Process Management*. LNCS, Vol. 3649, pp. 434-439. Springer, Berlin (2005)
34. Khatri, V., Vessey, I., Ramesh, V., Clay, P., Park, S.-J.: Understanding conceptual schemas: Exploring the role of application and IS domain knowledge. *Information Systems Research* 17, 81-99 (2006)
35. Cruz-Lemus, J.A., Genero, M., Morasca, S., Piattini, M.: Using Practitioners for Assessing the Understandability of UML Statechart Diagrams with Composite States. In: Hainaut, J.-L. et al. (eds.): *ER Workshops 2007*, LNCS, Vol. 4802, pp. 213-222. Springer, Berlin (2007)
36. Mendling, J., Reijers, H.A., Cardoso, J.: What makes process models understandable? In: Alonso, G., Dadam, P., Rosemann, M. (eds.): *Business Process Management*. LNCS, Vol. 4714, pp. 48-63. Springer, Berlin (2007)
37. Recker, J., Dreiling, A.: Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education. In: Toleman, M., Cater-Steel, A., Roberts, D. (eds.): *Proceedings of the 18th Australasian Conference on Information Systems 2007*, pp. 356-366, Toowoomba, Australia (2007)
38. Serrano, M., Trujillo, J., Calero, C., Piattini, M.: Metrics for data warehouse conceptual models understandability. *Inf. Softw. Technol.* 49, 851-870 (2007)
39. Burton-Jones, A., Meso, P.N.: The Effects of Decomposition Quality and Multiple Forms of Information on Novices' Understanding of a Domain from a Conceptual Model. *Journal of the AIS* 9, 748-802 (2008)
40. De Lucia, A., Gravino, C., Oliveto, R., Tortora, G.: Data model comprehension an empirical comparison of ER and UML class diagrams. *IEEE International Conference on Program Comprehension*, pp. 93-102. IEEE (2008)
41. Genero, M., Poels, G., Piattini, M.: Defining and validating metrics for assessing the understandability of entity-relationship diagrams. *Data & Knowledge Engineering* 64, 534-557 (2008)
42. Mendling, J., Strembeck, M.: Influence factors of understanding business process models. In: Abramowicz, W., Fensel, D. (eds.): *BIS 2008*. LNBIP, Vol. 7, pp.142-153. Springer, Berlin (2008)
43. Reijers, H.A., Mendling, J.: Modularity in process models: Review and effects. In: Marlon Dumas, M.R., Ming-Chien Shan (ed.): *Business Process Management*. LNCS, Vol. 5240, pp. 20-35. Springer, Berlin (2008)

44. Vanderfeesten, I., Reijers, H.A., Mendling, J., van der Aalst, W.M.P., Cardoso, J.: On a quest for good process models: The cross-connectivity metric. In: Bellahsene, Z., Léonard, M. (eds.): *Advanced Information Systems Engineering*. LNCS, Vol. 5074, pp. 480-494. Springer, Berlin (2008)
45. Fuller, R.M., Murthy, U., Schafer, B.A.: The effects of data model representation method on task performance. *Information & Management* 47, 208-218 (2010)
46. Sánchez-González, L., García, F., Mendling, J., Ruiz, F., Piattini, M.: Prediction of business process model quality based on structural metrics. In: Parsons, J., Saeki, M., Shoval, P., Woo, C., Wand, Y. (eds.): *Conceptual Modeling - ER 2010*. LNCS, Vol. 6412, pp. 458-463. Springer, Berlin (2010)
47. Figl, K., Laue, R.: Cognitive complexity in business process modeling. In: Mouratidis, H., Rolland, C. (eds.): *Advanced Information Systems Engineering*. LNCS, Vol. 6741, pp. 452-466. Springer, Berlin (2011)
48. Ottensooser, A., Fekete, A., Reijers, H.A., Mendling, J., Menictas, C.: Making sense of business process descriptions: An experimental comparison of graphical and textual notations. *Journal of Systems and Software* 85 (3), 596-606 (2012)
49. Parsons, J.: An Experimental Study of the Effects of Representing Property Precedence on the Comprehension of Conceptual Schemas. *Journal of the AIS* 12, 441-462 (2011)
50. Recker, J., Dreiling, A.: The Effects of Content Presentation Format and User Characteristics on Novice Developers Understanding of Process Models. *Communications of the AIS* 28, 65-84 (2011)
51. Reijers, H.A., Mendling, J.: A Study Into the Factors That Influence the Understandability of Business Process Models. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* 41, 449-462 (2011)
52. Reijers, H.A., Mendling, J., Dijkman, R.M.: Human and automatic modularizations of process models to enhance their comprehension. *Information Systems* 36, 881-897 (2011)
53. Sánchez-González, L., Ruiz, F., García, F., Cardoso, J.: Towards thresholds of control flow complexity measures for BPMN models. In: *Proceedings of the ACM Symposium on Applied Computing (SAC'11)*, pp. 1445-1450. TaiChung (2011)
54. Schalles, C., Creagh, J., Rebstock, M.: Usability of Modelling Languages for Model Interpretation: An Empirical Research Report. In: Bernstein, A., Schwabe, G. (eds.): *10th International Conference on Wirtschaftsinformatik*. pp. 787-796. Zurich, Switzerland (2011)
55. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 319-340 (1989)

Vom SOM-Geschäftsprozessmodell zur vollständig dokumentenorientierten RESTful SOA – Ein modellbasierter Ansatz

Matthias Wolf und Thomas Benker

Otto-Friedrich-Universität Bamberg, Lehrstuhl für Wirtschaftsinformatik, insbes.
Systementwicklung und Datenbankanwendung, Bamberg, Germany
{matthias.wolf,thomas.benker}@uni-bamberg.de

Abstract. Serviceorientierte Architekturen dienen als Aufgabenträger zur Automatisierung von Geschäftsprozessen. Um sicherzustellen, dass diese Aufgabenträger den Anforderungen der Geschäftsprozesse genügen, ist die modellbasierte Ableitung der softwaretechnischen Spezifikation, ausgehend vom Geschäftsprozessmodell, notwendig. Gegenstand der vorliegenden Arbeit ist die modellbasierte Abbildung von SOM-Geschäftsprozessen auf eine Zielarchitektur, die sich durch Dokumentenorientierung und den Architekturstil REpresentational State Transfer (REST) auszeichnet. Letzterer wird in der jüngsten Vergangenheit vermehrt zur Realisierung serviceorientierter Architekturen (RESTful SOA) diskutiert. Die softwaretechnische Spezifikation sieht dazu Vorgangs- und Entitätsservices vor, die über http-Verben GET, PUT, POST und DELETE mit auszutauschenden Zustandsinformationen aufgerufen werden. Für den Austausch werden JSON-Dokumente (Java Script Object Notation) verwendet. Die Durchgängigkeit der Dokumentorientierung wird dadurch erreicht, dass JSON-Dokumente nicht nur für den Austausch, sondern auch für die persistente Verwaltung herangezogen werden. Sie bilden den zentralen Entwicklungsgegenstand des vorgestellten modellbasierten Vorgehens, das anhand einer Fallstudie veranschaulicht wird.

Keywords: Modellbasierte Entwicklung, RESTful SOA, Dokumentorientierung

1 Einleitung

In aktuellen Untersuchungen werden serviceorientierte Architekturen (SOA) auf Basis des Architekturstils Representational State Transfer [1] (RESTful SOA) im Unternehmensumfeld intensiv diskutiert (vgl. [2-5]). SOAs im Allgemeinen dienen der Automatisierung von Geschäftsprozessen, welche häufig mit Geschäftsprozessmodellen gestaltet, analysiert und dokumentiert werden. Als Technologie zur Realisierung wird in der vorliegenden Arbeit REST zugrunde gelegt. Ein Vergleich mit SOAP/WSDL-Services wird ausführlich in den Beiträgen [2-4] diskutiert. Die Vorteile des Einsatzes von REST werden dabei hinsichtlich Flexibilität, Interoperabilität

und Skalierbarkeit für ad-hoc Web-Integrationsszenarien im betrieblichen Umfeld gesehen.

Bisherige Arbeiten zur Entwicklung von RESTful SOA konzentrieren sich primär auf die softwaretechnische Spezifikation der Aufgabenträgerebene. Geschäftsprozesse als Spezifikation der fachlichen Anforderungen auf Aufgabenebene und die davon ausgehende modellbasierte Ableitung serviceorientierter Anwendungssysteme finden aktuell in der Literatur nur punktuell Berücksichtigung (siehe Kapitel 2). Durch die modellbasierte Ableitung aus Geschäftsprozessmodellen wird jedoch zum einen die Konsistenz zwischen Implementierung und Dokumentation erhöht. Zum anderen wird ein Beitrag zur Beherrschung der Komplexität der Entwicklungsaufgabe geleistet, der sich ein Entwickler ausgesetzt sieht. Hierzu ist ein möglichst hochgradig automatisierbarer modellbasierter Entwicklungsprozess anzustreben. Als Teilaufgaben der Entwicklung sind v.a. die Identifikation von Ressourcen (REST-Services) als zentrale Komponenten der RESTful SOA sowie die Spezifikation zugehöriger Zustandsbeschreibungen in Form von Dokumenten zu nennen. Letztere sind innerhalb von Ressourcen persistent zu verwalten und zwischen diesen auszutauschen. Als Formate, nicht nur zum Austausch, werden häufig XML oder die JavaScript Object Notation (JSON) genutzt [3-4]. Jüngste Entwicklungen unter dem Begriff Dokumentendatenbanken ermöglichen die Persistierung der Zustandsinformationen in der Form ihrer Nutzung durch das Anwendungssystem. JSON-Dokumente sollen in der vorliegenden Arbeit das einzig zu nutzende Datenformat und einen zentralen Entwicklungsgegenstand bei der angestrebten modellbasierten Ableitung einer RESTful SOA darstellen. Dies manifestiert sich in dem Begriff der *vollständigen Dokumentenorientierung*.

Die daraus ableitbare Problemstellung dieser Arbeit lässt sich als Konstruktionsproblem nach [6] klassifizieren und innerhalb der gestaltungsorientierten Wirtschaftsinformatik nach [7] positionieren. Den Untersuchungsgegenstand bildet das modellbasierte Entwicklungsvorgehen zur Spezifikation einer vollständig dokumentenorientierten RESTful SOA ausgehend von einem Geschäftsprozessmodell. Hierzu wird das Semantische Objektmodell (SOM) nach Ferstl und Sinz [8] zugrunde gelegt und bzgl. des Untersuchungsziels, der methodisch geleiteten Überbrückung der semantischen Lücke zwischen Geschäftsprozessmodell und service-orientierter Aufgabenträgerbeschreibung, erweitert. Der Kernbeitrag der vorliegenden Arbeit setzt sich aus zwei wesentlichen Bestandteilen zusammen. Zum einen erfolgt die Spezifikation der SOA, von einem Geschäftsprozessmodell ausgehend, modellbasiert. Damit lassen sich Flexibilität, Nachvollziehbarkeit und der Beitrag zur Beherrschung der Komplexität der Entwicklung service-orientierter Anwendungssysteme steigern. Zum anderen werden die von Geschäftsprozessen zu verwaltenden Dokumente in den Mittelpunkt der Betrachtung gerückt, und die Ausgestaltung der SOA hinsichtlich der Vermeidung von Medienbrüchen an diesem Entwicklungsgegenstand ausgerichtet. Wesentliche Herausforderungen der Untersuchung bestehen dabei in der Spezifikation von Regeln (1) zur Identifikation von REST-Ressourcen, (2) zur Ableitung der Zustandsbeschreibung durch Dokumente sowie (3) zur Ableitung der softwaretechnischen Spezifikation der RESTful SOA.

Die Schritte der Problemlösung Analyse der Problemstellung (Kapitel 2: Stand der Literatur), Konzeption (Kapitel 3: Vorstellung des modellbasierten Vorgehens) und Validierung (Kapitel 4: Vorstellung der Fallstudie und Kapitel 5: Diskussion der Ergebnisse) spiegeln sich in der Gliederung dieser Arbeit wieder. Im Methodenspektrum der Wirtschaftsinformatik [9] lässt sich die Arbeit als qualitativ-konstruktiv, im Speziellen als konzeptionell-/argumentativ-deduktive Analyse, klassifizieren.

2 Stand der Literatur

REST wurde als Architekturstil für verteilte Systeme in der Dissertation von Roy Fielding [1] vorgestellt. Darin wird u. a. spezifiziert, wie existierende Web-Protokolle verwendet werden können, um Web-Services zu realisieren [3]. Der Architekturstil REST ist dabei durch folgende vier Prinzipien charakterisiert: (1) Als zentrale Komponenten werden Objekttypen und ihre Funktionalität als Ressourcen mit eindeutiger Identifikation (URI) bereitgestellt. (2) Aus Außensicht weisen diese Ressourcen eine einheitliche Schnittstelle in Form der http-Operatoren GET (lesen), PUT (aktualisieren), DELETE (löschen) und POST (erzeugen) auf. (3) Für die Darstellung ihrer Zustände kann eine Ressource in mehreren Repräsentationen vorliegen. Diese werden häufig in Dokumentformaten wie XML, HTML, CSV oder JSON verwaltet. (4) Beziehungen zwischen Ressourcen werden über das Konzept der Links (Verknüpfungen) realisiert. Im Rahmen von Serviceaufrufen können diese u.a. auch zur Steuerung der Anwendungslogik genutzt werden. Im Zusammenhang mit der Anwendung der beschriebenen Prinzipien steht auch die Forderung nach zustandsloser Kommunikation. REST sieht für Anfragen an Ressourcen vor, die für die Verarbeitung benötigten Zustandsinformationen als Parameter im Kontext der Anfrage mit zu übergeben.

Eng verbunden mit dem Architekturstil REST ist das Konzept der ressourcenorientierten Architektur (ROA). Overdick [10] charakterisiert eine ROA als SOA, welche den REST-Prinzipien genügt, also (all) ihre Entitäten (nicht nur Web-Services) als Ressource expliziert und somit global identifizier- und ansprechbar macht. Die Forderung der einheitlichen Schnittstelle ist eine weitere Einschränkung der klassischen SOA. Auch die im vorliegenden Beitrag entwickelte RESTful SOA ist demzufolge als ressourcenorientierte Architektur klassifizierbar und folgt mit dieser Einordnung auch dem Verständnis von Richardson und Ruby [4].

Mit der Modellierung und modellbasierten Entwicklung der Komponenten einer RESTful SOA beschäftigt sich eine Reihe von Arbeiten. Die dort beschriebenen Ansätze werden anhand folgender Kriterien analysiert: (1) Erfolgt die Identifikation von Ressourcen bzw. Dokumenten modellbasiert? (2) Auf welchen Modellebenen wird die RESTful SOA spezifiziert? (3) Erfolgt der Übergang zwischen den Modellebenen modellbasiert?

Den Ausgangspunkt für die Identifikation von Ressourcen bilden zum einen natürlich-sprachliche Beschreibungen, wie fachliche Anforderungen [3-4] oder Beschreibungen von Web-Service-Funktionalitäten [11]. Zum anderen erfolgt eine Ressourcenidentifikation auf Basis von Modellen durch die Analyse von Aktionen in Workflowschemata [12], [5] oder der funktionalen Beschreibung von Interaktionen in

einem Schema [13]. Dieses Vorgehen setzt damit bereits spezifizierte Client-Server-Interaktionen auf der Aufgabenträgerebene voraus. Die Ableitung selbst wird hier natürlich-sprachig erläutert.

Dokumente werden in den untersuchten Beiträgen primär unter dem Gesichtspunkt der Ableitung von Repräsentationsformaten diskutiert. Sie werden auf Basis von Ressourcenspezifikationen der REST-Anwendung (meist als Klassendiagramm) entwickelt [11-14]. Auch werden alternative Repräsentationsformate und Best Practices für die Dokumentenbeschreibung vorgestellt und diskutiert [3-5].

Die Spezifikation einer RESTful SOA beginnt in bestehenden Ansätzen meist auf der softwaretechnischen Ebene. Einerseits werden hierfür eigene Metamodelle für die Ressourcenbeschreibung entwickelt [15]. Andererseits erfolgen Struktur- bzw. Verhaltensspezifikationen der ROA mit (erweiterten) UML-Diagrammen [3], [11], [13-14]. Der Übergang auf und die Beschreibung der Implementierungsebene erfolgen durch die Ableitung von Schnittstellenbeschreibungen von Ressourcen unter Verwendung der Web Application Description Language (WADL) [11], [13] oder eigener Komponentenspezifikationen [14]. Zudem werden für die detaillierte Spezifikation von Softwarearchitektur und Implementierung der RESTful SOA eine Vielzahl an Hinweisen, Pattern und Best Practices bereitgestellt [3-4]. Die Überwindung der Lücke zwischen den Ebenen der Softwarearchitektur und der Implementierung erfolgt in den betrachteten Ansätzen nur teilweise regelbasiert.

Bezugnehmend auf die drei anfangs formulierten Kriterien ist zusammenfassend festzuhalten, dass in den betrachteten Arbeiten (1) die Identifikation von Ressourcen nicht auf Basis von Geschäftsprozessmodellen auf der Aufgabenebene erfolgt. Die Ressourcenidentifikation erfolgt durch Analyse natürlich-sprachiger Beschreibungen oder Workflow-/Interaktionsschemata. (2) Die Untersuchung der Entwicklung von RESTful SOA konzentriert sich primär auf die Spezifikation der softwaretechnischen Ebene. (3) Die meisten Ansätze beschreiben zudem einen modellbasierten Übergang hin zur Implementierung. Eine Betrachtung der Aufgabenebene in Form von Geschäftsprozessen findet nicht statt. Auch wird die Beziehung zwischen Aufgaben, die mit der SOA automatisiert werden sollen, und den korrespondierenden Komponenten der RESTful SOA unzureichend betrachtet. Die aufgezeigte semantische Lücke zwischen den Abstraktionsebenen der Geschäftsprozesse und der softwaretechnischen Spezifikation soll mit der vorliegenden Arbeit verringert werden.

3 Modellbasierte Spezifikation einer dokumentenorientierten RESTful SOA

Zielsetzung der vorliegenden Arbeit ist die modellbasierte Ableitung einer softwaretechnischen Spezifikation zur Automatisierung von Geschäftsprozessen durch eine dokumentenorientierte RESTful SOA. Als methodische Grundlage wird das Semantische Objektmodell (SOM) nach Ferstl und Sinz [8] gewählt. Die SOM-Methodik beinhaltet einen objekt- und geschäftsprozessorientierten Modellierungsansatz, der einen Geschäftsprozess sowohl auf Aufgaben- als auch auf Aufgabenträgerebene struktur- und verhaltensorientiert beschreibt. Die SOM-Methodik bildet das modell-

basierte Fundament der vorliegenden Arbeit, weil durch die explizite Spezifikation der strukturorientierten Sicht auf unterschiedlichen Ebenen ein wichtiges Instrument zur Identifikation und Spezifikation von Artefakten (Dokumente, Ressourcen) bzgl. der vorliegenden Zielsetzung bereitgestellt wird. Das in diesem Abschnitt nun detailliert vorgestellte Vorgehen kann zudem an eine bestehende modellbasierte Methodik anknüpfen und diese um eine plattform spezifische Ebene erweitern. Abbildung 1 zeigt die Modellarchitektur einschließlich der Erweiterung zur softwaretechnischen Beschreibung einer dokumentenorientierten RESTful SOA:

Geschäftsprozessmodellebene E1: Die SOM-Methodik sieht auf Aufgabenebene die Spezifikation von Geschäftsprozessen anhand einer Struktur- (Interaktionsschema, IAS) und einer Verhaltenssicht (Vorgangs-Ereignis-Schema, VES) vor. Die Koordination betrieblicher Objekte mittels Transaktionen wird im IAS erfasst. Das VES beschreibt die Durchführung der korrespondierenden Aufgaben als Ablaufdiagramm.

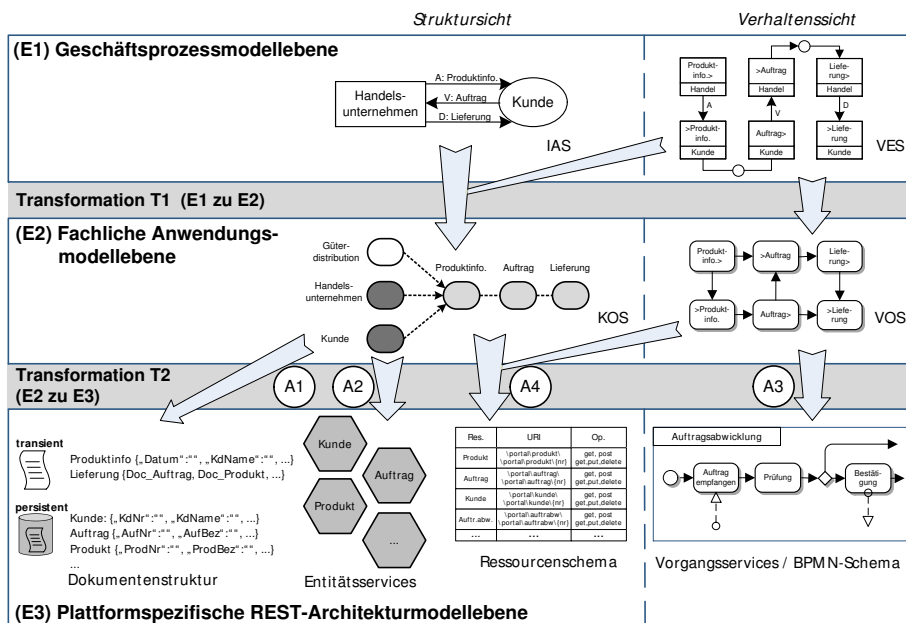


Abb. 1. Erweiterte Modellarchitektur der SOM-Methodik

Fachliche Anwendungsmodellebene E2: Auf Basis der Geschäftsprozessspezifikation wird ein plattformunabhängiges objektorientiertes Modell des Anwendungssystems (AwS), ebenfalls bestehend aus Struktur- (konzeptuelles Objektschema, KOS) und Verhaltenssicht (Vorgangsobjektschema, VOS), modellbasiert abgeleitet. Das KOS differenziert seine Objekttypen nach ihrem Ursprung im Geschäftsprozessmodell als leistungs-, objekt- oder transaktionsspezifisch. Für die Erläuterung der Schritte von Transformation T1 (E1 zu E2) sowie die zugehörige Schemakonsolidierung sei auf die Ausführungen in [8], [16] verwiesen.

REST-Architekturmodellebene E3: Die plattformspezifische Aufgabenträgerebene beschreibt die Erweiterung der SOM-Architektur zur Spezifikation einer dokumentenorientierten RESTful SOA durch Vorgangs- und Entitätsservices, die notwendigen transienten und persistenten Dokumente sowie die Schnittstellenspezifikation der Ressourcen. Die Transformation T2 der Objekttypen aus KOS und VOS (auf E2) in die Modellbausteine der Ebene E3 wird modellbasiert durch vier Ableitungsschritte (A1: Identifikation und Spezifikation von Dokumenten, A2: Identifikation und Spezifikation von Entitätsservices, A3: Identifikation und Spezifikation von Vorgangsservices, A4: Spezifikation des Ressourcenschemas) regelbasiert beschrieben. Die Modellebene (E3) und die Spezifikation der Transformation T2 stellen die Erweiterung der SOM-Methodik dar. Grundlage hierzu bildet die Arbeit von Wolf [16]. Das dort vorgestellte, fachlich orientierte Vorgehen wird um softwaretechnische Artefakte erweitert und hinsichtlich der Implementierung detailliert. Die Ableitungsschritte und zugehörigen Regeln werden nachfolgend zunächst allgemein eingeführt und im anschließenden Kapitel 4 anhand einer Fallstudie verdeutlicht:

Ableitungsschritt A1: Identifikation und Spezifikation von persistenten und transienten Dokumenten. Dokumente stellen den zentralen Entwicklungsgegenstand der vorliegenden Arbeit dar. In einer RESTful SOA werden diese einerseits transient zum Austausch zwischen Ressourcen (Vorgangs- und Entitätsservices), andererseits zur persistenten Verwaltung des Ressourcenzustands verwendet. In beiden Fällen werden die Dokumente über das Repräsentationsformat JSON beschrieben. Dieses eignet sich auch für die dokumentorientierte persistente Speicherung. In einem ersten Schritt werden Dokumente auf Grundlage des plattformunabhängigen KOS identifiziert und deren Struktur (Dokumenttyp) festgelegt. Hierzu wird *dokumentenorientiert konsolidiert*. Die Bildung von persistenten Dokumenten orientiert sich dabei an dem Grundsatz nach [17] (S. 13-24), Daten auf Grundlage ihrer gemeinsamen Nutzung zu Dokumenten zu aggregieren. Daten, die in einer ACID-Transaktion gemeinsam behandelt werden, sind demzufolge in einem gemeinsamen Dokument abzubilden. Zur weiteren Klassifikation erfolgt die Unterscheidung von persistenten Dokumenten nach Stamm- (A1a) bzw. Transaktionsdaten (A1b):

(A1a) Stammdatendokumente leiten sich aus den linksstehenden Objekttypen des KOS ab. Als von einem konkreten Geschäftsvorfall existenzunabhängige Objekttypen treten diese als objektspezifische oder leistungsspezifische Objekttypen auf. Damit werden Daten der am Geschäftsprozess beteiligten betrieblichen Objekte bzw. der Leistungsspezifikation beschrieben. Diesen werden Attribute zugeordnet und festgelegt, ob sie dauerhaft zu speichern sind. Je dauerhaft zu speichernden Objekttyp wird ein korrespondierender Stammdatendokumenttyp abgeleitet.

(A1b) Transaktionsdokumente bilden einen konkreten Geschäftsvorfall (Leistungserstellung und deren Koordination) ab. Zur Ableitung der Dokumenttypen werden alle existenzabhängigen transaktionsspezifischen Objekttypen des KOS betrachtet, ihre Dokumentenstruktur spezifiziert sowie die Persistenzeigenschaft (transient = t, persistent = p) festgelegt. Diese Objekttypen werden als Nachrichten interpretiert, die während der Durchführung des Geschäftsprozesses zur Koordination der Leistungserstellung zwischen den beteiligten betrieblichen Objekten ausgetauscht werden. Alle als persistent markierten Objekttypen dienen zudem der Identifikation persistent zu spei-

chernder Dokumente. Ein persistentes Transaktionsdokument stellt die Aggregation ausgetauschter Daten zur Koordination einer Leistungserstellung (D-Transaktion des Geschäftsprozessmodells auf detaillierter Ebene) dar. Das Schreiben und das Lesen von Transaktionsdaten erfolgt somit ausschließlich basierend auf einem Dokument. Ein Beispiel für unterschiedliche Transaktionsdokumenttypen innerhalb eines Geschäftsvorfalles stellen die Koordination und Durchführung der eigentlichen Leistungserstellung und die Abwicklung der Abrechnung dar.

Zur Abbildung des Ressourcenzustands wird jedem Transaktionsdokument ein Attribut „_STATUS“ zugewiesen und zulässige Statusausprägungen aus dem initialen KOS sowie der Ablaufspezifikation auf Aufgabenträgerebene abgeleitet. Die Ableitungsregeln der Stamm- und Transaktionsdatendokumente fasst Tabelle 1 zusammen.

Tabelle 1. Ableitungsregeln der dokumentenorientierten Konsolidierung des KOS

KOS	Dokumenttyp	p/t
Objektspezifischer Objekttyp	Stammdatendokument	p
Leistungsspezifischer Objekttyp	Stammdatendokument	p
Transaktionsspezifischer Objekttypen	Aggregation zu Transaktionsdatendokument (differenziert nach Status)	p/t

Ableitungsschritt A2: Identifikation und Spezifikation von Services zur Verwaltung der Dokumente. Es werden Services identifiziert, die der Verwaltung, also Speicherung von und Zugriff auf die Repräsentation persistenter Dokumente dienen. Diese werden im Folgenden als Entitätsservices bezeichnet und als Ressourcen interpretiert. Für jeden in Schritt A1 identifizierten Dokumenttyp wird zu diesem Zweck ein Service spezifiziert. Aus Innensicht wird je zulässigem http-Verb ein Lösungsverfahren basierend auf Attributausprägungen der übergebenen Dokumente spezifiziert. Für Entitätsservices zur Verwaltung von Transaktionsdaten sind die Ausführungsstatus der Transaktion als Zustandsautomat zu berücksichtigen. Über das Attribut „_STATUS“ werden je nach Operation zu lesende bzw. schreibende Daten und zum jeweiligen Zeitpunkt gültige und zu prüfende Integritätsbedingungen festgelegt.

Ableitungsschritt A3: Identifikation und Spezifikation von Services zur Vorgangsteuerung. Vorgangsservices sind zur Realisierung der Geschäftslogik notwendig und können ebenfalls als Ressource nach dem REST-Verständnis interpretiert werden. Die Funktionalität wird aus Außensicht dementsprechend über eine REST-Schnittstelle angeboten und in Webclient-basierten Anwendungssystemen aus der Kommunikationsschicht heraus aufgerufen. Die Identifikation der Vorgangsservices erfolgt in Anlehnung an die Arbeit von Krücke und Sinz [18]. Vorgangsservices werden hier anhand von Message Exchange Patterns (MEP) [19] innerhalb des VOS abgegrenzt. Als relevant werden die Muster *In-Out* (Request-Response) und *In-Only* angesehen, über die die Kommunikation zwischen Servicenutzer und Serviceanbieter beschrieben wird. Aus Innensicht wird der identifizierte Service unter Verwendung einer ausführbaren Workflowsprache, im Fall der vorliegenden Arbeit mit BPMN (Business Process Model and Notation, [20]) beschrieben. Dabei werden die unter Schritt (A2) identifizierten Entitätsservices durch REST-Aufrufe orchestriert.

Ableitungsschritt (A4): Spezifikation des Ressourcenschemas. Das Ressourcenschema spezifiziert einerseits die Schnittstellen einer RESTful-SOA mit URI, Zugriffoperatoren und Dokumentenstruktur für jede abgeleitete Ressource und dient andererseits der Ergebnisdokumentation parallel zu den Schritten (A1) bis (A3). Als Ressourcen werden Entitäts- und Vorgangsservices unterschieden. Ihre Bereitstellung erfolgt dabei jeweils als Listen- und Individualressource. Listenressourcen bieten Zugriff auf alle Instanzen, Individualressourcen auf eine konkrete Instanz eines Ressourcentyps. Für den Zugriff auf eine Ressource sind, dem REST-Paradigma folgend, aus Außensicht die zulässigen http-Verben festzulegen. Standardmäßig werden als Operatoren für Ressourcenlisten GET und POST festgelegt. GET ermöglicht das Abfragen der Instanzen, POST das Anlegen neuer Instanzen. Dies gilt sowohl für das Anlegen neuer Entitäten als auch das Starten eines neuen Vorgangs. Die Abfrage und Manipulation ausgewählter Instanzen erfolgt über die Schnittstelle der Individualressource mit GET, PUT und DELETE. Abschließend ist eine URI je Ressource zu spezifizieren. Diese setzt sich aus der (optionalen) Angabe des Hosts als möglichen Einstiegspunkt und Wurzel der REST-Anwendung sowie den Namen der Services/Ressourcen zusammen. Die Angabe der {id} identifiziert eine Instanz. Über den Status lassen sich Ressourcen nach dem Ausführungsstand filtern. Die Dokumentenstruktur, die für Stamm- und Transaktionsdaten festgelegt wurde (A1), bildet die Grundlage für die Beschreibung der Dokumentenstruktur, die eine Ressource empfangen und senden kann. Tabelle 2 fasst die Ableitungsregeln für das Ressourcenschema zusammen.

Tabelle 2. Ableitungsregeln Ressourcenschema

Quelle	Ressource	URI	Operator	Dokument
Entitätsservice (Stammdaten)	Liste	\host\servicename\	get, post	JSON {...} Definition der Struktur nach (A1)
	Individual	\host\servicename\{id}\	get, put, delete	
Entitätsservice (Transaktions- daten)	Liste	\host\servicename\ ?status=transact.status	get, post	
	Individual	\host\servicename\{id}\ ?status=transact.status	get, put, delete	
Vorgangsservice	Liste	\host\servicename\	get, post	
	Individual	\host\servicename\{id}	get, put, delete	

4 Fallstudie

Anhand einer Fallstudie wird nachfolgend das in Kapitel 3 beschriebene Vorgehen zur Ableitung der softwaretechnischen Spezifikation einer dokumentenorientierten RESTful SOA veranschaulicht und die Anwendbarkeit nachgewiesen.

4.1 Einführung der Fallstudie

Gegenstand der Fallstudie ist der Geschäftsprozess einer Flugvermittlung. Die Dienstleistung wird über ein Online-Portal erbracht und beinhaltet das Vermitteln von Flügen einschließlich der Abfrage von Fluginformationen und der Reservierung.

Der Ablauf des Geschäftsprozesses lässt sich wie folgt grob skizzieren: Nachdem der Kunde eine Suchanfrage an das Flugportal gestellt und die Ergebnisse durchgesehen hat, kann er einen Flug auswählen. Wurden die Suchergebnisse noch aus dem lokalen Datenbestand bedient, erfolgt jetzt bei der Fluglinie die Abfrage aktueller und gesicherter Fluginformationen durch das Flugportal. Auf diesen Informationen aufbauend wird ein Angebot erstellt und dem Kunden übermittelt. Nach erfolgter Buchung des Angebots werden die Plätze bei der Fluglinie reserviert und der Kunde erhält abschließend eine Bestätigung sowie Buchungsunterlagen. Die Struktur des Geschäftsprozesses stellt Abbildung 2 als IAS und die dazu korrespondierende Verhaltenssicht als VES dar.

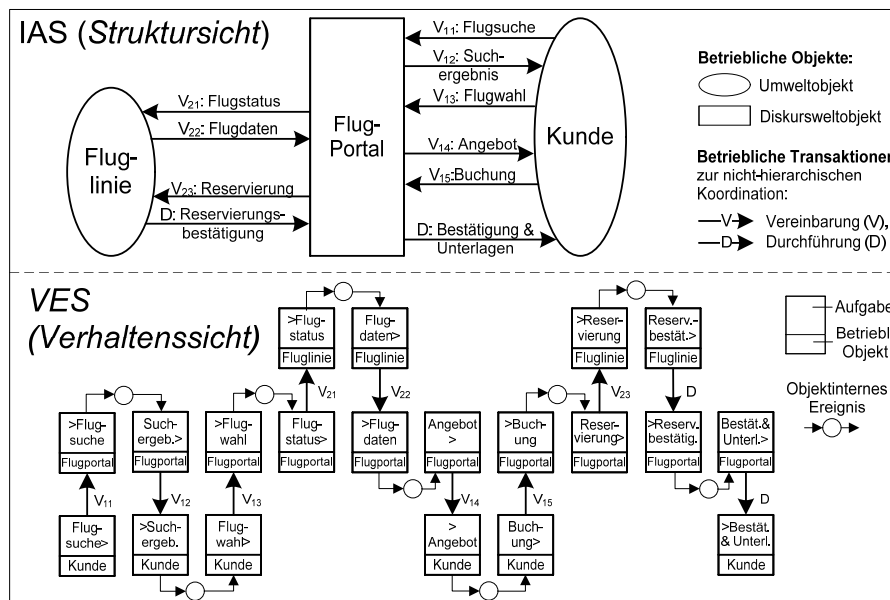


Abb. 2. Geschäftsprozessmodell der Fallstudie Flugvermittlung

Die Schritte Vereinbarung (V) und anschließende Durchführung (D) der Flugvermittlung werden in Form betrieblicher Transaktionen erfasst. Diese Transaktionen stellen Kommunikationsbeziehungen zwischen den beteiligten betrieblichen Objekten dar und weisen auf ausgetauschte Informationen hin. Über Transaktion V₁₂ wird beispielsweise eine Liste mit Flügen (Suchergebnis) von dem Flugportal an den Kunden übermittelt. Die daran beteiligten Aufgaben werden im VES erfasst und sind „Suchergebnis senden“ (Flugportal: Suchergeb. >) sowie „Suchergebnis empfangen“ (Kunde: > Suchergeb.).

Der modellierte Geschäftsprozess wird nun in eine fachliche AWS-Spezifikation transformiert, welche als Ausgangsbasis für die Identifikation und die Spezifikation von Dokumenten, Entitätsservices und Vorgangsservices verwendet wird.

Das IAS wird in ein initiales konzeptuelles Objektschema (KOS) transformiert. Hierbei werden die beteiligten betrieblichen Objekte (Kunde, Flugportal und Fluglinie) sowie die erbrachte Leistung (Flugvermittlung) links im KOS als existenzunabhängige Objekttypen angeordnet. Die Transaktionen werden entsprechend dem Ablauf ihrer Ausführung als transaktionsspezifische Objekttypen rechts davon angeordnet. Die Verhaltenssicht des Anwendungsmodells wird im Vorgangsobjektschema (VOS) dargestellt. Hierzu wird jede Aufgabe des VES in einen Vorgangsobjekttyp transformiert und die Schnittstellen zu den aufrufenden oder aufgerufenen betrieblichen Objekten bestimmt. In der Fallstudie werden die Aufgaben des Flugportals automatisiert, weshalb nur dessen Vorgangsobjekttypen im Folgenden zu betrachten sind. Die Ergebnisse der initialen Ableitung von KOS und VOS zeigt Abbildung 3.

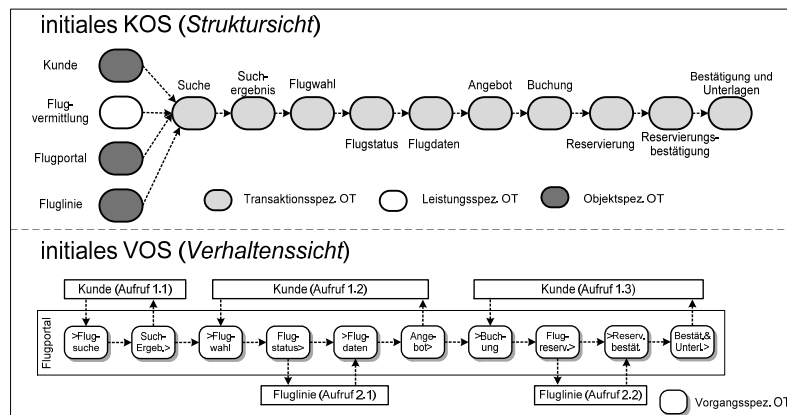


Abb. 3. Plattformunabhängige AWS-Spezifikation der Flugvermittlung

4.2 Identifikation und Spezifikation von Dokumenten (A1)

Ausgehend vom initialen konzeptuellen Objektschema (Abb.3) wird die dokumentenorientierte Konsolidierung mit den Schritten (A1a) *Ableitung von Stammdatendokumenten* sowie (A1b) *Ableitung von Transaktionsdatendokumenten* vorgenommen:

(A1a) In Tabelle 3 wird das Ergebnis der Konsolidierung der Stammdaten dargestellt. Kunde und Fluglinie beschreiben Geschäftspartner, die im Rahmen der Durchführung des Geschäftsprozesses kontaktiert werden und über das Flugportal Leistungen anbieten bzw. abrufen. Das Flugportal an sich dient der Vermittlung von Flügen, die lokal verwaltet werden. Insofern besteht für das Flugportal die Notwendigkeit zur Spezifikation und Speicherung eines Dokumententyps „Flug“. Der Objekttyp „Flugvermittlung“ ergibt sich aus der eigentlichen Leistung, die das Flugportal gegenüber dem Kunden erbringt. Diese Leistung lässt sich anhand verschiedener Bestandteile, wie

Kostensätze, AGBs usw. näher beschreiben. Zur persistenten Verwaltung dieser Leistungsspezifikation wird ebenfalls ein eigener Dokumenttyp erzeugt.

Tabelle 3. Ergebnis der Stammdatenkonsolidierung

Objekttyp	Persistent?	Dokumenttyp	Dokumentformat
Kunde	Ja	Kunde	Kundennummer, Name, Anschrift, usw.
Flugportal	Ja	Flug	FlugNr., Start-/Zielzeit, Start-/Ziel-flughafen, Plätze usw.
Fluglinie	Ja	Fluglinie	LinienNr., Bezeichnung, Anschrift, Kon-takt-daten, usw.
Flugver-mittlung	Ja	AGB	Rahmenbedingungen zur Flugvermitt-lung: Kosten, Rücktritt, usw.

(A1b) Zur Identifikation von Transaktionsdatendokumenten werden die Persistenzeigenschaften (transient = t, persistent = p) der transaktionsspezifischen Objekttypen festgelegt (vgl. Tabelle 4). Die dabei als persistent markierten Dokumenttypen werden zur Bildung des persistenten Transaktionsdatendokumenttyps „FlugBuchung“ aggregiert. Als Struktur ergibt sich somit das folgende Dokument-format: Transaktionsnummer (TNr), Status, Datum, Kundendaten (siehe Dokument-
typ „Kunde“), Flugdaten (siehe Dokumenttyp „Flug“). Die Überprüfung der Daten-konsistenz obliegt den Entitätsservices, die nachfolgend in Abschnitt 4.3 vorgestellt werden.

Tabelle 4. Analyse transaktionsspezifischer Objekttypen

Objekttyp	Dokumentstruktur	p/t
Suche	Datum, Uhrzeit, Start-/Zielflughafen	t
Suchergebnis	Liste mit Flügen (siehe Dokumenttyp Flug)	t
Flugwahl	TNr, Flug, _Status = „FLUG-AUSGEWAHLT“	p
Flugstatus	Flugnummer	t
Flugdaten	TNr, Flug, _Status = „FLUGSTATUS-GEPRUEFT“	p
Angebot	TNr, Flug, _Status = „FLUG-ANGEBOTEN“, Datum	p
Buchung	TNr, Flug, _Status = „FLUG-GEBUCHT“, Datum, Kunde	p
Reservierung	TNr, Flug, _Status = „FLUG-RESERVIERT“, Datum, Kunde	p
Reservierungsbestä-tigung	TNr, Flug, _Status = „BUCHUNG-BESTAETIGT“, Datum, Kunde	p
Best./Unterlagen	Flug, Kunde	t

4.3 Identifikation und Spezifikation von Entitätsservices (A2)

Für jeden unter Schritt (A1) identifizierten Dokumenttyp wird ein Entitätsservice erzeugt und die zulässigen http-Operatoren festgelegt. Dienste zur Stammdatenver-waltung ergeben sich für die Dokumenttypen „Flug“ (GET, PUT, POST), „Kunde“ (GET, PUT, POST), „Fluglinie“ (GET) und „AGBs“ (GET). Die Zugriffsoperatoren ergeben sich durch Analyse der Ablaufspezifikation im VOS (Abb. 3) und dem Kon-text der Servicenutzung. So kann es während der Ausführung des Geschäftsprozesses

nur zum lesenden Zugriff auf die Dokumente „Fluglinie“ und „AGB“ kommen. Das Löschen von Stammdaten im Rahmen des Geschäftsprozesses ist nicht vorgesehen.

Zur Verwaltung von Transaktionsdaten wird ebenfalls ein Entitätsservice angelegt. Für Dokumente des Typs „FlugBuchung“ ist lesender (GET), schreibender (POST, PUT) und löschender (DELETE) Zugriff zu definieren. Die zulässigen Status der Ressource zur Steuerung des Lösungsverfahrens sind in Tabelle 4 mit hinterlegt.

4.4 Identifikation und Spezifikation von Vorgangsservices (A3)

Auf Basis des initialen VOS aus Abbildung 3 werden die Vorgangsservices der RESTful SOA identifiziert (A3a) und ihre Innensicht spezifiziert (A3b).

Nach den Message Exchange Pattern „In-Out“ und „In-Only“ werden für die Fallstudie die Vorgangsservices „Suche“, „Angebot“ und „Buchung“ identifiziert. Über diese findet die vom Kunden initiierte Interaktion mit selbigem statt. Die Vorgangstypen des VOS werden bei der Spezifikation der Innensicht detaillierter beschrieben. Ein Kunde nutzt z. B. den Buchungsservice durch das Starten einer neuen Vorgangsserviceinstanz mit POST auf „\flugportal\angebot“ und Übergabe der korrespondierenden Dokumentdaten „FlugBuchung“.

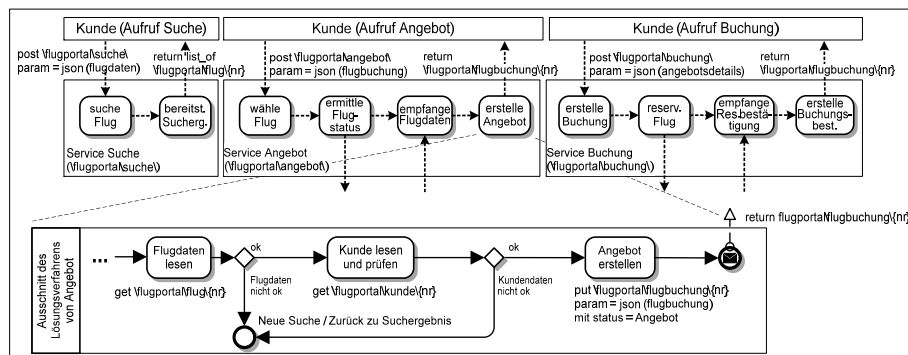


Abb. 4. Schema der Vorgangsservices und Ausschnitt des Lösungsverfahrens (BPMN)

Für das Beispiel wird ein Ausschnitt der Innensicht (Lösungsverfahren) zum Vorgangsservice „Angebot“ betrachtet (vgl. Abbildung 4). Die Funktionalität zur Erstellung des Angebots für den Kunden beginnt mit dem Lesen der Flugdaten mittels GET auf die Ressource „\flugportal\flug\{id}“ sowie dem Lesen und Prüfen der Kundendetails (GET auf „\flugportal\kunde\{nr}“). Zur Erstellung des Angebots wird die korrespondierende Flugbuchung durch PUT und Übergabe eines Dokuments mit den erstellten Angebotsdaten aktualisiert. Als Antwort auf seine Anfrage erhält der Kunde die URI auf das neu erstellte Angebot.

4.5 Spezifikation des Ressourcenschemas (A4)

Für die Fallstudie ergibt sich das in Tabelle 5 dargestellte Ressourcenschema. Als Host wird „flugportal“ festgelegt. Die Entitätsservices der Stammdaten „Flug“, „Kunde“, „Fluglinie“ (betriebliche Objekte) und „AGB“ (Leistung) werden jeweils als Listen- und Individualressource bereitgestellt. Der Zugriff auf die Fluglinie und die AGBs erfolgt dabei nur lesend (vgl. Abschnitt 4.3). Der Entitätsservice „FlugBuchung“ verwaltet die Flugbuchungstransaktion anhand ihrer Status. Entsprechend ist der Stand einer Flugbuchung nach Status geordnet zugreif- und manipulierbar. Die Vorgangsservices „Suche“, „Angebot“ und „Buchung“ werden ebenfalls als Listen- und Individualressourcen bereitgestellt.

Tabelle 5. Ressourcenschema der Fallstudie

Ressource	URI		Dokument
Flug	\flugportal\flug\ \flugportal\flug\{id}	get, post get, put, delete	JSON {...} vgl. Tab. 3
Kunde	\flugportal\kunde\ \flugportal\kunde\{id}	get, post get, put, delete	
Fluglinie	\flugportal\fluglinie\ \flugportal\fluglinie\{id}	get get	
AGBs	\flugportal\AGBs\ \flugportal\flugbuchung\ \flugportal\flugbuchung\{id}	get get, post get, put, delete	
FlugBuchung (je Status)	\flugportal\flugbuchung\?status=X \flugportal\flugbuchung\{id}\?status=X	get, post get, put, delete	
Suche	\flugportal\suche\ \flugportal\suche\{id}	get, post get, put, delete	
Angebot	\flugportal\angebot\ \flugportal\angebot\{id}	get, post get, put, delete	
Buchung	\flugportal\buchung\ \flugportal\buchung\{id}	get, post get, put, delete	

5 Zusammenfassung der Ergebnisse und Ausblick

In der Arbeit wird ein mehrstufiger modellbasierter Ansatz vorgeschlagen, der die schrittweise Ableitung der plattformspezifischen Beschreibung einer dokumentorientierten RESTful SOA ausgehend von SOM-Geschäftsprozessmodellen zum Gegenstand hat. Hierzu wird ein methodisch geleitetes Vorgehen zur Spezifikation von Vorgangs- und Entitätsservices vorgeschlagen, das das Dokument als Entwicklungsartefakt in den Mittelpunkt rückt. Beide Servicearten werden als Ressourcen im REST-Sinne interpretiert und sind über http-Verben und transiente JSON-Dokumente aufrufbar. Entitätsservices dienen der Verwaltung von Stamm- und Transaktionsdaten. Sie gehen aus persistenten Dokumenten hervor, die mittels dokumentenorientierter Konsolidierung des konzeptuellen Objektschemas abgeleitet werden. Die Konzepte des SOM-Geschäftsprozesses werden dabei über drei Abstraktionsebenen hinweg modellbasiert in dazu korrespondierende Bausteine der REST-Architektur überführt.

Das Geschäftsprozessmodell stellt somit nicht nur eine zur Implementierung konsistente Dokumentation dar, sondern auch den Ausgangspunkt des Entwicklungsprozesses. Das bereitgestellte Regelwerk unterstützt den Entwickler bei der Bewältigung Komplexität zur Überbrückung der semantischen Lücke zwischen Geschäftsprozess und softwaretechnischer Spezifikation. Dazu trägt nicht zuletzt die Trennung von struktur- und verhaltensorientierter Sicht auf den verschiedenen Modellebenen der SOM-Methodik bei. Über eine Fallstudie wird das Vorgehen näher erläutert und die Anwendbarkeit aufgezeigt. Die Skalierbarkeit des beschriebenen Ansatzes auf komplexere Anwendungsszenarien wird durch SOM als modellbasierter Geschäftsprozessmodellierungsansatz über Aufgaben- und Aufgabenträgerebene mit REST als Zielarchitektur der SOA gefördert. Gerade REST zielt als Architekturstil auf den Entwurf verteilter Architekturen für hoch-skalierbare Systeme, z. B. dem WWW, ab.

Anhand des beschriebenen Regelwerks wurde ein Prototyp am Fallbeispiel implementiert und erfolgreich gegen seine Anforderungen validiert. In diesem Rahmen wurden auch Technologien hinsichtlich der Unterstützung von REST und Dokumentenorientierung evaluiert. Vorgangsservices wurden mit BPMN auf der Plattform Activiti¹, Entitätsservices auf Basis von Java EE und dessen API JAX-RS² implementiert. Zur Realisierung der durchgängigen Dokumentenorientierung wurde das dokumentenorientierte Datenbanksystem Couchbase³ ausgewählt, das zur Datenverwaltung ebenfalls das Format JSON nutzt. Somit wird der Bruch zwischen Anwendungslogik und relationaler Datenhaltung vermieden.

Künftige Forschungsarbeiten lassen sich an den Eckpunkten der vorliegenden Arbeit identifizieren. Zur Erhöhung des Automatisierungsgrades der Entwicklung sollte der Ansatz hinsichtlich einer metamodellbasierten Integration der Modellebenen erweitert werden. Ausgehend von den Ergebnissen bzgl. persistenter Dokumente sind Dokumentdatenbanken als Basistechnologie sowie Einsatzmöglichkeiten und Architekturen von nicht-relationalen Datenbanken in Unternehmen, wie auch von McKinsey [21] motiviert, zu untersuchen.

Literatur

1. Fielding, R.T.: Architectural styles and the design of network-based software architectures. PhD thesis, University of California, Irvine (2000)
2. Pautasso, C., Leymann, F., Zimmermann, O.: Restful web services vs. "big" web services. In: Proceedings of the 17th international conference on World Wide Web - WWW '08, p. 805-814. ACM Press (2008)
3. Tilkov, S.: REST und HTTP. Einsatz der Architektur des Webs für Integrationsszenarien. dpunkt, Heidelberg, Neckar (2011)
4. Richardson, L., Ruby, S.: RESTful web services. O'Reilly, Farnham (2007)

¹ Activiti BPM Plattform (<http://activiti.org/>)

² JAX-RS (<http://jcp.org/en/jsr/detail?id=311>)

³ Couchbase (<http://www.couchbase.com/>)

5. Erl, T., Carlyle, B., Pautasso, C., Balasubramanian, R.: SOA with REST. Principles, patterns & constraints for building enterprise solutions with REST. Prentice Hall, Upper Saddle River, NJ (2013)
6. Sinz, E.J.: Konstruktionsforschung in der Wirtschaftsinformatik: Was sind die Erkenntnisziele gestaltungsorientierter Wirtschaftsinformatikforschung? In: Österle, H., Winter, R., Brenner, W. (eds.): Gestaltungsorientierte Wirtschaftsinformatik. Ein Plädoyer für Rigor und Relevanz. Infowerk, Nürnberg (2010)
7. Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Kremer, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J.: Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. Zeitschrift für betriebswirtschaftliche Forschung, 664–669 (2010)
8. Ferstl, O.K., Sinz, E.J.: Grundlagen der Wirtschaftsinformatik. Oldenbourg, München (2013)
9. Wilde, T., Hess, T.: Forschungsmethoden der Wirtschaftsinformatik. Wirtschaftsinformatik 49 (4), 280–287 (2007)
10. Overdick, H.: The Resource-Oriented Architecture. In: 2007 IEEE Congress on Services (Services 2007), pp. 340–347. IEEE (2007)
11. Porres, I., Rauf, I.: Modeling Behavioral RESTful Web Service Interfaces in UML. In: Chu, W., Wong, W.E., Palakal, M.J., Hung, C.-C. (eds.): Proceedings of the 2011 ACM Symposium on Applied Computing - SAC '11, pp. 1598–1605. ACM Press (2011)
12. Xu, X., Zhu, L., Liu, Y., Staples, M.: Resource-Oriented Architecture for Business Processes. In: 2008 15th Asia-Pacific Software Engineering Conf., pp. 395–402. IEEE (2008)
13. Laitkorpi, M., Selonen, P., Systa, T.: Towards a Model-Driven Process for Designing ReSTful Web Services. In: IEEE Int.'l Conf. on Web Services, pp. 173–180. IEEE (2009)
14. Pérez, S., Durao, F., Meliá, S., Dolog, P., Díaz, O.: RESTful, resource-oriented architectures: a model-driven approach. In: Chiu, D.K. (ed.): WISE 2010. LNCS, Vol. 6724, pp. 282–294. Springer, Berlin et al. (2011)
15. Schreier, S.: Modeling RESTful applications. In: Proceedings of the Second International Workshop on RESTful Design - WS-REST '11, p. 15-21. ACM Press (2011)
16. Wolf, M.: Modellbasierte Spezifikation von RESTful SOA auf Basis von Geschäftsprozessmodellen. In: Mattfeld, D.C., Robra-Bissantz, S. (eds.): Multikonferenz Wirtschaftsinformatik 2012, pp. 1649–1660. GITO, Berlin (2012)
17. Fowler, M., Sadalage, P.J.: NoSQL distilled. Addison-Wesley, Boston, Mass. (2012)
18. Krücke, A., Sinz, E.J.: Entwurf partieller SOA auf der Grundlage von Geschäftsprozessmodellen. In: Sinz, E.J., Bartmann, D., Bodendorf, F., Ferstl, O.K. (eds.): Dienstorientierte IT-Systeme für hochflexible Geschäftsprozesse. Univ. of Bamberg Press, Bamberg (2011)
19. W3C: Web Services Description Language (WSDL) Version 2.0 Part 2: Adjuncts, <http://www.w3.org/TR/wsd20-adjuncts/>
20. OMG: Business Process Model and Notation (BPMN). Version 2.0, <http://www.omg.org/spec/BPMN/2.0/PDF>
21. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A.: Big Data: The next frontier for innovation, competition, and productivity, http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

Using a Generic Model Query Approach to Allow for Process Model Compliance Checking – An Algorithmic Perspective

Sebastian Bräuer, Patrick Delfmann, Hanns-Alexander Dietrich,
and Matthias Steinhorst

University of Münster, ERCIS, Leonardo Campus 3, 48149 Münster, Germany
{braeuer,delfmann,dietrich,steinhorst}@ercis.uni-muenster.de

Abstract. Increased regulation forces financial companies to assure their business processes' compliance with legal and company-internal rules. In this paper, we introduce a model-driven business process compliance checking approach. It allows for defining compliance rules and identifying their occurrences in process models based on a graph theory-based approach. We outline the challenges to be met in the conceptualization of the approach and especially its implementation through suitable algorithms. Furthermore, we present an according modeling tool and evaluate the approach against related work.

Keywords: Business Process Management, Process Model Compliance Checking, Pattern Matching, GMQL

1 Introduction

In the wake of recent financial and economic crises, financial institutions are faced with a steady increase in regulations [1]. Against this backdrop, compliance checking has become of major interest in both Business Process Management research [2] and practice [3]. Compliance checking means determining if all business processes of a company comply with existing internal and external regulations [4]. It impacts process modeling, because regulations need to be represented in the process models of a company. This is commonly referred to as design time compliance checking [3].

In this context, a regulation can be understood as a restriction on the control flow of a process model. This restriction can be represented as a subsection or pattern of the overall model graph. For example, a simple, yet commonly occurring compliance rule states that a particular activity A must be preceded by an activity B. To apply for a loan, for instance, a customer's financial background first needs to be investigated. For a process model to comply with this rule, every instance of A (apply for loan) must consequently have a predecessor B (investigate financial background). To check whether a given set of process models does indeed comply with this rule, all corresponding pattern instances need to be found. A respective pattern can thereby either represent a compliance violation where A is *not* preceded by B or a compliant model

subsection where A is *indeed* preceded by B. Design time compliance checking thus corresponds to pattern matching in process models.

As many companies develop and maintain large repositories of process models [6], a manual pattern search is unfeasible. Implementing an automated solution, however, proves challenging for three reasons:

- First, a company may use different modeling languages to document its business processes. A compliance checking approach therefore should support different graph-based modeling languages.
- Second, as design time compliance checking equals the problem of graph pattern matching, a compliance checking approach must be able to identify arbitrarily complex structures within the overall model graph. In particular, compliance regulations often translate to (cyclic) paths of elements, which need to be found.
- Third, attribute values of model objects need to be compared to one another in the matching process. An attribute of a model object can for instance be its label or additional information about the object that need to be captured in the model. Consider the example of the 4-eye-principle which dictates that two particularly critical business activities need to be executed by two different employees or organizational units. Fig. 1 contains a violation of this rule in a BPMN-like process model, because activities “Check loan application” (A) and “Verify loan application” (B) are executed by the same organizational unit. In terms of a pattern query, this violation refers to a path from activity A to activity B such that the label attributes of the organizational units directly related to these activities carry the same value.

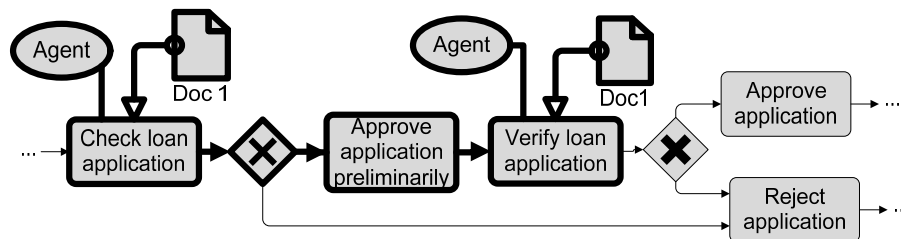


Fig. 1. Violation of the 4-eye-principle

To meet these challenges, a generic model query language (GMQL) has recently been proposed [7] and applied in the context of design-time compliance checking [8]. As our previous work focusses on conceptually specifying GMQL, the purpose of this paper is to provide an algorithmic specification of the matching process. We explain how GMQL is able to interpret variables and variable conditions. The paper also presents a working implementation of the algorithmic specification. As some of the compliance checking approaches have not been implemented, the paper contributes to the proliferation of applicable compliance checking approaches in research and practice.

As the development of a model query language falls into the realm of design science research, the remainder of this paper is structured according to the phases of a design science research process as outlined by PEFFERS ET AL. [9]. To define the objectives of our solution, we briefly introduce the concept of GMQL (Section 2). We

design and develop the solution by proposing an algorithmic specification of GMQL in Section 3. This specification is implemented and applied in the context of design time compliance checking in Section 4. This section also provides a discussion of the prerequisites and limitations of GMQL. We evaluate GMQL against the backdrop of existing literature in Section 5. The paper closes with a summary of its main findings and an outlook to future research in Section 6. Figure 2 contains a description of the research process. The figure lists each phase, the research method used to complete the phase, and the section of the paper at hand containing the respective findings.

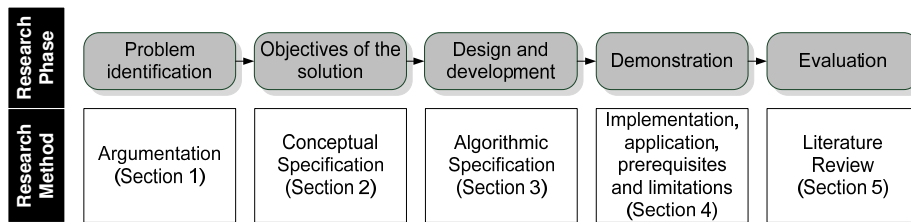


Fig. 2. Research methodology and structure of paper

2 Objectives of the Solution

GMQL recognizes any conceptual model as two basic sets. These are the set O of its objects and the set R of its relationships. The set E of elements is defined as the union of O and R . GMQL provides set-modifying functions and operators that perform operations on these basic sets. To be able to specify pattern queries, we define four classes of such functions. First, we have to be able to recognize elements of a particular type of the language and elements having attributes (of a particular type or value):

- *ElementsOfType*(X, a) returns all elements of the input set $X \in E$ that belong to a particular type a . The respective elements are put into a single output set.
- *ElementsWithTypeAttributeOfValue*(X, a, b) returns a set containing all elements of the input set of elements $X \in E$ that have an associated type attribute of name a with the attribute value b . A type attribute is assigned to all instances of a given object type. An example of a type attribute is the label of a BPMN activity. As all activities have a label the according attribute is assigned to the type specification.
- *InstanceAttributesOfValue*(Y, a, b) returns a set containing all instance attributes of the input object set $Y \in O$ whose attribute is of name a and whose values are equal to b . An instance attribute is assigned to an object instances. An example of an instance attribute is a clause of a particular law that needs to be adhered to in order to compliantly execute a process activity.
- *ElementsWithTypeAttributeOfDataType*(X, a) returns a set containing all elements of the set of elements $X \in E$ that are assigned type attributes of a particular data type. The data type of a type attribute can be STRING, INT, DOUBLE, BOOL, and ENUM. In case the second parameter is set to INT the functions will therefore return all elements that have type attributes of type INT. In case of a process activ-

ity, respective type attributes can for instance be execution time, costs, or number of employees involved to execute the activity.

- *InstanceAttributesOfDataType(Y,a)* returns a set containing all instance attributes of the object set $Y \in O$ whose data types equal the parameter a . Again, the above-mentioned data types are supported.

Furthermore, to examine neighborhood relationships, we need to identify all elements having (a particular number of) ingoing or outgoing relations (of a particular type):

- *ElementsWith{In|Out}Relations(X,Z)* return all elements of $X \in E$ and their {ingoing | outgoing} relationships defined in $Z \in R$. These functions each return a set of sets. Each inner set contains an element of X and all its relationships of Z .
- *ElementsWith{In|Out}RelationsOfType(X,Z,c)* return all elements of $X \in E$ and their {ingoing | outgoing} relationships of $Z \in R$ that are of type c . Again, these functions return a set of sets with each inner set containing one element of X as well as its {ingoing | outgoing} relationships of Z that belong to type c .
- *ElementsWithNumberOf{In|Out}Relations(X,Z,n)* return all elements of $X \in E$ that have a predefined number n of relationships of $Z \in R$. These functions return a set of sets with each inner set containing one element and its n relationships.
- *ElementsWithNumberOf{In|Out}RelationsOfType(X,Z,n,c)* are a combination of the two latter groups of functions. They return elements having a predefined number n of relationships of $Z \in R$ that are of type c . These functions return a set of sets.

In addition, we want to be able to find particular elements, their immediate neighbors, and the relationship(s) between them:

- *ElementsDirectlyRelated(X₁,X₂)* and *DirectSuccessors (X₁,X₂)* return all elements of $X_1 \in E$, their neighboring elements of $X_2 \in E$, as well as the relationships between the respective elements. *ElementsDirectlyRelated (X₁,X₂)* only works on undirected graph sections whereas *DirectSuccessors (X₁,X₂)* only works on directed graph sections.

Lastly, to be able to find structures representing element paths of arbitrary length, we included the following functions in the pattern matching mechanism:

- *{Directed}Paths(X₁,X_n)* return all {directed} paths between all elements of $X_1 \in E$ and all elements of $X_n \in E$. One inner set of the resulting set of sets contains one path from one element of X_1 to one element of X_n .
- *{Directed}Paths{Not}ContainingElements(X₁,X_n,X_c)* return all paths from elements of $X_1 \in E$ to all elements of $X_n \in E$ that contain at least one or no element of X_c .

For all paths-functions GMQL offers versions that determine only the *shortest* or *longest paths* as well as *loops*. As its theoretical basis is set theory, GMQL furthermore incorporates the set operators *Union*, *Intersect*, and *Complement* that perform the standard set operations on two sets of elements. Analogously to the *Intersect*- and *Complement*-Operators, the approach offers versions working on sets of sets (*InnerIntersect* and *InnerComplement*). The *Join*-operator unifies two inner input sets if

they have at least one element in common. The *SelfUnion* operator turns a set of sets into a single set, whereas a *SelfIntersect* operator performs an intersection on all inner sets resulting in one single set that holds all elements contained in every inner set.

These set-modifying functions and operators allow for constructing arbitrary pattern queries recursively, as the result set of one particular function/operator call serves as input for another function/operator (cf. Section 4.1).

In the context of design time compliance checking not only the graph structure needs to be analyzed but also information that is stored in object variables. Consider, again, the 4-eye-principle illustrated in Fig. 1. This rule requires the query language to treat values of particular attributes (in this case the labels of two organizational units) as variables that are to be compared to one another. Variables act as wildcards in the matching process and are of a predefined type. Available variable types are *ElementType*, *RelationshipType*, *Integer*, *AttributeDataType*, *AttributeName*, and *AttributeValue*. This implies that all functions that take instances of these data types as input can also take a variable of the respective type as input. The *ElementsOfType* function, for instance, is called with a set of elements and an element type specification. Instead of this type specification the function alternatively can be fed a variable of type *ElementType*. To define variable conditions only those variables can be used that have the same type. In other words, only variables of the same type can be compared to one another. Possible condition types are equal ($=$), unequal (\neq), smaller-than ($<$), greater-than ($>$), smaller-than-or-equal (\leq), and greater-than-or-equal (\geq).

3 Design and Development

A pattern query is represented in a tree structure. The matching algorithm is implemented using the visitor design pattern known from software engineering [10]. The visitor walks through the query expression in a bottom-up fashion calculating the leaf nodes first. In the following, this tree structure and the visitor calculating the node results will be referred to as the standard pattern compiler.

If a pattern query is run on a model, the matching algorithm distinguishes three cases. First, the query does not contain any variables or conditions. Second, the query contains variables but no conditions. Third, the query contains both variables and conditions. In case the query does not contain either variables or conditions, the standard pattern compiler is executed. If, however, the query contains either a variable and/or a condition a pre-compiler is executed that replaces the variable with the concrete values that can be found in the model the query is executed on. Each variable instantiation is then used to create a standard pattern query (without variables or conditions) that is fed to the standard pattern compiler. As this increases runtimes by the number of created pattern queries, a caching mechanism was implemented that allows for caching previously calculated (fragments of) pattern queries. The caching mechanism is implemented as a hash table allowing access to intermediary matching results in constant time.

For evaluating an arbitrary query q against a model m the function *CompilePatternQuery* (cf. Figure 3) can be used. As an output a set of sets S_r is calculated that

contains all pattern matches. As a first step, the query is analyzed and all occurring variables are stored in a set S_v . A variable thereby at least consists of a type and a label. In case that no variables are discovered within the query, no pre-compiling is necessary and the query is included in the set S_q that is directly handed over to the actual compiler for evaluating the query against the model.

```

PROCEDURE CompilePatternQuery(q, m)
  INPUT: Query q with potential variable expression and optional condition(s); model m to be analysed
  OUTPUT: Set  $S_r$  that contains all pattern matches

  Create a set  $S_v$  containing all variables in q
  IF  $S_v$  is empty THEN
    Add q to the set  $S_q$  containing all queries
  ELSE IF q contains at least one condition THEN
    Create set  $S_c$  holding all conditions of q
     $S_q \leftarrow \text{PrecompilePatternQueryWithConditions}(q, m, S_v, S_c)$ 
  ELSE q contains no conditions
     $S_q \leftarrow \text{PrecompilePatternQueryWithVariables}(q, m, S_v)$ 
  END IF
  Create  $S_r$  by executing all patterns in  $S_q$  on model m
END PROCEDURE CompilePatternQuery

```

Fig. 3. Pseudo-code of the function `CompilePatternQuery`

In case that the query includes at least one variable ($S_v \neq \emptyset$) the compiler analyses whether the query also includes any conditions. If this is the case, the set S_c is created, which contains all conditions that are included in q . A condition thereby consists of a left and right value, as well as an operator that is used for comparing the included values. With the query q , the model m , the set of variables S_v and the set of conditions S_c as parameters, the function *PrecompilePatternQueryWithConditions* is called (cf. Figure 4). The goal of this call is the creation of one or several queries that do not contain any variables or conditions but only those concrete value occurrences for which the conditions are fulfilled. In analogy to the conventional approach, these queries are then finally handed over to the standard pattern compiler as the set of sets S_q .

For generating the set of queries the algorithm requires the creation of an auxiliary empty set S_a . After that we use a loop over the set of variables S_v to determine the set of value occurrence S_{vo} per variable s_v and store S_{vo} in S_a . These execution steps are exemplarily depicted in Figure 5. a-c for variables of type Integer. The potential value occurrences for a variable s_v are identified by searching for all possible value occurrences within the whole model. In case of a variable of type Integer the algorithm for instance collects the number of incoming and outgoing relationships per object. For determining possible value occurrences of a variable of type *AttributeValue* the attribute values of instance attributes and type attributes are collected. For the other attributes the identification of value occurrences is performed accordingly.

```

PROCEDURE PrecompilePatternQueryWithConditions (q, m, Sv, Sc)
  INPUT: Query q with variable expression and condition(s);
        model m to be analysed; set Sv containing all variables;
        set Sc containing all conditions
  OUTPUT: Set of precompiled pattern queries Sq
  Create empty auxiliary set Sa
  FOREACH sv in Sv
    Get set of value occurrences Svo for sv in model m
    Add Svo to Sa
  END FOREACH
  Create a set of sets Sqvc that contains the Cartesian
  products of all Svo in Sa to obtain all possible value
  combinations
  FOREACH sqvc in Sqvc
    IF all conditions in Sc hold true on sqvc THEN
      Add set of value occurrences sqvc to a set of
      verified value combinations Svvc
    END IF
  END FOREACH
  FOREACH svvc in Svvc
    Create a clone q' of the original query
    Replace variables sv in q' with identified valid
    value occurrences from svvc
    Add q' to set of precompiled pattern queries Sq
  END FOREACH
END PROCEDURE PrecompilePatternQueryWithConditions

```

Fig. 4. Pseudo-code of the function PrecompilePatternQueryWithConditions

On the basis of the determined value occurrences that are stored per variable in S_a , the Cartesian product of all sets of value occurrences S_{v_o} in the auxiliary set S_a is generated for obtaining sets with all possible value combinations (cf. Figure 5 d). These sets are stored in the set $S_{q_{vc}}$. For determining those value combinations that fulfill the associated conditions, a loop over the set $S_{q_{vc}}$ is used (cf. Figure 5 e). Thereby, a set of value occurrences $s_{q_{vc}}$ is only written to the set of verified value combinations $S_{v_{vc}}$ if all conditions are fulfilled by the according values (cf. Figure 5 f). Finally, for each of these verified value combinations $s_{v_{vc}}$ a copy q' of the original query q is generated in which all variables s_v are replaced by their concrete value occurrences. This clone q' is then added to the set of precompiled queries S_q that is returned by the function.

<p>a) Derived sets:</p> <p>$S_v = \{A, B, C\}$ $S_c = \{\{A \neq B\}, \{C=4\}\}$</p>	<p>b) Value occurrences for each variable:</p> <p>$S_{vo} = \{1, 2, 3\} \mid s_v = A$ $S_{vo} = \{1\} \mid s_v = B$ $S_{vo} = \{4, 5\} \mid s_v = C$</p>	<p>c) Auxiliary set of value occurrences:</p> <p>$S_a = \{\{1, 2, 3\}, \{1\}, \{4, 5\}\}$</p>
<p>d) Creation of the Cartesian product:</p> <p>$S_{qvc} = \{\{1, 1, 4\}, \{1, 1, 5\}, \{2, 1, 4\}, \{2, 1, 5\}, \{3, 1, 4\}, \{3, 1, 5\}\}$</p>	<p>e) Checking of conditions:</p> <p>$A \neq B \wedge C=4$</p> <p>$1 \neq 1 \wedge 4=4$ ✗ $1 \neq 1 \wedge 5=4$ ✗ $2 \neq 1 \wedge 4=4$ ✓ $1 \neq 1 \wedge 5=4$ ✗ $3 \neq 1 \wedge 4=4$ ✓ $3 \neq 1 \wedge 5=4$ ✗</p>	<p>f) Set of validated value occurrences:</p> <p>$S_{vvc} = \{\{2, 1, 4\}, \{3, 1, 4\}\}$</p>

Fig. 5. Exemplary execution steps of the function `PrecompilePatternQueryWithConditions`

In all other cases a query is present that includes variables ($S_v \neq \emptyset$), but that does not have any associated conditions ($S_c \neq \emptyset$). Again, a pre-compiling of this query is required. In this case, we execute the function `PrecompilePatternQueryWithVariables` with the query q , the model m , and the set of variables S_v as input (cf. Figure 6).

```

PROCEDURE PrecompilePatternQueryWithVariables(q, m, Sv)
  INPUT: Query q with variable expression; model m to be
  analysed; Set Sv containing all variables
  OUTPUT: Set of precompiled pattern queries Sq
  Create temporary set of possible patterns Sq and add q to
  it
  Create an additional empty temporary set of possible que-
  ries Sq'
  FOREACH sv in Sv
    Get set of value occurrences Svo for variable sv in
    model m
    FOREACH svo in Svo
      FOREACH sq in Sq
        Replace sv in sp by svo and add sq to Sq'
      END FOREACH
    END FOREACH
  Sq = Sq'
  Clear Sq'
  END FOREACH
END PROCEDURE PrecompilePatternQueryWithVariables

```

Fig. 6. Pseudo-code of the function `PrecompilePatternQueryWithVariables`

As described above, a set of queries S_q is finally returned that includes all adjusted queries. For generating this set the algorithm first requires the creation of two temporary sets S_q and S_q' that are used for handling the original query q as well as the cleaned instances of this query where some or all of the variables have been replaced by concrete values. After that, we use a loop over the set of variables S_v to determine the set of value occurrences S_{vo} analogously to the approach with conditions. Two

embedded loops (the outer one over the set of value occurrences S_{vo} , the inner one over the temporary set of possible patterns S_q) are employed for replacing the variables within the queries of the set S_q . Thereby, for each inserted value occurrence an altered query is added to the other temporary set of possible queries S_q' . Inside the outermost loop the set S_q is assigned to the set S_q' that now contains the partly replaced queries (if it is the i^{th} iteration, the variable s_{vi} is replaced). The temporary set S_q' is then cleared and the loop starts over. Finally, the set S_q contains all pre-compiled patterns that do not include any more variables. This set is then fed to the standard pattern compiler.

4 Demonstration

4.1 Implementation and Application

The pattern query given below represents the violation of the 4-eye-principle outlined in the introductory section:

```
DirectedPaths (
  InnerIntersect (Join (
    DirectSuccessorsInclRelations (
      ElementsWithTypeAttributeOfValue (
        ElementsOfType (O, Document), Caption, A)
        ElementsOfType (O, Activity))
    ElementsDirectlyRelatedInclRelations (
      ElementsWithTypeAttributeOfValue (
        ElementsOfType (O, OrgaUnit), Caption, C)
        ElementsOfType (O, Activity)))
    ElementsOfType (O, Activity))
  InnerIntersect (Join (
    DirectSuccessorsInclRelations (
      ElementsWithTypeAttributeOfValue (
        ElementsOfType (O, Document), Caption, B)
        ElementsOfType (O, Activity))
    ElementsDirectlyRelatedInclRelations (
      ElementsWithTypeAttributeOfValue (
        ElementsOfType (O, OrgaUnit), Caption, D)
        ElementsOfType (O, Activity)))
    ElementsOfType (O, Activity)))
```

A=B; C=D

The query complies with the BPMN-like process modeling language depicted in Figure 1. It returns process paths of arbitrary length that start and end in an activity that is directly connected to both an organizational unit as well as a document. In addition to the label values of the organizational units being equal, the query further restricts the

result set to include only those paths whose start and end activities are directly related to document types that also have equal labels.

Both parameters of the *DirectedPaths* call are structured analogously. The join-construct returns all activities, their adjacent organizational units and documents, as well as the relationships between these objects (note that the *DirectSuccessors* call is necessary, because the relationship between an activity and a document is directed (cf. Figure 1)). This construct is then inner-intersected with the set of all activities to only feed activities to the *DirectedPaths* call. The set of organizational units and the set of documents that appear in both input parameters of the *DirectedPaths* call are further restricted to include only those objects that have captions of a particular value. These values take variables A to D as input that form two conditions specifying that all corresponding attributes need to have the same value.

This query is only an example of how GMQL can be used in the domain of design time compliance checking. Further examples are introduced in the following section as well as in [8]. In order to specify a query for a different modeling language the type information only needs to be adapted to that language (e.g. in an EPC an activity is called a function). GMQL is thus generic in the sense that it can be used for models created in arbitrary graph-based modeling languages.

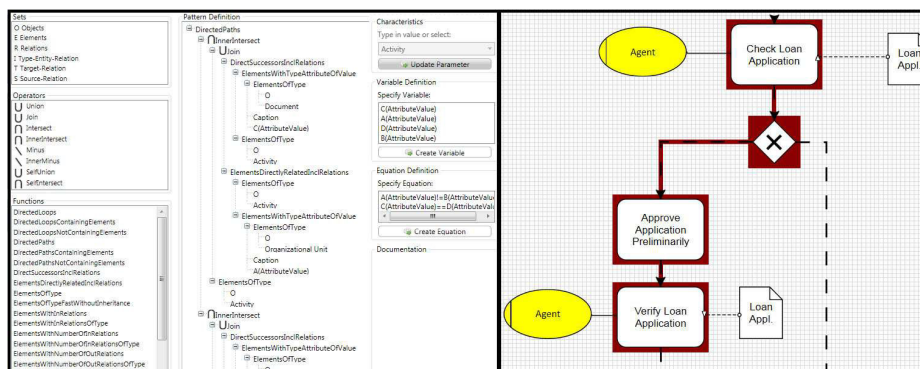


Fig. 7. Query specification environment (left) and model editor (right)

GMQL was implemented as a plugin for an existing meta-modeling tool that was available from a prior research project. Figure 7 depicts a screenshot of the meta-modeling tool's language editor that contains a query specification environment (left-hand part of Figure 7). The various functions and operators are available on the left-hand side. Variables and conditions can be defined on the right-hand side. These elements can be dragged and dropped to the pattern definition field where the query can be constructed. The right-hand part of Figure 7 contains an excerpt of the tool's modeling environment, which allows for selecting and running a previously defined pattern query on a model. All identified pattern occurrences are highlighted. Note that in this case a path is found that starts and end in an activity. Start and end objects are both directly related to an organizational unit and a document that have identical labels.

4.2 Prerequisites and Limitations

GMQL is able to identify structural patterns in process models of arbitrary modeling languages. However, design time compliance checking cannot solely be based on a structural analysis of the model graph. In addition, the semantics of element labels need to be considered as well. As indicated above, GMQL considers a label to be an attribute of an element. Respective information can thus be included in the matching process. Particular semantic issues like synonyms or homonyms, however, are not directly addressed by GMQL. However, GMQL allows for integrating approaches supporting terminological unambiguousness easily. Parts of the underlying algorithms comparing attribute labels are built generically. This means that, for example, it can be determined whether two nodes carry labels with the same *meaning* by checking their underlying ontological or linguistic concepts (e.g., “check invoice” ~ “bill audit”). Corresponding approaches maintaining the meaning of element labels can be found in [11] or [12] and can be used together with GMQL.

An additional prerequisite of using GMQL is an appropriate level of detail for a given model. In particular, the model needs to be syntactically correct and terminologically unambiguous. Element paths, for instance, can only be found if the model graph is not fragmented. The application of process modeling guidelines as proposed by [13] is thus advisable.

To apply GMQL in an enterprise, process managers have to learn GMQL in order to define the queries representing the appropriate compliance rules. Alternatively, an enterprise-wide role can be set up that is responsible for defining queries. In any way, we expect that this will require a considerable learning effort.

In terms of limitations, GMQL is limited to design time compliance checking in conceptual process models. In this context design time refers to an analysis of the model graph structure. Therefore, compliance violations that do not translate to specific model subsections cannot be found by our approach. A common compliance rule, for instance, states that particular business documents need to be stored for a given period of time. Such a rule cannot be checked in the model graph structure and can thus not be found by GMQL. In addition, compliance violations that occur during process runtime cannot be detected by GMQL. To determine if a given process instance contains a compliance violation respective mining techniques need to be applied that are beyond the scope of this paper. Such approaches are discussed in the area of business process intelligence [14].

Furthermore, GMQL does not consider the execution semantics of a process model. It analyses the graph structure and includes type and label information in its matching process. GMQL does not recognize the semantics of a given element type (e.g. an XOR split carries a different meaning than an AND join). In our experience, this, however, is not necessary in many cases. Consider, for example, the predecessor/successor compliance rule as depicted in Figure 8. Let the rule require that activity B needs to be performed after activity A has been executed. In case a) the path from A to C represents a violation of this rule, while in case b) it does not (because of the AND split/join). The GMQL query given in the right-hand part of Figure 8 identifies corresponding structures. It determines directed paths from A to C that do not contain

B. From the resulting set of sets, all paths that contain AND split nodes are subtracted. An AND split is a node having at least two outgoing edges. Thus, the set of all ANDs that have one outgoing edge is subtracted from the set of all ANDs resulting in the set of all ANDs that have at least two outgoing edges (i.e. AND splits). In case a), the query thus identifies the path from A to C as a compliance violation; while in case b) the empty set is returned (i.e. no violation is found). This argument demonstrates that including execution semantics in the matching process is not always necessary and analyzing the graph structure suffices to detect compliance violations.

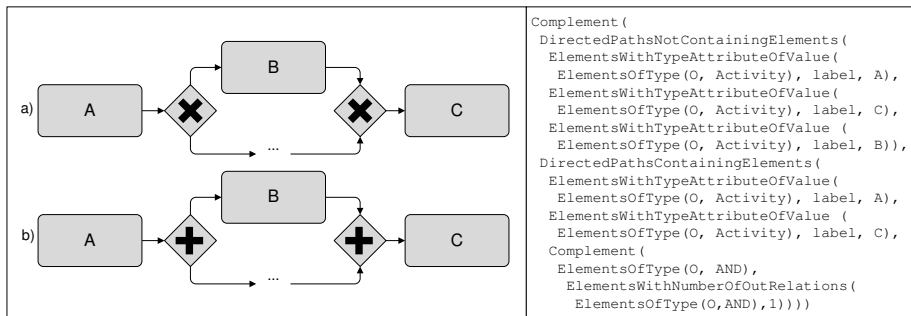


Fig. 8. Predecessor/successor compliance rule (left) and according GMQL query (right)

5 Evaluation

Design-time compliance checking requires an approach that is able to (a) identify arbitrarily complex structures within a model graph, (b) define pattern queries containing variables and conditions, and (c) support multiple modeling languages.

Many approaches discussed in the literature transform a process model into some form of finite-state automaton. This aims at checking the temporal logic of the task execution. The approach of GOVERNATORI ET AL. and the related research approaches [15-18] fall into this category. As they transform the models to a state-based event system, it is possible to map control flow restrictions to these states. This transformation, however, does not allow for finding more complex structures like highly-branched structures. This drawback is overcome in [19], as the transformation is a preprocessing step to an SQL-based query approach that is able to find more complex structures. Thus, element paths can be calculated. However, SQL suffers from performance problems when calculating paths, because node and edge tables need to be joined a potentially huge number of times [20]. In contrast, GMQL allows for calculating element paths without complex join operations. Furthermore, as GMQL provides specialized paths functions determining paths that (do not) contain particular elements a transformation to finite-state automata can be avoided which further increases runtime performance (see example above for more details). In addition, the approaches introduced above allow for defining compliance violations as formulae that are fed to the state automaton. These formulae require extensive knowledge about the used symbols and operators. GMQL queries in contrast are, albeit complex at

times, comparatively easy to understand, because the names of the functions and operators are self-explanatory.

Other approaches are restricted to particular modeling languages or language types [21-24]. The work of MONAKOVA ET AL. [25], for instance, transforms BPEL models to logical representations to allow for control- and data-flow analysis. The logical representation of a BPEL model stores the execution path and variables that allow for verifying compliance rules. WOLTER ET AL. [26-27] allow conditions to support different access control and authorization strategies in extended BPMN models. It is furthermore not possible to define arbitrary pattern queries. Moreover, OCL can be used to define and check compliance violations [28]. However, OCL is also restricted to UML. In contrast, GMQL can be used on models created in arbitrary graph-based modeling languages, because it is based on the idea that any model can be represented as the set of its objects and relationships.

Therefore, GMQL is similar to many approaches put forth in the area of general pattern matching in (process model) graphs. In graph theory, the problem of pattern matching is known as the problem of subgraph isomorphism (SGI). A plethora of algorithms has been developed in recent years that perform well in practical runtime scenarios despite the theoretical intractability of this problem (e.g., [29]). SGI, however, is concerned with finding one-to-one mappings between a given pattern graph and a subsection of a model graph. This requires knowing the exact structure of the pattern before searching it, which is unrealistic in many compliance checking scenarios, because, for instance, the length of a paths between particular elements is not always a priori known (see 4-eyes-principle above). The problem of finding similar patterns in a graph is known as subgraph homeomorphism (SGH). This problem, however, covers aspects of minor containment leading to an abundance of possible (mostly not suitable) pattern occurrences. These go too far for the purpose of pattern matching in process models [30]. Also, the theoretical runtime complexity of this problem is even worse than that of SGI [30]. Neither SGI nor SGH are able to define restrictions in the form of variables and conditions as it is possible with GMQL.

This literature analysis demonstrates that GMQL advances the current state of the art in design time compliance checking, because it is able to find arbitrarily complex patterns in models of any graph-based modeling languages. At the same time it allows for defining restrictions on attributes of particular model elements while avoiding computationally expensive model transformations to finite-state automata.

6 Summary and Outlook

We presented an algorithmic specification of a generic model query language. Besides the graph structure of a model, the language compiler is able to process attributes, variables, and variable conditions. Future research will continue along multiple consecutive lines. We plan to extend GMQL to allow for searching pattern occurrences that stretch across multiple models. This implies that GMQL must also allow for defining pattern queries that apply to models of more than one language. So far, a pattern query can only be executed on a model of the language the query was created

for. Relaxing this restriction will allow for verifying, if a hierarchy structure captured in an organizational chart is properly represented in a process model. Both extensions can be easily implemented by augmenting the set of elements that is initially fed into matching algorithm. We aim at gaining a deeper understanding of the compliance checking domain in order to determine if GMQL is indeed sufficient for this field. This may require including additional information in the matching process which so far has not been necessary (e.g. execution semantics). We will explore the possibility of compliance checking during modeling by enriching a process modeling language with GMQL-based concepts supporting the ex-ante specification of compliance rules.

References

1. Opromolla, G.: Facing the Financial Crisis: Bank of Italy's Implementing Regulation on Hedge Funds. *Journal of Investment Compliance* 10 (2), 41-44 (2009)
2. Abdullah, S.N., Indulska, M., Sadiq, S.: A Study of Compliance Management in Information Systems Research. In: *Proceedings of the 17th European Conference on Information Systems*, pp. 1-10 (2009)
3. Sadiq, S., Governatori, G., Namiri, K.: Modeling Control Objectives for Business Process Compliance. In: Alonso, G., Dadam, P., Rosemann, M. (eds.): *BPM 2007*. LNCS, Vol. 4714, pp. 149-164. Springer, Heidelberg (2007)
4. El Kharbili, M., De Medeiros, A.K.A, Stein, S., van der Aalst, W.M.P.: Business process compliance checking: Current state and future challenges. In: *MoBIS 2008*. LNI, Vol. 141, pp. 107-113. GI, Bonn (2008)
5. Cabanillas, C., Resinas, M., Ruiz-Cortés, A.: Hints on how to face business process compliance. *Ac-tas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos* 4 (4), 26-32 (2010)
6. Uba, R., Dumas, M., García-Bañuelos, L., La Rosa, M.: Clone Detection in Repositories of Business Process Models. In: Rinderle-Ma, S., Toumani, F., Wolf, K. (eds.): *Business Process Management*. LNCS, Vol. 6896, pp. 248–264. Springer, Heidelberg (2011)
7. Delfmann, P., Herwig, S., Lis, L., Stein, A., Tent, K., Becker, J.: Pattern Specification and Matching in Conceptual Models. A Generic Approach Based on Set Operations. *Enterprise Modelling and Information Systems Architectures* 5 (3), 24-43 (2010)
8. Becker, J., Bergener, P., Delfmann, P., Weiß, B.: Modeling and Checking Business Process Compliance Rules in the Financial Sector. In: *Proceedings of the 32nd International Conference on Information Systems* (2011)
9. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. *Journal of Management Information Systems* 24 (3), 45-77 (2007)
10. Gamma, E., Helm, R., Johnson, R. E.: *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Longman, Amsterdam (1995)
11. Thomas, O., Fellmann, M.: Semantic Process Modeling – Design and Implementation of an Ontology-based Representation of Business Processes. *Business and Information Systems Engineering* 1 (6), 438-451 (2009)
12. Delfmann, P., Herwig, S., Lis, L.: Unified Enterprise Knowledge Representation with Conceptual Models - Capturing Corporate Language in Naming Conventions. In: *Proceedings of the 30th International Conference on Information Systems* (2009)

13. Mendling, J., Reijers, H., van der Aalst, W.: Seven Process Modeling Guidelines. *Information and Software Technology* 52 (2), 127-136 (2010)
14. Jiafei, L., Jagadeesh, C., van der Aalst, W.: Mining Context-Dependent and Interactive Business Process Maps using Execution Patterns. In: *Proceedings of the 6th International Workshop on Business Process Intelligence* (2010)
15. Governatori, G., Milosevic, Z.: A formal analysis of a business contract language. *International Journal of Cooperative Information Systems* 15 (4), 659-685 (2006)
16. Hoffmann, J., Weber, I., Governatori, G.: On compliance checking for clausal constraints in annotated process models. *Information Systems Frontiers* 43, 1-23 (2009)
17. Lu, R., Sadiq, S., Governatori, G.: Compliance aware business process design. In: ter Hofstede, A., Benatallah, B., Paik, H.-Y. (eds.): *Business Process Management Workshops*. LNCS, Vol. 4928, pp. 120-131, Springer, Heidelberg (2008)
18. Lu, R., Sadiq, S., Governatori, G.: Measurement of Compliance Distance in Business Processes. *Information Systems Management* 25 (4), 344-355 (2008)
19. Awad, A., Weske M.: Visualization of Compliance Violation in Business Process Models. In: Awad, A., Weske, M. (eds.): *BPM 2009. LNBP 43*, pp. 182-193. Springer (2010)
20. Sakr, S.: Storing and Querying Graph Data Using Efficient Relational Processing Techniques. In: *Proceedings of the 3rd International United Information Systems Conference* (2009)
21. Liu, Y., Muller, S., Xu, K.: A static compliance-checking framework for business process models. *IBM Systems Journal* 46 (2), 335-361 (2007)
22. Knuplesch, D., Ly, L., Rinderle-Ma, S., Pfeifer, H., Dadam, P.: On Enabling Data-Aware Compliance Checking of Business Process Models. In: *Conceptual Modeling – ER 2010*. LNCS, Vol. 6412, pp. 332-346. Springer, Heidelberg (2010)
23. Trčka, N., van der Aalst, W., Sidorova, N.: Data-Flow Anti-patterns: Discovering Data-Flow Errors in Workflows. In: *Advanced Information Systems Engineering*. LNCS, Vol. 5565, pp. 425-439. Springer, Heidelberg (2009)
24. Kumar, A., Liu, R.: A Rule-Based Framework Using Role Patterns for Business Process Compliance. In: Bassiliades, N., Governatori, G., Paschke, A. (eds.): *Rule Representation, Interchange and Reasoning on the Web*. LNCS, Vol. 5321, pp. 58-72, Springer, Heidelberg (2008)
25. Monakova, G., Kopp, O., Leymann, F., Moser, S., Schäfers, K.: Verifying Business Rules Using an SMT Solver for BPEL Processes. In: *Proceedings of the Business Process and Services Computing Conference*, pp. 81-94 (2009)
26. Wolter, C., Miseldine, P., Meinel, C.: Verification of business process entailment constraints using SPIN. In: Massacci, F., Redwine, S.T., Zannone, N. (eds.): *Engineering Secure Software and Systems*, pp. 1-15 (2009)
27. Wolter, C., Meinel, C.: An approach to capture authorisation requirements in business processes. *Requirements Engineering* 15 (4), 359-373 (2010)
28. Warmer, J., Kleppe, A.: *Object Constraint Language 2.0*. mitp, Bonn 2004
29. Cordella, L.P., Foggia, P., Sansone, C., Vento, M.: A (sub)graph isomorphism algorithm for matching large graphs. *IEEE transactions on pattern analysis and machine intelligence* 26 (10), 1367-1372 (2004)
30. Lingas, A., Wahlen, M.: An exact algorithm for subgraph homeomorphism, *Journal of Discrete Algorithms* 7 (4), 464-468 (2009)

Nutzung des Viewpoint-Konzepts zur Unterstützung kollaborativer Modellierung – Konzeption und prototypische Implementierung

Julian Krumeich, Dirk Werth, und Peter Loos

Institut für Wirtschaftsinformatik (IWi),
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI),
Saarbrücken, Deutschland
{julian.krumeich, dirk.werth, peter.loos}@iwi.dfki.de

Abstract. In der Wirtschaftsinformatik werden Modelle zur Entwicklung von Informationssystemen eingesetzt. Trotz der inhärenten Komplexitätsreduzierung weisen diese häufig einen komplexen Charakter auf. Der wahrgenommene Nutzen durch den Modellnutzer ist daher vermindert. Dies wird intensiviert, wenn mehrere Subjekte aus unterschiedlichen Domänen bei der Modellerstellung und -nutzung im Sinne einer kollaborativen Modellierung involviert sind. Um die Modellierung hierbei zu unterstützen, gilt das Viewpoint-Konzept als erfolgsversprechend. Es ermöglicht eine Reduzierung der Komplexität von Modellen, indem diese in Stakeholder-spezifische Fragmente heruntergebrochen werden. Der vorliegende Beitrag konzipiert eine Viewpoint-basierte Modellierungsmethode, die das Viewpoint-Konzept mit der kollaborativen Modellierung verknüpft. Hierdurch wird den typischen Problemstellungen der kollaborativen Modellierung ebenfalls entgegengewirkt. Neben der Konzeption der Viewpoint-basierten Modellierungsmethode präsentiert der Beitrag überdies einen Proof-of-Concept anhand einer prototypischen Implementierung.

Keywords: Kollaborative Modellierung, Sichtenkonzept, Viewpoint-basierte Modellierung, konstruktionsorientierte Forschung

1 Einleitung

1.1 Motivation

Bei der Entwicklung von Systemen – im Kontext der Wirtschaftsinformatik typischerweise Informationssystemen – sind Modelle ein weitverbreitetes Mittel, um ein besseres Verständnis auf ein zugrundeliegendes Artefakt zu erhalten. Folgt man der allgemeinen Modelltheorie Stachowiaks [1], so sind Modelle „stets Modelle von etwas, nämlich Abbildungen, Repräsentationen natürlicher oder künstlicher Originale, die selbst wieder Modelle sein können.“

Auch wenn Modelle durch ihr inhärentes Verkürzungsmerkmal eine Komplexitätsreduzierung anstreben, so ist ihnen dennoch häufig – insbesondere bei der Entwick-

lung äußerst komplexer Systeme – ein weiterhin komplexer Charakter zuzusprechen. Die wahrgenommene Zweckdienlichkeit durch den Modellnutzer bleibt daher vermindert. Dies wird noch intensiviert, wenn mehrere Subjekte – die gegebenenfalls noch aus unterschiedlichen Domänen stammen – in die Modellerstellung und spätere -nutzung involviert sind. Bieten Modelle in einem solchen Umfeld lediglich eine einzelne, starre Perspektive auf das jeweilige Modell, so ist der Modellnutzen nur relativ gering für das einzelne beteiligte Subjekt erfahrbar [2].

Daher gilt es, die einzelnen Stakeholder, die bei der Modellerstellung als auch -nutzung beteiligt sind, stärker in den Fokus zu rücken, um damit den Nutzen der Modellierung zu erhöhen. Da die einzelnen Stakeholder in der Regel spezifische Fähigkeiten, Verantwortlichkeiten sowie Wissen und Expertise vorweisen [3], benötigen sie individuelle Perspektiven auf die Entwicklung des zugrundeliegenden Systems. Konsequenterweise sollte eine Betrachtung und Manipulation der Modelle und Modellfragmente nach den jeweils spezifischen Bedürfnissen ermöglicht werden [4].

1.2 Forschungsbeitrag

Um die Modellierung unter Berücksichtigung dieser Anforderungen zu unterstützen, gilt das Viewpoint-Konzept als erfolgversprechend. Dieses ermöglicht eine Reduzierung der Komplexität von Modellen, indem diese in Stakeholder-adäquate Fragmente heruntergebrochen werden [5]. Dadurch können die an der Modellierung beteiligten Akteure stärker in den Fokus des Modellierungsprozesses gerückt und so das Verständnis und die Produktivität gesteigert werden [3]. Gleichzeitig trägt das Viewpoint-Konzept zur Erstellung von besseren, das zugrundeliegende System adäquater abbildenden, Modellen bei. Dies konnte bereits durch Studien belegt werden [6].

Durch die Involvierung mehrerer Stakeholder in den Modellierungsprozess resultiert allerdings auch eine Komplexitätssteigerung durch die erschwerte Koordination des zugrundeliegenden Modellierungsprojektes. Bei einer solchen Konstellation wird von kollaborativer Modellierung gesprochen [7-8], die als ein „*process where a number of people actively contribute to the creation of a model*“ definiert wird [9]. Die Anwendung dieses Modellierungsprinzips ist jedoch nicht als trivial zu betrachten. [9] führte in diesem Kontext eine Expertenbefragung unter IT-Beratern durch, um herauszufinden, was die größten Probleme bei der kollaborativen Modellierung sind. Als Ergebnis ergab sich u. a. „low model acceptance“, „participants feel misunderstood“, „slow progress, facilitator overload“ und „limited model comprehension“.

Betrachtet man diese Probleme, so lässt sich feststellen, dass das Viewpoint-Konzept diese aber gerade zu lösen vermag und damit einen erfolgsversprechenden Beitrag zur kollaborativen Modellierung leisten könnte. Daher entwickelt der vorliegende Beitrag eine Viewpoint-basierte Modellierungsmethode, bei der das Viewpoint-Konzept genutzt wird, um eine effektive Einbeziehung mehrerer Stakeholder in den kollaborativen Modellierungsprozess zu erreichen. Überdies zeigt der Beitrag einen Proof-of-Concept des Konzeptes anhand einer prototypischen Implementierung.

1.3 Forschungsansatz und Aufbau des Beitrages

Dieser Beitrag folgt dem konstruktionsorientierten Forschungsansatz, bei dem ein betriebliches, d. h. prinzipiell praktisches Problem motivierend für die Entwicklung eines innovativen IT-Artefakts steht, mit dessen Hilfe eine Lösung der zugrundeliegenden Problemstellung erreicht werden soll. Diesem Forschungsansatz folgend entwickelt dieser Beitrag zunächst eine konzeptionelle Methode zur in Abschnitt 1.1 vorgestellten Ausgangsproblematik der betrieblichen Praxis. Darauf aufbauend wird überdies eine konkrete Instanz des Artefaktes in Form eines Prototyps vorgestellt.

Nach [10] sind bei der konstruktionsorientierten Forschung sieben Guidelines zu beachten, um eine stringente und problemrelevante Forschung durchzuführen. In diesem Sinne wird durch die Einleitung und Motivation die Relevanz des Forschungsproblems dargestellt (*Guideline 2*). Das in Abschnitt 3 entwickelte Konzept stellt ein IT-Artefakt (*Guideline 1*) dar, welches durch die prototypische Implementierung – wie sie in Abschnitt 4 vorgestellt wird – ein nachvollziehbares und prüffähiges Artefakt darstellt (*Guideline 4*). Dieses gilt es noch nach *Guideline 3* zu evaluieren (siehe hierzu Abschnitt 5). Um auf vorhandenem Wissen aufzubauen (*Guideline 6*), greift der in Abschnitt 3 entwickelte Ansatz auf eine zuvor durchgeführte Arbeit zurück, die wiederum auf weiteren existierenden Arbeiten aufbaut. Durch die Vorstellung der verfolgten Forschungsmethodik ist der *Guideline 5* ebenfalls Rechnung getragen. Letztlich wird durch die Einreichung dieses Beitrags die Erfüllung der letzten *Guideline 7* angestrebt, die eine Verbreitung des Konzepts vorsieht.

Der Aufbau des vorliegenden Beitrags ist wie folgt gegliedert: Nachdem in Abschnitt 1 die Motivation sowie der angewendete Forschungsansatz aufgezeigt wurden, werden in Abschnitt 2 verwandte Arbeiten im Bereich des Viewpoint-Konzepts und der kollaborativen Modellierung vorgestellt. Im Anschluss daran wird in Abschnitt 3 das Viewpoint-Konzept eingeführt und mit der kollaborativen Modellierung zur Viewpoint-basierten Modellierungsmethode verbunden. Diese wird im nachfolgenden Abschnitt 4 anhand einer prototypischen Implementierung instanziiert. Der Beitrag schließt in Abschnitt 5 mit einem Ausblick auf zukünftige Arbeiten.

2 Verwandte Arbeiten

Der Ursprung des Viewpoint-Konzepts geht auf die 1985 veröffentlichte Arbeit von [11] zurück [12]. Der dort entwickelte „MultiView“ Ansatz zielt auf die Unterstützung des Entwicklungsprozesses von computergestützten Informationssystemen ab, indem dieser komplexe Prozess in fünf unterschiedliche Perspektiven bzw. Viewpoints unterteilt wird. Ein weiterer Grundstein auf dem Gebiet der Viewpoint-Forschung stellt die in der Literatur häufig zitierte Arbeit von [3] dar. Während [11] mit ihrem Ansatz primär auf den Einsatz von Viewpoints in der Softwareentwicklung abzielen, ist der Ansatz von [3] potenziell bei der Entwicklung jedes Artefakts, welches nicht in einem trivialen Entwicklungsprozess konstruierbar ist, anwendbar.

Mit „IEEE 1471:2000“ existiert zudem eine Standarddefinition des Viewpoint-Konzepts, die auch von der ISO/IEC als „ISO/IEC 42010:2007“ Standard im Jahr 2007 angenommen wurde [13]. Demzufolge lässt sich das Viewpoint-Konzept bzw.

ein Viewpoint als „a specification of the conventions for constructing and using a view“ definieren. Abgesehen von der eigentlichen Definition des Viewpoint-Konzepts als auch der verschiedener Viewpoints in unterschiedlichen Anwendungsdomänen – hier ist bspw. im Bereich des Enterprise Architecture Managements das TOGAF-Rahmenwerk zu nennen [14] – sind auch Arbeiten zu erwähnen, die stärker auf die tatsächliche Nutzung des Viewpoint-Konzepts fokussieren.

Hier sei die Arbeit von [15] genannt. Diese schlagen ein methodisches Rahmenwerk vor, welches das Zusammenwirken zwischen einzelnen Viewpoints zu verbessern versucht. In diesem Kontext ist auch der „Hybrid Multi-View Modeling“ Ansatz von [16] aufzuführen. Dieser verfolgt das Ziel, eine (aus technischer Sicht) effiziente Synchronisation unterschiedlicher Views im Modellierungsprozess zu ermöglichen. Dies bedeutet, dass Änderungen, die in einer View eines Viewpoints durchgeführt wurden, mittels bestimmter Techniken erkannt und auf inkrementelle Weise an die weiteren existierenden Views propagiert werden. Hiermit wird das Ziel verfolgt, insbesondere bei größeren Modellen, keinen übermäßigen Overhead durch die Neuberechnung der durch die Modelländerung beeinflussten Views zu generieren.

Mit einem stärkeren Fokus auf die konkrete Umsetzung des Viewpoint-Konzepts in Modellierungswerkzeugen ist auf den erst kürzlich vorgestellten Beitrag von [5] zu verweisen. In diesem werden Anforderungen an die tatsächliche Umsetzung des Viewpoint-Konzepts bei der Entwicklung von Modellierungswerkzeugen untersucht. Auch eine durchgeführte Vorarbeit zu diesem Beitrag betrachtet solche werkzeugspezifischen Anforderungen an das Viewpoint-Konzept [2].

Zwar wird von manchen, in der Literatur zu findenden Ansätzen, die Zusammenführung verschiedener nach dem Viewpoint-Konzept erstellten Modelle erforscht, kollaborative Aspekte hinsichtlich der tatsächlichen Modellierung lassen sich dagegen in der Viewpoint-Forschung nicht erkennen.

In der Forschung zu kollaborativer Modellierung entwickelte [17] mit dem COMA-Werkzeug eine konkrete Werkzeugunterstützung zur Umsetzung der kollaborativen Modellierung. COMA steht für „Collaborative Modeling Architecture“, mit dessen Hilfe die Koordination von Gruppen bei der Modellierung von UML-Modellen erreicht werden soll. Ein weiterer Ansatz wird von [18] vorgeschlagen. Ihre „Fragment-Driven Process Modeling Methodology“ erlaubt die Entwicklung eines Prozessmodells durch mehrere Akteure auf kollaborative Weise. Hierzu modelliert jeder Akteur seine eigenen Aktivitäten, die sodann vom System zu einem Modell zusammengefügt werden. Allerdings wird hierbei lediglich ein Bottom-up Ansatz verfolgt, der synonym als Fragment-getriebener Ansatz bezeichnet wird. Folglich werden lediglich die einzelnen Prozessfragmente ablaforientiert zusammengesetzt. Eine kollaborative Modellierung am gleichen Prozessablauf bzw. Fragment und vor allem im Sinne der Viewpoint-basierten Modellierung wird nicht erzielt.

Ein Ansatz, der stärker auf das Zusammenfügen von verteilt erstellten Modellbereichen abzielt, ist der von [19] entwickelte Ansatz. Dieser zielt darauf ab, Modellierer beim Zusammenfügen und Integrieren verschiedener konzeptioneller Modelle zu unterstützen. Da hierbei ein Grad der Unsicherheit existiert, wird dieser mithilfe des Einsatzes von Unsicherheitsformalismen und der Nutzung von Erkenntnissen aus dem Bereich der „Belief Theory“ zu lösen versucht.

So zeigt sich bei der Betrachtung verwandter Arbeiten, dass bisher kein Ansatz existiert, der eine Verbindung des Viewpoint-Konzepts – der die Förderung der Involvierung verschiedener Akteure unterstützt – mit dem Forschungsfeld der kollaborativen Modellierung – das gerade die Koordinierung der Modellierung durch mehrere Akteure verfolgt – verbindet.

3 Methode zur Nutzung des Viewpoint-Konzepts zur Unterstützung kollaborativer Modellierung

3.1 Unterschiedliche Prinzipien der Modellbildung

Betrachtet man die beiden in der Betriebswirtschaftslehre überwiegend verwendeten Modelltheorien, nämlich den abbildungs- und konstruktionsorientierten Modellbegriff, so wird eine starke Bedeutung des subjektiven Einflusses auf das Modellverständnis deutlich (vgl. auch im Folgenden [20]). Während erstere das Subjekt in Bezug zum jeweiligen Modellzweck begreift und ihm einen beeinflussenden Einfluss auf die Modellgestaltung zuschreibt, hebt der konstruktionsorientierte Modellbegriff die Bedeutung des Modellierers als Subjekt hervor und unterstreicht dabei dessen Wechselwirkung mit dem Modellnutzer als Subjekt während des Konstruktionsprozesses von Modellen. Somit lassen sich zwei primäre Gestaltungsfaktoren erkennen, das „modellierende Subjekt“ und das „modellnutzende Subjekt“. In diesem Zusammenhang gilt es allerdings zu betonen, dass beide Ansätze jeweils von einem singulären Subjekt ausgehen bzw. eine singuläre Handlung eines subjektiven Kollektivs voraussetzen (vgl. Abbildung 1, links).

In der Realität sind bei Modellierungsvorhaben allerdings im Allgemeinen mehrere Einzelpersonen an der Modellbildung beteiligt. Auch ist zweifelsohne die Nutzung des Modells durch mehrere Personen zu attestieren. In der Praxis ist daher eher von einer kollaborativen Modellbildung auszugehen, d. h. dem Fall, in dem mehrere, nicht als Kollektiv handelnde Subjekte, gemeinsam ein Modell erzeugen (vgl. Abbildung 1, Mitte oben). Soweit sich eine solche multisubjektive Umgebung durch starke Autonomie der beteiligten Subjekte auszeichnet, kann nicht mehr von einer kooperativen Modellerstellung ausgegangen werden. Dies trifft insbesondere dann zu, wenn sich der Gegenstandsbereich des Originals zwischen den Modellierern unterscheidet. In diesem Fall muss eine kooperative Modellerstellung scheitern, da kein gemeinsamer Modellierungsgegenstand vorhanden ist. Vielmehr wird hier jeder Modellierer seine eigene Modellbildung durchführen, sodass für eine Menge von in Beziehung stehenden Originalen eine Menge von in Beziehung stehenden Modellsystemen entsteht. Eine solche Konstellation wird als multisubjektive Modellbildung bezeichnet und steht synonym für eine Modellbildung in multisubjektiver Umgebung autonomen Charakters (vgl. Abbildung 1, Mitte unten).

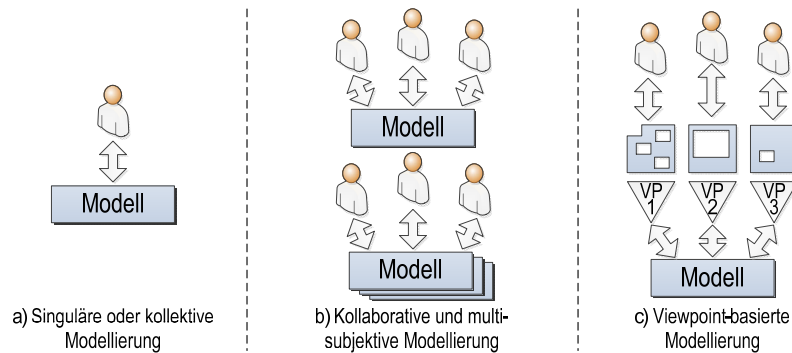


Abb. 1. Prinzipien der Modellbildung

Das in diesem Beitrag zugrundeliegende Modellierungsverständnis folgt dem sogenannten Viewpoint- bzw. Sichten-bezogenen Ansatz. Hier wird davon ausgegangen, dass im Modellierungsprozess – der Sichtweise der kollaborativen Modellierung folgend – mehrere Modellierer und Modellnutzer ein zu modellierendes System konstruieren und nutzen, die allesamt allerdings eine andere Zielsetzung mit dem Modell verfolgen können. Typischerweise lässt sich ein solches Szenario bei der Modellierung von Systemen unter Einbezug mehrerer Domänenexperten erkennen, die gemeinsam ein Gesamtsystem abzubilden versuchen. Folglich verlangt jedes beteiligte Subjekt – im Kontext des Viewpoint-Konzepts häufig als Stakeholder bezeichnet (vgl. bspw. [5]) – unterschiedliche Betrachtungsschwerpunkte und ggf. Repräsentationen zur Modellnutzung und -erstellung. Im Gegensatz zur multisubjektiven Modellierung, bei der mehrere Modelle durch die beteiligten Modellierer erstellt werden, wird beim Viewpoint-Ansatz hauptsächlich von der Erstellung eines Modells ausgegangen. Durch die Anwendung spezifischer Viewpoints auf dieses Modell wird sowohl dessen nutzerspezifische Darstellung als auch Anpassung erlaubt und dadurch der gemeinsame Modellierungsprozess adäquat auf die beteiligten Nutzer angepasst. Somit ermöglicht die Nutzung des Viewpoint-Konzepts eine Unterstützung der kollaborativen Modellierung. Die Verbindung beider Ansätze wird als Viewpoint-basierte Modellierung bezeichnet (vgl. Abbildung 1, rechts). Hierbei ist aus Mikrosicht die Viewpoint-spezifische Darstellung und Modellierung und aus Makrosicht die dynamische Koordination der Modellierung unter Beachtung der Anzahl (teilweise Vielzahl) an beteiligten Viewpoints zu berücksichtigen.

3.2 Statische Sicht der Viewpoint-basierten Modellierung

Wenngleich das Viewpoint-Konzept in diversen Domänen, wie dem Software Engineering oder dem Enterprise Architecture Management, Anwendung findet, so lässt sich kein domänenübergreifendes Verständnis des Konzepts attestieren [2]. Aus diesem Grund wurde in einer Vorarbeit zu diesem Beitrag eine domänenübergreifende Definition erarbeitet, die in diesem Abschnitt als Grundlage für die statische Konzeption der Viewpoint-basierten Modellierungsmethode genutzt wird. Anknüpfend daran

erfolgt im nachfolgenden Abschnitt 3.3 eine dynamische Auseinandersetzung mit der Viewpoint-basierten Modellierung.

Zur Konzeption der statischen Sicht werden zunächst grundlegende Begriffe eingeführt, basierend auf denen Viewpoint-spezifische Konzepte definiert werden.

Die Grundlage der Viewpoint-basierten Modellierung bilden *Metamodelle*. Diese definieren einen allgemeinen Rahmen zur Erzeugung von Modellen mittels einer Menge an Elementen sowie deren Beziehungen zueinander. Diese Elemente werden dabei als *Modellkonzepte* bezeichnet. Modellkonzepte sind daher Teil eines bestimmten Metamodells und dienen als Grundlage für die Instanziierung von Modellelementen in Modellinstanzen. Zur Erweiterung existierender Metamodelle können *Profile* eingesetzt werden. Diese nehmen ein Metamodell als Referenz und definieren die darin abstrakt spezifizierten Modellkonzepte für eine bestimmte Anwendungsdomäne um. Hierdurch wird die adäquate Adressierung einer Domäne durch das Metamodell ermöglicht. Ein Modell bzw. eine *Modellinstanz* stellt eine konkrete Ausprägung eines Metamodells dar und beinhaltet demnach eine bestimmte Menge an Modellelementen inkl. deren Beziehungen zueinander, die den im zugrundeliegenden Metamodell definierten Regeln folgen. Ein *Modellelement* ist somit als konkrete Ausprägung eines Modellkonzepts charakterisiert und stellt entweder ein Domänenobjekt oder eine Beziehung zwischen zwei oder mehreren Objekten dar.

Auf Basis dieser grundlegenden Begrifflichkeiten, welche die Basis für die Erzeugung von Modellen für eine bestimmte Anwendungsdomäne ermöglichen, werden nachfolgend Viewpoint-spezifische Begriffe eingeführt.

Entsprechend der Zielsetzung der Viewpoint-basierten Modellierung dienen *Viewpoints* der Unterstützung bestimmter Stakeholder, um ihren Beitrag zum zugrundeliegenden Systemdesign unter Berücksichtigung ihrer verschiedenen (fach-) spezifischen Perspektiven zu leisten. Unter Bezugnahme der zuvor eingeführten Modellbegriffe definiert ein Viewpoint demnach, welche Modellkonzepte von ausgewählten Metamodellen relevant für den jeweiligen Anwendungszweck, d. h. die Erzeugung, Nutzung und Manipulation von Modellinstanzen, eines Stakeholders sind. Somit bezieht sich ein Viewpoint auf eine Menge an Metamodellen und ggf. deren Erweiterungen durch Profile oder eine Teilmenge dieser. Folglich schränkt ein Viewpoint die ursprünglich zugrundeliegenden Metamodelle Stakeholder-spezifisch ein – die Definition lässt allerdings auch eine vollständige 1:1 Abdeckung zu. Um ein System durch seine zugrundeliegenden Modellinstanzen zielführend abzubilden, ist ein Viewpoint durch eine Menge an *ViewTypes* definiert, die er seinen Stakeholdern zur Abbildung von Modellinstanzen anbietet und die mittels Views instanziiert und visualisiert werden. Für jeden Viewpoint ist eine Menge an ViewTypes definiert, die jeweils eine bestimmte Perspektive eines Modells Stakeholder-spezifisch abbilden. Eine *View* – als konkrete Instanz eines ViewTypes – definiert ferner die Darstellung von Modellelementen für bestimmte Stakeholder und das Vorgehen, wie diese modifiziert werden können – dies wird für gewöhnlich mittels bestimmter Diagramme ermöglicht. Somit erlauben Views ihren Stakeholdern die Interaktion mit bestimmten Aspekten einer oder mehrerer Modellinstanzen, die dem Viewpoint zugrundeliegenden Metamodelle folgen. Während eine View die Darstellung von Modellelementen einer Modellinstanz realisiert, lässt die *Projektion* – als Spezialisierung einer View – über-

dies eine Stakeholder-spezifisch angepasste Visualisierung einzelner Modellelemente zu. Hierdurch ist nicht nur eine spezifische Selektion und Deselektion von Modellkonzepten möglich, sondern auch deren domänenadäquate Darstellung.

Die textuelle Darstellung der einzelnen Konzepte lässt bereits eine starke Verketten untereinander erkennen. Um den Zusammenhang des übergreifenden Viewpoint-Konzepts zu verdeutlichen, sind die einzelnen Konzepte daher in Abbildung 2 visualisiert.

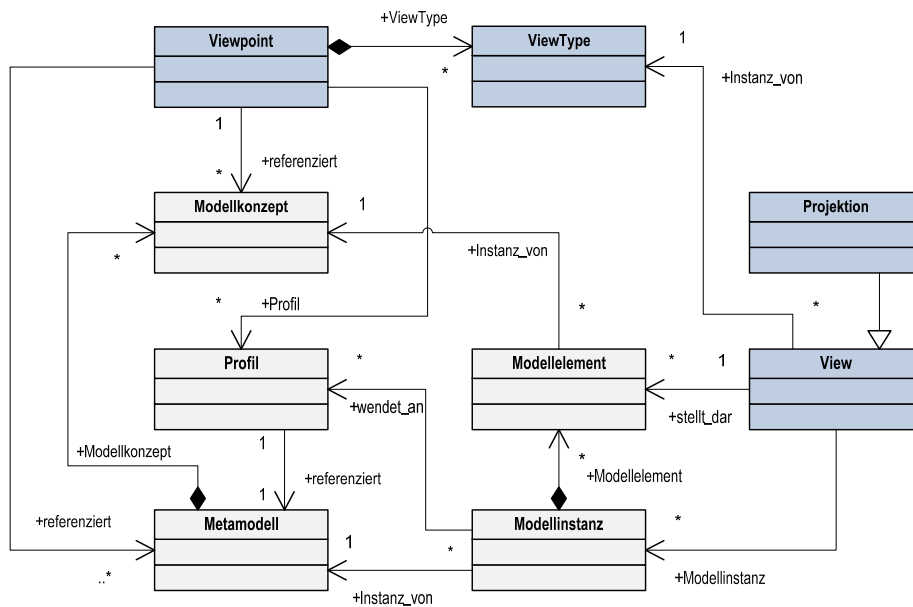


Abb. 2. Statische Sicht auf die Viewpoint-basierte Modellierung (in Anlehnung an [2])

3.3 Dynamische Sicht der Viewpoint-basierten Modellierung

Während die statische Konzeption der Viewpoint-basierten Modellierung primär die Aufspaltung von Modellen nach Stakeholder-spezifischen Anforderungen abdeckt, um diese zielgruppengerecht darzustellen, so muss überdies die dynamische Komponente im Sinne der Erzeugung und Manipulation von Modellinstanzen, d. h. die modellnutzende Komponente, ebenfalls konzipiert werden.

Zur Realisierung der Viewpoint-basierten Modellierung sollen die einzelnen Modellierungstätigkeiten so verteilt werden, dass das eigentliche Ziel der gemeinschaftlichen Schaffung einer konsistenten Modellinstanz erreicht wird. Die hierzu notwendigen Modellierungstätigkeiten werden demnach vom Verantwortlichen des jeweiligen Modellierungsprojektes in einzelne Modellierungsaufgaben heruntergebrochen. Ein reines Herunterbrechen ist allerdings nicht ausreichend, denn die einzelnen Aufgaben müssen den entsprechenden Modellierern zugeordnet werden.

Wird einem Stakeholder S_1 eine Modellierungsaufgabe zugeteilt, so wird ihm in seiner Modellierungsumgebung die zugrundeliegende Modellinstanz M basierend auf seinem Viewpoint VP_1 in der View $v_{VP_1,1}$ dargestellt (vgl. Abbildung 3, 1). Modifiziert er nun das zugrundeliegende Modell M mit seinem Domänenwissen, so kann hierdurch entweder eine konfliktfreie oder konfliktbehaftete Anpassung der Modellinstanz resultieren. Eine konfliktfreie Anpassung bedeutet (vgl. Abbildung 3, 2a), dass die Änderungen in seiner View $v_{VP_1,1}$ weder in Konflikt zu anderen Views $v_{VP_1,2-n}$ seines Viewpoints VP_1 noch zu anderen Viewpoints VP_{2-n} und deren Views $v_{VP_{2-n},1-m}$ als auch den kompletten zugrundeliegenden Metamodellen MM an sich stehen (vgl. hierzu [2]). In diesem Fall lässt sich die Modellanpassung ohne notwendige syntaktische Anpassung als konfliktfreie Modellinstanz M' speichern (vgl. Abbildung 3, 3a). Abgeleitet aus den zugrundeliegenden Metamodellen MM , die jeweils die zulässigen Beziehungen zwischen den Modellkonzepten definieren (vgl. Abschnitt 3.2), lassen sich syntaktische und zum Teil semantische Konflikte identifizieren. In der Literatur sind verschiedene Algorithmen hierzu publiziert worden (vgl. bspw. [21]), daher wird an dieser Stelle auf weitere technische Details verzichtet.

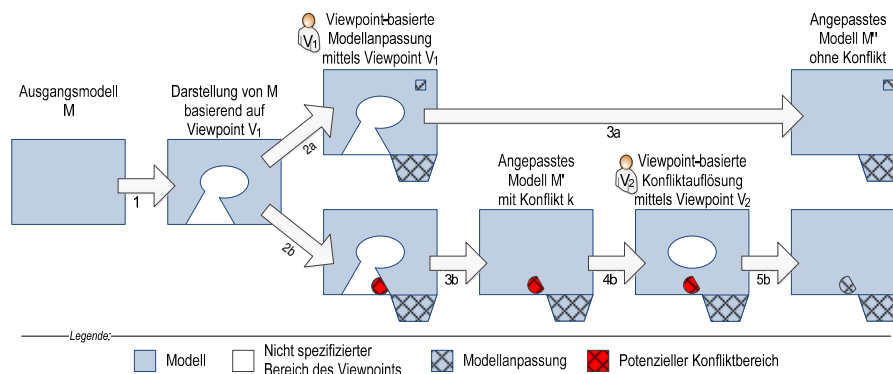


Abb. 3. Modellanpassungen bei der Viewpoint-basierten Modellierungsmethode

Während die konfliktfreie Modifikation als trivialer Fall anzusehen ist, gilt es für eine zielgerichtete kollaborative Modellierung auch den konfliktbehafteten Fall zu behandeln (vgl. Abbildung 3, 2b). An dieser Stelle sei explizit darauf hingewiesen, dass das Viewpoint-Konzept inkonsistente Zwischenzustände von Modellinstanzen erlaubt (vgl. bspw. [22]). Modifiziert Stakeholder S_1 daher die Modellinstanz M in seiner View $v_{VP_1,1}$ derart, dass seine Modellanpassungen einen Konflikt k in der angepassten Modellinstanz M' zu einem anderen Viewpoint VP_x oder einem zugrundeliegenden Metamodell MM_y darstellt (vgl. Abbildung 3, 3b) (globale Inkonsistenz), so muss eine syntaktische Anpassung erfolgen, die als zu erledigende Modellierungsaufgabe definiert werden kann. Sofern das konfliktbehaftete Modell M' konsistent zum gewählten Viewpoint VP_1 ist, kann der Stakeholder S_1 – sofern er nicht über weitere Viewpoints verfügt – eine globale Konsistenz nicht herstellen. Falls das Modell aber bereits inkonsistent zum gewählten Viewpoint ist (lokale Inkonsistenz), sollte Stakeholder S_1 in der Lage sein in einer anderen View v_x des ViewTypes VT_x den

Konflikt zu lösen. Bei lokaler Konsistenz, aber globaler Inkonsistenz kann somit Stakeholder S_1 die Inkonsistenz nicht lösen. Dies muss daher durch einen Stakeholder S_2 erfolgen, der mit seinem Viewpoint VP_2 die konfliktbehafteten Zustände k in M' auflösen kann (vgl. Abbildung 3, 4b). Hierzu muss ihm die entsprechende Modellierungsaufgabe zugeordnet werden. Nach der Anpassung der Modellinstanz M' durch Stakeholder S_2 ergibt sich nun erneut entweder ein konfliktbehafteter Zustand der Instanz oder eine konfliktfreie Modellinstanz M'' (vgl. Abbildung 3, 5b).

Um die Zuordnung solcher Modellierungsaufgaben zu unterschiedlichen Stakeholdern und deren Viewpoints zu ermöglichen – die dazu noch weitestgehend automatisch erfolgen soll bzw. dem Projektverantwortlichen in Form einer Empfehlung ausgesprochen wird – muss eine geeignete Datenbasis existieren. Diese muss Informationen hinsichtlich der Erfahrung von Stakeholdern im Kontext der Viewpoints, der Erfahrung mit ähnlichen Aufgaben und die derzeitige Modellierungsauslastung beinhalten. Auf dieser Grundlage kann eine Empfehlungsgenerierung erfolgen. Da es nicht Ziel dieses Beitrages ist, eine feingranulare Darstellung jedes Teilaspektes der Viewpoint-basierten Modellierung zu liefern, sondern primär die Schaffung eines Verständnisses über das Zusammenspiel der drei unterschiedlichen Teilfunktionen im Sinne einer integrierten Gesamtfunktionalität, wird an dieser Stelle auf eine detaillierte Darstellung verzichtet. Die Autoren verfolgen derzeit die Veröffentlichung eines weiteren Beitrags, der die Empfehlungsgenerierung expliziert.

3.4 Veranschaulichung der Viewpoint-basierten Modellierung anhand eines Szenarios

Zur Veranschaulichung des Zusammenspiels der einzelnen Aspekte des Konzepts ist in Abbildung 4 ein beispielhaftes kollaboratives Modellierungsszenario aufgezeigt.

Der Projektmanager des Modellierungsprojektes erstellt eine Aufgabenstruktur, welche die Modellierung des Projektes abdeckt. Bei der Zuweisung der Modellierer zu den einzelnen Aufgaben unterstützt ihn das Empfehlungssystem und schlägt ihm für jede umzusetzende Modellierungsaufgabe den geeignetsten Viewpoint vor. Die einzelnen Stakeholder erhalten daraufhin in ihrer Modellierungsumgebung eine Mitteilung über eine neue Modellierungsaufgabe. Auch hier kommt das Empfehlungssystem zum Tragen, indem es die günstigste Aufgabensequenz für den Modellierer empfiehlt, anhand derer er seine Aufgaben erledigen soll. Durch die Wahl des Viewpoints erhält der Modellierer überdies bereits die geeignete Benutzerumgebung, um die Modellierungsaufgabe zu lösen. Nach Abschluss der Aufgabe speichert der Modellierer seine Modellinstanz ins zentrale Modellrepository. Die im Hintergrund verlaufende Konsistenzprüfung anhand der Metamodelle generiert daraufhin bei Bedarf neue Modellierungsaufgaben für ihn oder andere Stakeholder.

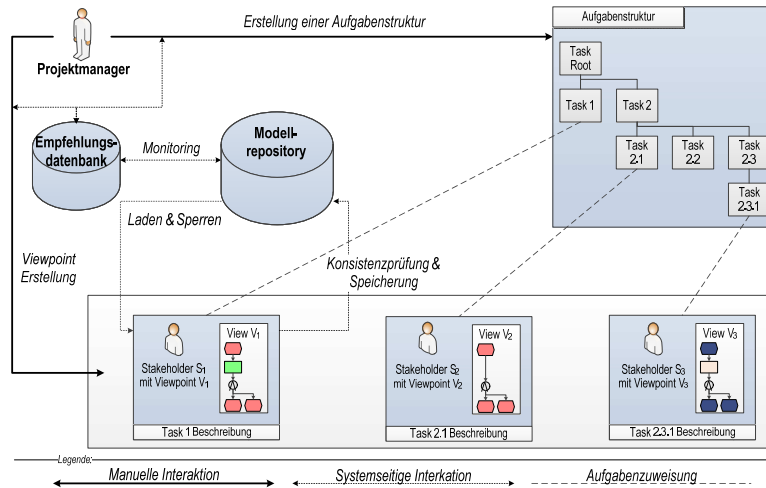


Abb. 4. Darstellung eines Viewpoint-basierten Modellierungsszenarios

4 Prototypische Konzeptimplementierung

Dem konstruktionsorientierten Forschungsansatz folgend, der diesem Beitrag zugrunde liegt (vgl. Abschnitt 1.3), wird in diesem Abschnitt das zuvor konzeptionell entwickelte Artefakt in seiner prototypischen Umsetzung *CIMFlex4CM* vorgestellt. *CIMFlex4CM* ist ein auf Eclipse aufbauendes Modellierungstool, das Benutzern eine eigene View auf gemeinsame Modelle erlaubt. Zudem realisiert das Tool die beschriebene Zuteilung von Modellierungsaufgaben. Abbildung 5 zeigt eine Viewpoint-spezifische Darstellung zweier Prozessmodelle in *CIMFlex4CM*, die auf den Metamodellen der EPK und BPMN beruhen und in getrennten Views dargestellt werden. Zudem ist ein Wizard zur Erzeugung von Viewpoints im unteren Bereich sichtbar.

Grundlegend basiert der entwickelte Prototyp auf den für *Eclipse* bereitgestellten Java-Frameworks *Eclipse Modeling Framework* (EMF) und *Graphical Modeling Framework* (GMF). Diese ermöglichen es, Modellierungstools in der Eclipse IDE zu entwickeln, ohne alle modellierungsspezifischen Funktionen von Grunde auf neu implementieren zu müssen. Hierzu erlaubt EMF anhand von strukturierten Metamodellen automatisch Java-basierten Quellcode zu erzeugen. Als Metamodelle dienen Ecore-Modelle, die auf dem EMOF (Essential Meta-Object Facility)-Standard basieren. Auf dieser Grundlage ist es nun mit GMF möglich, grafische Editoren auf den EMF-aufbauenden Modellen zu generieren (für weitere Details vgl. z. B. [23]).

Zur Realisierung der Viewpoint-basierten Modellierungsmethode wurde aufbauend auf EMF/GMF mittels zweier Generatoren die Erzeugung und Anpassung von Viewpoints und den dazugehörigen ViewTypes realisiert (vgl. Abbildung 6). Diese wurden basierend auf den in Eclipse bekannten Wizards implementiert.

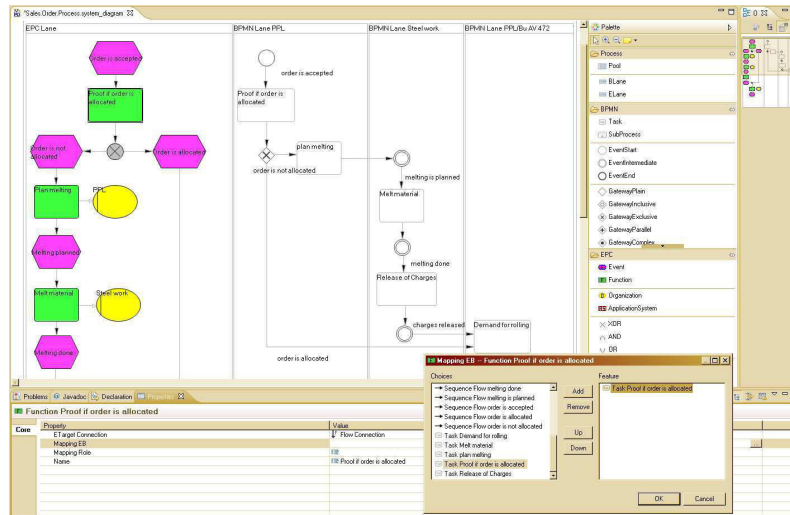


Abb. 5. Screenshot des CIMFlex4CM-Prototyps

Der entwickelte Viewpoint-Generator lädt hierzu das jeweilige Metamodell als Ecore-Datei und speichert die vom Viewpoint-Ersteller getroffene Selektion an Modellkonzepten in einer XML-basierten Datei ab. Neben der reinen Selektion von Konzepten können zudem sogenannte Flags gesetzt werden, die definieren, ob die ausgewählten Konzepte nur schreibgeschützt verwendet werden sollen und ob eine Hinterlegung mit weiteren Modellinstanzen möglich sein soll. Im letzten Fall muss zusätzlich das gewünschte Metamodell ausgewählt werden, um den hinterlegbaren Modelltyp zu definieren. Der Generator ermöglicht ebenfalls das Laden und Abändern bereits erstellter Viewpoints.

Zu den erstellten Viewpoints lassen sich mit Hilfe des zweiten Generators dazugehörige ViewTypes erzeugen. Hierzu wird die erzeugte Viewpoint-Datei als auch eine dazugehörige Default-View-Datei, welche die Standarddarstellung der Metamodellkonzepte definiert, geladen. Der ViewType-Generator ermöglicht es nun, die Viewpoint-spezifische Konzeptauswahl weiter einzuschränken als auch die Selektion von bestimmten Modellelementen zu ermöglichen. Falls gewünscht kann überdies eine View-spezifische Darstellungsform gemäß der Projektion definiert und ebenfalls als XML-basierte Datei gespeichert werden. Sind sowohl Viewpoint als auch eine Menge an ViewTypes erstellt, kann ein entsprechendes Eclipse Plugin generiert und an den jeweiligen Endnutzer ausgerollt werden. Dieser kann das Plugin laden und von seiner Eclipse-Instanz die gewünschte Modellinstanz aus einem zentral angelegten Modellrepository öffnen und sodann Viewpoint-spezifisch darstellen und bearbeiten.

Um die Viewpoint-basierte Modellierung (vgl. Abschnitt 3.3) zu realisieren und den Aspekt der Modellierungsaufgaben, die es innerhalb eines (kontinuierlichen) Modellierungsprojektes zu bewältigen gibt, in den Prototyp zu integrieren ist dieser mit dem Eclipse Mylyn Plug-In verknüpft. Mylyn stellt ein *Task and Application Lifecycle Management* Rahmenwerk dar, das ein aufgabenfokussiertes Interface sowie verschiedene *Issue Tracking* Funktionalitäten in Eclipse implementiert.

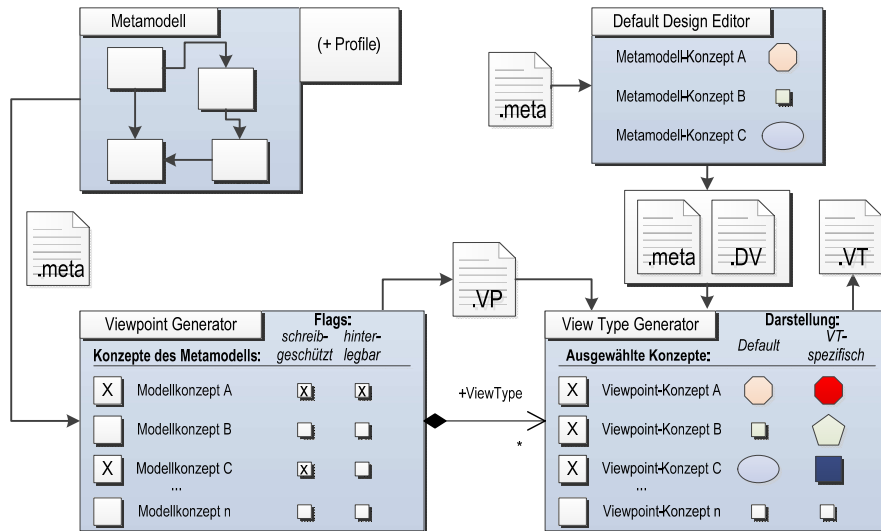


Abb. 6. Erzeugung von Viewpoints und ViewTypes

Zur optimalen Verteilung der einzelnen Modellierungsaufgaben an die geeigneten Stakeholder, ist das Mylyn Plugin überdies an das Empfehlungssystem Apache Mahout geknüpft, das die optimale Zuteilung der einzelnen Aufgaben basierend auf den vorhandenen Viewpoints und dazugehörigen ViewTypes realisiert und so bestimmten Stakeholdern zuordnet. Um eine Datengrundlage für das eingesetzte Empfehlungssystem zu schaffen, beobachtet das Issue Tracking System von Mylyn u. a. die Zuteilung und das Lösen von Aufgaben als auch entstandene Probleme hierbei. Hierdurch wird ebenfalls implizit ein Reporting innerhalb eines Modellierungsprojektes erzielt. Zur Empfehlungsgenerierung werden dann die einzelnen Stakeholder nach ihrer Erfahrung im Kontext des Viewpoints, ihrer Erfahrung mit ähnlichen Aufgaben und wie ihre derzeitige Auslastung mit Modellierungsaufgaben ist, eingestuft. Es gilt allerdings anzumerken, dass das eingesetzte System wie alle Recommender eine gewisse Datenbasis benötigt, um zufriedenstellende Empfehlungen geben zu können.

Somit ermöglicht der Prototyp eine gezielte Zuordnung einzelner Modellierungsaufgaben unter den kollaborierenden Teilnehmern. Überdies empfiehlt das integrierte Empfehlungssystem zusätzlich dem Nutzer indirekt die geeignete Modellierungsumgebung, die im Viewpoint bzw. den ViewType definiert ist. Als dritte Funktionalität ermöglicht das Empfehlungssystem zudem die optimale Aufgabenreihenfolge sowohl aus Projekt- als auch Stakeholder-bezogener Perspektive.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag konzipierte eine Viewpoint-basierte Modellierungsmethode, die das sogenannte Viewpoint-Konzept mit der kollaborativen Modellierung verknüpft. Durch die Nutzung von Stakeholder-spezifischen Viewpoints werden typische

Problemstellungen der kollaborativen Modellierung – z. B. die häufig geringe Akzeptanz von Modellen unter beteiligten Akteuren aus unterschiedlichen Domänen aufgrund deren eingeschränkten Verständnisses der übergreifenden Modelle – angegangen. Resultierend erleichtert und verbessert das Viewpoint-Konzept die Einbeziehung unterschiedlicher Stakeholder der verschiedensten Domänen in den kollaborativen Modellierungsprozess. Neben der Konzeption der Viewpoint-basierten Modellierungsmethode zeigte der Beitrag überdies dessen Proof-of-Concept anhand einer prototypischen Implementierung auf.

Das in diesem Beitrag publizierte Forschungsergebnis ist innerhalb des europäischen Forschungsprojektes *ViBaM* (Viewpoint-Based Modeling) entstanden. In Folge des Projektes planen die beiden beteiligten Industriepartner *SOFTEAM* und *IKV++ TECHNOLOGIES* die Integration des entwickelten Konzepts in deren kommerziellen Modellierungswerkzeuge *MODELIO* und *MEDINI ANALYZE*. Hierdurch wird die kollaborative Modellierung in den beiden Anwendungsdomänen der Projektpartner – die Unternehmensmodellierung sowie die Analyse funktionaler Sicherheit bei der Entwicklung von Automobilelektronik – unterstützt.

Danksagung. Die hier vorgestellten Forschungsergebnisse wurden im Rahmen des Forschungsprojektes *ViBaM* (Viewpoint-Based Modeling) erarbeitet, das vom Bundesministerium für Bildung und Forschung (BMBF) (FKZ 01QE1106B/C) innerhalb des Europäischen Förderprogramms Eurostars (E! 5529) gefördert wurde. Die Autoren möchten sich weiterhin bei den anonymen Gutachtern für die konstruktiven Anmerkungen bedanken, die zur Verbesserung dieser Artikel beigetragen haben.

Literatur

1. Stachowiak, H.: Allgemeine Modelltheorie. Springer, Berlin Heidelberg (1973)
2. Fischer, K., Panfilenko, D., Krumeich, J., Born, M., Desfray, P.: Viewpoint-Based Modeling - Towards Defining the Viewpoint Concept and Implications for Supporting Modeling Tools. In: EMISA 2012 - International Workshop on Enterprise Modelling and Information Systems Architectures, pp. 123-136. Köllen GmbH, Bonn (2012)
3. Finkelstein, A., Kramer, J., Nuseibeh, B., Finkelstein, L., Goedicke, M.: Viewpoints: A Framework for Integrating Multiple Perspectives. *Syst. Development* 2 (1), 31-57 (1992)
4. Dijkman, R.M., Quartel, D.A.C., van Sinderen, M.J.: Consistency in multi-viewpoint design of enterprise information systems. *Information and Software Technology* 50 (7-8), 737-752 (2008)
5. Goldschmidt, T., Becker, S., Burger, E.: Towards a Tool-Oriented Taxonomy of View-Based Modelling. In: Sinz, E.J., Schürr, A. (eds.): *Modellierung 2012*. LNI, Vol. 201, pp. 59-74. Gesellschaft für Informatik, Bonn (2012)
6. Easterbrook, S., Yu, E., Ar, J., Fan, Y., Horkoff, J., Leica, M., Qadir, R.A.: Do viewpoints lead to better conceptual models? An exploratory case study. In: *RE '05 Proceedings of the 13th IEEE International Conference on Requirements Engineering*, pp. 199-208. IEEE Computer Society, Los Alamitos (2005)
7. Frost & Sullivan: *Meetings Around the World: The Impact of Collaboration on Business Performance*. Frost & Sullivan White Papers (2007)

8. Renger, M., Kolfshoten, G.L., de Vreede, G.-J.: Challenges in Collaborative Modeling: A Literature Review. In: Dietz, J.L.G., Albani, A., Barjis, J. (eds.): *Advances in Enterprise Engineering I. LNBIP*, Vol. 10, pp. 61-77. Springer, Berlin Heidelberg (2008)
9. Rittgen, P.: Collaborative Modeling - A Design Science Approach. In: *HICSS '09 42nd Hawaii International Conference on System Sciences*. IEEE Computer Society, Los Alamitos (2010)
10. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. *MIS Quarterly* 28 (1), 75-105 (2004)
11. Wood-Harper, A.T., Antill, L., Avison, D.E.: *Information systems definition: the Multiview approach*. Blackwell Scientific Publications, Oxford (1985)
12. Lankhorst, M.: *Enterprise Architecture at Work: Modelling, Communication and Analysis*. Springer, Berlin Heidelberg (2009)
13. Software Engineering Standards Committee of the IEEE Computer Society: *IEEE Recommended Practice for Architectural Description of Software-Intensive Systems*. IEEE Computer Society, Los Alamitos (2007)
14. The Open Group. Welcome to TOGAF Version 9 – An Open Group Standard, <http://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>
15. Andrade, J., Ares, J., García, R., Pazos, J., Rodríguez, S., Silva, A.: A Methodological Framework for Viewpoint-Oriented Conceptual Modeling. *IEEE Transactions on Software Engineering* 30 (5), 282-294 (2004)
16. Cicchetti, A., Ciccozzi, F., Leveque, T.: Supporting incremental synchronization in hybrid multi-view modeling. In: *MODELS'11 Proceedings of the 2011th international conference on Models in Software Engineering*. LNCS, Vol. 7167, pp. 89-103. Springer, Berlin Heidelberg (2012)
17. Rittgen, P.: Collaborative Business Process Modeling - Tool Support for Solving Typical Problems. In: *CONF-IRM 2010 Proceedings*, paper 45. AIS Electronic Library (2010)
18. Kim, K.-H., Won, J.-K., Kim, C.-M.: A Fragment-Driven Process Modeling Methodology. In: *Computational Science and Its Applications – ICCSA 2005 Proceedings*. LNCS, Vol. 3482, pp. 817-826. Springer, Berlin Heidelberg (2005)
19. Bagheri, E., Ghorbani, A.A.: A belief-theoretic framework for the collaborative development and integration of para-consistent conceptual models. *Journal of Systems and Software* 82 (1), 707-729 (2008)
20. Werth, D.: *Modellierung unternehmensübergreifender Geschäftsprozesse: Modelle, Notationen und Vorgehen für prozessorientierte Unternehmensverbände*. EHV, Bremen (2005)
21. Paige, R.F., Brooke, P.J., Ostroff, J.S.: Metamodel-based model conformance and multiview consistency checking. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 16 (3), article 11 (2007)
22. Finkelstein, A., Gabbay, D., Hunter, A., Kramer, J., Nuseibeh, B.: Inconsistency Handling in Multiperspective Specifications. *IEEE Transactions on Software Engineering* 20 (8), 569-578 (1994)
23. Rempp, G., Akermann, M., Löffler, M., Lehmann, J.: *Model Driven SOA*. Springer, Berlin Heidelberg (2011)

Eine Untersuchung der Potentiale automatisierter Abstraktionsansätze für Geschäftsprozessmodelle im Hinblick auf die induktive Entwicklung von Referenzprozessmodellen

Jana-Rebecca Rehse, Peter Fettke, und Peter Loos

Institut für Wirtschaftsinformatik (IWi)
im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI),
Universitätscampus, Geb. D3.2, 66123 Saarbrücken, Germany
{Jana-Rebecca.Rehse, Peter.Fettke, Peter.Loos}@iwi.dfk.de

Abstract. In der Literatur wurde bereits eine Reihe von Ansätzen zur Abstraktion für Geschäftsprozessmodelle identifiziert und darüber hinaus verschiedene allgemeine Anwendungsszenarien vorgestellt. Die induktive Entwicklung von Referenzmodellen stellt eine weitere mögliche Anwendung bereits bestehender Abstraktionsansätze dar. Dieser Beitrag zielt darauf ab, Abstraktion in Geschäftsprozessmodellen im Allgemeinen sowie konkret einige dieser Ansätze bezüglich ihrer Eignung zur Entwicklung von induktiven Referenzprozessmodellen zu bewerten. Dazu wird ein generischer Algorithmus entwickelt, der aus einer Menge von Prozessmodellen ein Referenzprozessmodell gewinnt, indem die Modelle gemäß der Vorgehensweise des verwendeten Ansatzes abstrahiert und die abstrahierten Modelle anschließend zusammengefügt werden. Dieses Vorgehen wird genutzt, um acht verschiedene Abstraktionsansätze hinsichtlich ihrer Potentiale im Rahmen der induktiven Entwicklung von Referenzmodellen zu evaluieren, indem sie auf eine Menge von Modellen angewendet werden. Die resultierenden Referenzmodelle werden bewertet und verglichen, um einen Überblick über die verschiedenen Ansätze und ihre Eignung zu erhalten.

Keywords: induktive Entwicklung von Referenzprozessmodellen, Abstraktion in Geschäftsprozessmodellen

1 Einführung

Abstraktion als Idee der Vereinfachung oder Verallgemeinerung von Sachverhalten existiert im wissenschaftlichen Diskurs nicht nur als erkenntnistheoretisches Konzept, sondern auch als spezifische Anwendung in einzelnen Disziplinen. Eine mögliche Interpretation der Abstraktion in Geschäftsprozessmodellen ist, relevante Eigenschaften eines Modelles zu bewahren, während unwesentliche Details ausgeblendet werden können. Diese Interpretation der Abstraktion soll im Folgenden anstelle des allgemeinen Abstraktionsbegriffs verwendet werden. Da bereits vorhandene Arbeiten aufge-

griffen und der Fokus auf die Automatisierung gelegt wird, wird eine mathematisch-formale Definition des Begriffs zugrunde gelegt (vgl. [16]).

Das Konzept ist bereits von verschiedenen Autoren untersucht worden. Neben einer Reihe konkreter Ansätze wie diese Art der Abstraktion zu realisieren ist [4], [10-11], [13] gibt es verschiedene konzeptuelle Arbeiten zum Thema [14], [10] sowie eine Veröffentlichung, die nicht nur ein entsprechendes Rahmenwerk ausführlich definiert, sondern auch Anwendungsszenarien und ausgewählte Ansätze vorstellt [16]. In [14] werden außerdem eine Reihe praktischer Anwendungsfälle für Abstraktionsansätze identifiziert. Diese Publikationen stellen entweder die spezifische Vorgehensweise ihres Ansatzes vor oder ausgewählte, bestehende Ansätze allgemein gegenüber. Die induktive Entwicklung von Referenzprozessmodellen ist dagegen bisher noch nicht als mögliches Anwendungsgebiet von Abstraktionsansätzen identifiziert worden. Da ein induktives Vorgehen, bei dem ausgehend von einer Reihe unterschiedlicher Prozessmodelle ein Referenzprozessmodell entwickelt wird, gezielt Gemeinsamkeiten der Modelle betont und von unternehmensspezifischen Besonderheiten absieht, bieten sich zur Realisierung dieser Modifikationen und zur Generalisierung eines Modells Abstraktionsansätze an, die gegebenenfalls auch weitgehend automatisiert werden können. Die Potentiale bestehender Ansätze bezüglich dieser Anwendung sollen hier betrachtet werden. Dabei ist das primäre Ziel des Beitrags nicht, einen weiteren Abstraktionsansatz zu definieren, sondern detailliert zu evaluieren, wie die Vielzahl an bestehenden Ansätzen allgemein für die induktive Entwicklung von Referenzprozessmodellen verwendet werden kann und was die Vor- und Nachteile der einzelnen Ansätze sind. Ausgehend von dieser Evaluation können dann allgemeine Anforderungen an einen Abstraktionsansatz, der für die induktive Entwicklung von Referenzprozessmodellen verwendet werden soll, formuliert und konkrete Handlungsempfehlungen abgeleitet werden.

Dazu werden zunächst in Abschnitt 2 einige Grundlagen der Abstraktion in Geschäftsprozessmodellen aufgegriffen und kurz erläutert. Abschnitt 3 geht auf die Evaluation der einzelnen Ansätze ein, indem die Auswahl- und Bewertungskriterien erklärt, ein generischer Algorithmus zur Anwendung der Ansätze definiert und die verwendeten Beispielmodelle illustriert werden. In Abschnitt 4 wird die konkrete Anwendung am Beispiel von drei exemplarischen Abstraktionsansätzen vorgestellt, deren Resultate in Abschnitt 5 verglichen und diskutiert werden. Abschnitt 6 beschließt den Beitrag mit einer Schlussfolgerung und der Betrachtung weiterführender Fragen.

2 Grundlagen der Abstraktion in Geschäftsprozessmodellen

Im Folgenden werden die in [16] eingeführten und formal definierten Begrifflichkeiten für die Beschreibung von Abstraktionsansätzen verwendet. Diese sollen hier noch einmal kurz erklärt und in Abbildung 1 illustriert werden. In den betrachteten Arbeiten werden Modelle auf zwei Arten abstrahiert: Entweder wird der Abdeckungsgrad der Modelle reduziert, indem man Objekte entfernt (*Eliminierung*) oder die Granularität wird erhöht, indem man Objekte zusammenfasst (*Aggregation*), sodass das abstrahierte Modell weniger Informationen zur Verfügung stellt als das Ausgangsmodell.

Abstraktionsobjekte sind die Teile eines Modells, die relevante Informationen beschreiben und daher für die Abstraktion in Frage kommen, hier also Knoten und Kanten. *Abstraktionskriterien* dienen dazu, Abstraktionsobjekte zu beschreiben, indem sie damit assoziierte Informationen bereitstellen. Die dazugehörigen *Signifikanzfunktionen* treffen basierend auf den Kriterien die Entscheidung, ob Objekte abstrahiert werden sollen.

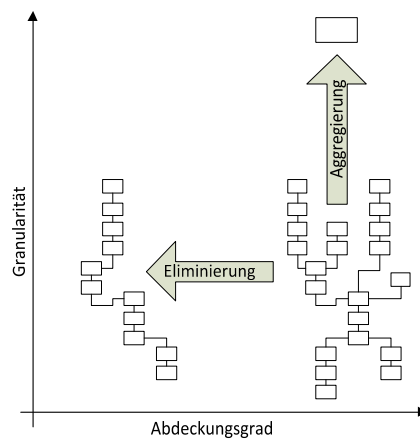


Abb. 1. Illustration von Aggregation und Eliminierung (vgl. [14])

3 Konzeption der Evaluation

3.1 Einordnung der ausgewählten Abstraktionsansätze

Die ausgewählten Abstraktionsansätze werden miteinander verglichen, um die Eignung für die hier vorgestellte Anwendung einzuschätzen, eine detaillierte Analyse durchzuführen und die Automatisierung zu ermöglichen. Diese Evaluation erfolgt anhand des im Folgenden erklärten Kriterienkatalogs, mit dem die Ansätze eingeordnet und bewertet wurden. Diese Kriterien entstammen zum einen dem verwendeten Rahmenwerk zur Beschreibung der Abstraktion [16], sodass sie einen Vergleich der Vorgehensweisen ermöglichen, zum anderen den identifizierten Anforderungen hinsichtlich der induktiven Entwicklung von Referenzprozessmodellen, sodass die tatsächlichen Potentiale der Ansätze bewertet werden können.

1. Prinzip: Welches der möglichen Prinzipien wird verwendet? Dabei wurden bei der Betrachtung der Ansätze das syntaktische Prinzip, das allein anhand syntaktischer Kriterien Abstraktionsentscheidungen trifft, das strukturelle Prinzip, das Informationen über die Prozessstruktur beachtet, und das semantische Prinzip, das versucht, Abstraktionsentscheidungen anhand der Semantik des Modells oder der Modellierungssprache zu treffen, identifiziert.
2. Nutzerorientierung: Wie gut lässt sich die Abstraktion auf die unterschiedlichen Bedürfnisse einzelner Nutzer anwenden?

3. Übergeordnetes Ziel: Wie motivieren die Autoren die Abstraktion und wofür soll sie verwendet werden?
4. Abstraktionsobjekte: Welche Objekte des Modells sollen abstrahiert werden?
5. Abstraktionskriterien: Welche Eigenschaften des Objekts betrachtet man?
6. Signifikanzfunktionen: Wie wird entschieden, ob Objekte abstrahiert werden?
7. Abstraktionsoperationen: Wird Aggregation oder Eliminierung verwendet?
8. Implementierung: Ist eine Implementierung im Forschungsprototyp vorhanden?
9. Anpassung: Wie gut lässt sich der Ansatz auf verschiedene Modellierungssprachen anwenden?
10. Benötigte Informationen: Welche Informationen werden zusätzlich zum Prozessmodell verwendet, um den Ansatz anzuwenden?
11. Mögliche Ergänzungen: Wie kann der vorgestellte Ansatz sinnvoll ergänzt werden?

Neben diesen allgemeinen Kriterien wurden einige Maßstäbe ergänzt, die konkret die Eignung zur induktiven Entwicklung von Referenzprozessmodellen bewerten sollen.

12. Vernachlässigung von Besonderheiten: Wie gut können unternehmensspezifische Besonderheiten ausgeblendet werden?
13. Betonung gemeinsamer Strukturen: Wie gut werden Gemeinsamkeiten der einzelnen Modelle identifiziert und ins Referenzmodell übernommen?
14. Sprachliche Aspekte: Wird auf die Gestaltung sprachlicher Konstrukte (bspw. Knotenlabels) eingegangen?

3.2 Generischer Algorithmus zur Evaluation

Aus einer Menge von Prozessmodellen soll heuristisch ein Referenzmodell entwickelt werden, indem mithilfe eines beliebigen Abstraktionsansatzes jedes Modell einzeln abstrahiert und die Modelle anschließend zusammengefügt werden. Die jeweiligen Abstraktionsobjekte, Abstraktionskriterien und Abstraktionsoperationen müssen also bestimmt werden. Dazu wird ein generischer Algorithmus definiert, der, gegeben eine Menge von Prozessmodellen, mithilfe von Abstraktionsschritten ein Referenzmodell generiert. Ein Abstraktionsschritt besteht aus einem Prozessmodell, einer Menge von darin enthaltenen Abstraktionsobjekten und einer Abstraktionsoperation, die auf diesen Objekten durchzuführen ist. Der generische Algorithmus stellt den Abstraktionsschritt als Schnittstelle bereit, führt die entsprechenden Operationen aus und fügt die Modelle anschließend zusammen. Letzteres kann sowohl manuell als auch mithilfe eines geeigneten Algorithmus wie beispielsweise [8] geschehen.

Das Ziel ist, mithilfe von Abstraktion gezielt unternehmensspezifische Besonderheiten auszublenden, Gemeinsamkeiten der Modelle zu erkennen und die Granularität der Modelle zu vereinheitlichen. Dies ist in Abbildung 2 zu sehen: Während sich Modell A und B nicht sofort zusammenfügen lassen, kann man mittels Abstraktion die Funktionen „Auftrag fachlich prüfen“ und „Auftrag kaufmännisch prüfen“ zu der höher granularen Funktion „Auftrag prüfen“ aggregieren. Da die beiden Modelle danach gleich sind, ergibt sich unmittelbar ein Referenzmodell (die Zuordnung der einzelnen Knoten ist farblich gekennzeichnet).

Der Abstraktionsgrad der Modelle kann vereinheitlicht werden, indem man das Abstraktionskriterium entsprechend wählt. So kann bei gleicher Granularität die gleiche Anzahl Abstraktionsschritte durchgeführt werden. Allerdings ist der Abstraktionsgrad eines Modelles automatisiert nur schwer zu erkennen, daher muss dies im Zweifelsfall manuell festgelegt werden.

Programm: Generischer Algorithmus zur Generierung von Referenzmodellen mittels Abstraktion

```

Menge von Prozessmodellen  $p_1, \dots, p_n$ 
begin
  for all  $p_i$  in  $p_1, \dots, p_n$ 
    while ( $p_i$  contains insignificant objects)
       $S =$  find insignificant objects in  $p_i$ ;
       $op =$  find operation in  $p_i$ ;
      Apply  $op$  onto  $S$  in  $p_i$ ;
  merge  $p_1, \dots, p_n$  into  $p$ ;
  return  $p$ ;
end

```

Die unterstrichenen Operationen müssen durch eine Schnittstelle Abstraktionsalgorithmus definiert werden.

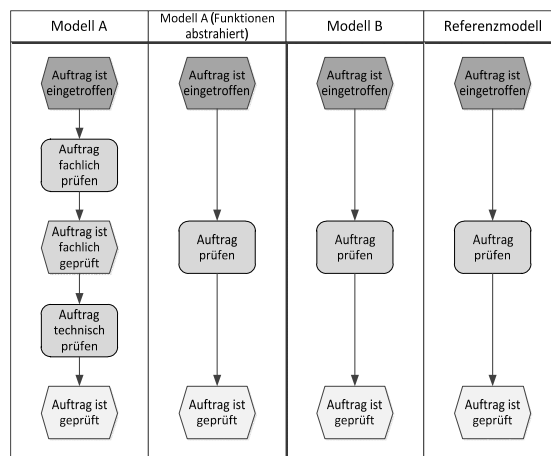


Abb. 2. Beispiel zur Erläuterung des generischen Algorithmus

3.3 Auswahl der betrachteten Algorithmen

In der Literatur existiert eine Reihe von Veröffentlichungen, die sich mit Abstraktion in Geschäftsprozessmodellen beschäftigt, wobei Abstraktion auch implizit zur Generierung von Prozesssichten [2], [7] oder zur Prozessvereinfachung [11] verwendet wird. Eine solche Auswahl befindet sich bereits in [16], allerdings wurde hier nur ein

Teil der dort betrachteten Ansätze übernommen und um andere, dort nicht erwähnte Ansätze ergänzt. Schlussendlich wurden insgesamt acht Ansätze nach folgenden Kriterien ausgewählt:

- **Automatisierung:** Die gezielte Automatisierung der Abstraktion und damit der induktiven Entwicklung der Referenzmodelle steht hier im Mittelpunkt. Die Abstraktion ist daher möglichst mathematisch-formal definiert worden und es wurden solche Ansätze ausgewählt, die dieser Definition entsprechen und die Automatisierung ermöglichen.
- **Anwendung auf Geschäftsprozessmodelle bzw. ereignisgesteuerte Prozessketten [5]:** Da die verschiedenen Ansätze später anhand von als EPK modellierten Prozessmodellen aus [1] evaluiert werden sollen, wurde die Auswahl so eingeschränkt, dass nur solche Ansätze betrachtet wurden, die auf einfachen, allgemeinen Modellierungssprachen oder EPKs definiert oder die leicht auf solche anzupassen sind.
- **Möglichst allgemeine Anwendung:** Um die Vergleichbarkeit der Algorithmen zu gewährleisten, sollten sie möglichst allgemein motiviert und somit auch im hier gewählten Kontext anwendbar sein. Daher wurden solche Ansätze, die ein sehr spezifisches Ziel (bspw. Abstraktion von Prozesssichten für externe Partner, vgl. [3]) verfolgen, hier nicht betrachtet.
- **Definition von Abstraktionsobjekten, -kriterien und -operationen:** Damit sie überhaupt sinnvoll angewendet werden können, sollten die gewählten Ansätze dem hier verwendeten Rahmenwerk der Abstraktion in Geschäftsprozessmodellen entsprechenden und dessen formale Anforderungen möglichst gut erfüllen.
- **Gleichmäßige Verteilung der Prinzipien:** Syntaktische, strukturelle und semantisch eingeordnete Ansätze sollten gleichermaßen betrachtet werden.

3.4 Vorgehensweise

Das verwendete Beispiel besteht aus den Modellen drei verschiedener Rechnungsprüfungsprozesse, die in [1] dokumentiert sind und hier als Modell A, B und C bezeichnet werden (siehe Abbildung 3). Die Zuordnung der einzelnen Prozessbestandteile wird durch die Identität der Knotenlabels beschrieben, das heißt, dass Knoten mit gleicher Beschriftung auch als identisch betrachtet werden. In Abbildung 3 sind exemplarisch die Knoten, die in allen Modellen enthalten sind, farblich hervorgehoben. Hier wurde die Zuordnung aufgrund der Knotenlabels gewählt, für die vorgestellte Anwendung kann allerdings auch eine beliebige andere Zuordnung der Modellteile gewählt werden. Aufgrund der gewählten Zuordnung weisen die Modelle denselben Abstraktionsgrad auf. Sie werden im Folgenden mithilfe eines bzw. mehrerer Schritte abstrahiert und anschließend möglichst passend zusammengefügt.

Um zu evaluieren, inwieweit sich die ausgewählten Ansätze zur Entwicklung eines Referenzprozessmodells eignen, werden im folgenden Abschnitt drei Abstraktionsansätze auf ein Beispiel angewendet. Es wurden ein syntaktischer, ein semantischer und ein „hybrider“ Ansatz, dessen Prinzip durch die Kriterien bestimmt werden kann, gewählt.

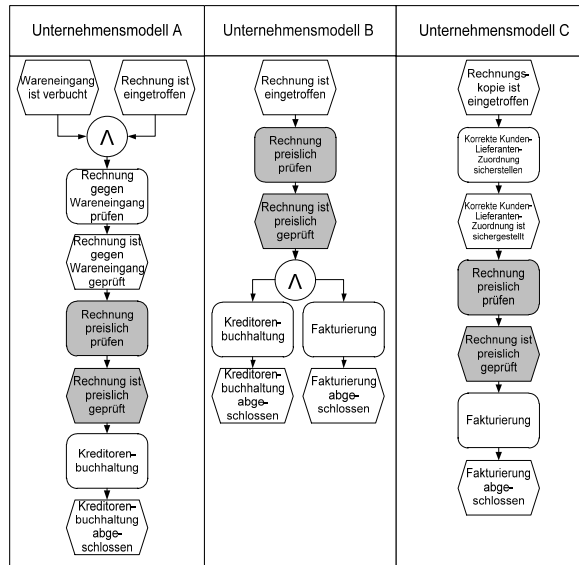


Abb. 3. Beispiele für individuelle Unternehmensmodelle aus [1]

4 Evaluierung der Abstraktionsansätze bezüglich induktiver Referenzprozessmodellentwicklung

4.1 Polyvyanyy et al.: Reducing the Complexity of Large EPCs [11]

Die Operationen dieses Ansatzes sind bereits auf EPKs definiert und können entsprechend der Modelltopologie angewendet werden. Prinzipiell sind bei unterschiedlicher Reihenfolge der verschiedenen Operationen unterschiedliche Modelle zu erwarten. Als Abstraktionskriterium dient hier die relative Häufigkeit der Knoten innerhalb der drei Modelle. In jedem Modell werden die Knoten mit der niedrigsten relativen Häufigkeit identifiziert und die Modellteile, in denen sie sich befinden, gemäß den vorgegebenen Regeln abstrahiert, wobei die Ereignisse entsprechend an die neugebildeten Funktionen angepasst werden müssen.

Diese Knoten in Modell A sind „Wareneingang ist verbucht“ und „Rechnung gegen Wareneingang prüfen“. Ersterer bleibt bestehen, da es keine entsprechende Abstraktionsoperation in diesem Ansatz gibt. Letzterer wird innerhalb der Sequenz mit „Rechnung preislich prüfen“ in die allgemeinere Funktion „Rechnung prüfen“ aggregiert. In Modell B tritt zwar jeder Knoten in mindestens zwei Modellen auf, der Operatorblock aus „Kreditorenbuchhaltung“ und „Fakturierung“ ist dabei am seltensten. Er wird in die Funktion „Rechnung verbuchen“ aggregiert. Modell C enthält den Knoten „Korrekte Kunden-Lieferanten-Zuordnung sicherstellen“, der ebenfalls mit „Rechnung preislich prüfen“ zu „Rechnung prüfen“ aggregiert wird.

Hier ergibt sich direkt eines der Probleme des generischen Algorithmus. Zwar lässt sich durch die elementar gehaltenen Abstraktionsschritte der Abstraktionsgrad des

resultierenden Modells gut kontrollieren. Ist aber in einem Modell nur ein Teil eines größeren Prozesses enthalten, können diese weder abstrahiert noch miteinander identifiziert werden. Hier wurde stattdessen eine linguistische Abstraktion durchgeführt, die Knoten also nach der syntaktischen Abstraktion gemäß dem generischen Prozess umbenannt. Dieser Schritt verfälscht die Abstraktion und ist somit kritisch zu betrachten, allerdings ist dies die einzige Möglichkeit, die abstrahierten Modelle überhaupt zusammenzufügen. Diese Schwierigkeiten ergeben sich auch aus dem hier gewählten Abstraktionskriterium und aus der geringen Größe der Modelle. Das vorgeschlagene Kriterium der Transitionswahrscheinlichkeiten konnte hier nicht angewendet werden, sodass weitere Kriterien überprüft werden sollten. Auf die Benennung der neuen Knoten wird hier ebenfalls nicht eingegangen, sodass die Vorgänge nur schwer automatisiert werden können.

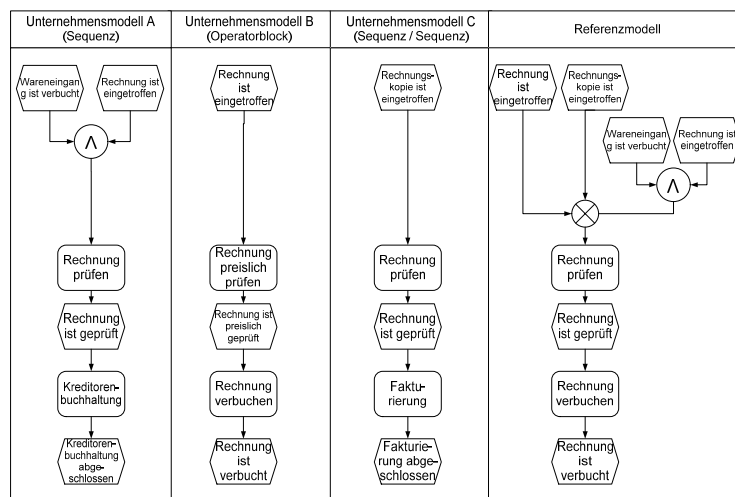


Abb. 4. Illustration des Ansatzes aus [11]

4.2 Smirnov et al.: Meronymy-Based Aggregation of Activities in Business Process Models [13]

Die Anwendung dieses Ansatzes benötigt zunächst die Einordnung der einzelnen Funktionen der Modelle in einem kontrollierten Vokabular, das bestimmt, aus welchen Funktionen sich eine einzelne Funktion zusammensetzt. Dies ist hier durch einen Funktionsbaum (in Abbildung 5) dargestellt.

Auf dieser Grundlage kann nun mithilfe der im Artikel beschriebenen Metrik berechnet werden, welche Menge von Funktionen in eine übergeordnete Funktion aggregiert werden sollen. Dafür berechnet man, wie gut eine Menge von Knoten einen übergeordneten Knoten repräsentiert, wobei die Knotenmenge mit der besten Bewertung in ihren gemeinsamen Vorgängerknoten aggregiert wird. In Modell C ist das der Wurzelknoten, der einen Abdeckungsgrad von 0,6 aufweist. Damit ist der Abstraktionsgrad dieses Modells und damit der des Referenzmodelles sofort sehr hoch, die

anderen Modelle können nur noch daran angepasst werden. Dies geschieht sowohl in Modell A als auch in Modell B durch jeweils zwei Abstraktionsschritte, die die Bewertung 1 bzw. 0,6 aufweisen. Im Modell A werden zunächst „Rechnung gegen Wareneingang prüfen“ und „Rechnung preislich prüfen“ zu „Rechnung prüfen“, in Modell B „Kreditorenbuchhaltung“ und „Fakturierung“ zu „Rechnung verbuchen“ aggregiert. Danach werden in beiden Modellen die verbliebenen Funktionen zum Wurzelknoten „Rechnungsprüfung“ aggregiert. Damit bestehen alle Modelle nur noch aus dieser sehr generischen Funktion und den entsprechenden Startereignissen.

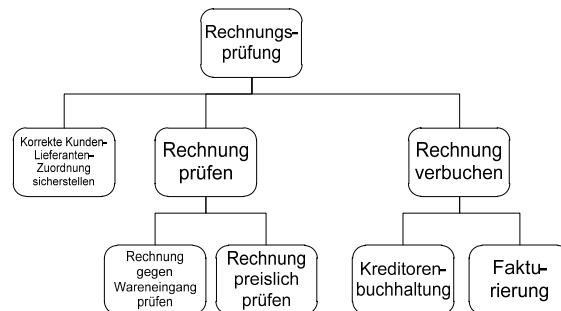


Abb. 5. Funktionsbaum der Modelle A, B und C

Aufgrund des klar definierten Abstraktionskriteriums ist dieser Ansatz am besten zu automatisieren. Allerdings ist dadurch auch der Abstraktionsgrad der Modelle nur schwer zu kontrollieren. Man könnte hier eventuell das Signifikanzkriterium anpassen, sodass nicht immer die Operation mit der besten höchsten Bewertung durchgeführt werden muss. Gleichzeitig ist hier das Ziel, semantisch sinnvoll zu abstrahieren. Dadurch, dass das kontrollierte Vokabular nur auf Funktionen definiert ist, bleiben Ereignisse entweder so bestehen oder werden gemäß der Funktion ausgetauscht, sodass sie an Relevanz verlieren. Außerdem stellt das geforderte kontrollierte Vokabular einen erheblichen Mehraufwand dar, der vor allem bei großen Modellmengen nicht realisierbar ist. Liegt es vor, ist die Abstraktion einfach zu berechnen. Auch die Bezeichnung des neu aggregierten Prozesselements ist durch das kontrollierte Vokabular gegeben.

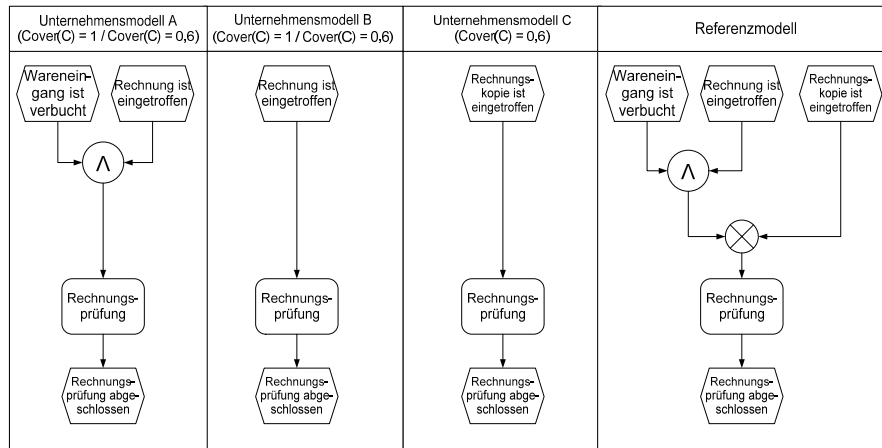


Abb. 6. Illustration des Ansatzes aus [13]

4.3 Günther und van der Aalst: Fuzzy Mining - Adaptive Process Simplification Based on Multi-perspective Metrics [4]

Die hier vorausgesetzte relative Häufigkeit von Knoten und Kanten wird als der Anteil der Modelle definiert, in dem der Knoten bzw. die Kante zwischen zwei Knoten auftritt. Weiterhin sind ein Signifikanz- und ein Korrelationsniveau zu definieren. Um die Vorgehensweise hier anschaulich zu gestalten, wurden beide Zahlen auf 50% gesetzt, sodass alle Objekte, die nur in einem Modell auftreten, abstrahiert werden sollen. Da hier auch das Korrelationsniveau bei 50% liegt, ist keins der Objekte stark mit einem anderen korreliert, alle Objekte werden also aus dem Modell eliminiert. Dies sind drei Knoten im Modell A („Wareneingang ist verbucht“, „Rechnung gegen Wareneingang prüfen“ sowie „Rechnung ist gegen Wareneingang geprüft“). Da alle Knoten aus Modell B in mindestens zwei Modellen auftreten, wird keiner von ihnen abstrahiert. Aus Modell C dagegen werden zwei Knoten („Korrekte Kunden-Lieferanten-Zuordnung sicherstellen“ und „Korrekte Kunden-Lieferanten-Zuordnung ist sichergestellt“) eliminiert.

Ein Vorteil dieser Vorgehensweise ist, dass der Anwender durch die Definition der Niveauewerte entscheiden kann, ob aggregiert oder eliminiert werden soll, während die Operation bei allen anderen Ansätzen vorgegeben war. Bei der hier gewählten relativen Häufigkeit kommt in einem Schritt allerdings nur eine Operation zur Geltung, da eine Kante nicht häufiger sein kann als ihre zugehörigen Knoten. Da die Abstraktionskriterien frei wählbar sind, kann dies aber anders gestaltet werden. Tatsächlich ist dieser Ansatz von allen vorgestellten am flexibelsten: Es werden zwar Kriterien definiert, die konkrete Auswahl bleibt aber dem Nutzer überlassen.

5 Überblick und Diskussion der Ergebnisse

Zum abschließenden Vergleich der Abstraktionsansätze werden die Ergebnisse der Evaluation in Tabelle 1 präsentiert und die drei resultierenden Referenzprozessmodelle in Abbildung 8 dargestellt. Es fällt auf, dass sie sich nicht nur im Abstraktionsgrad, sondern auch in der Modellstruktur erheblich unterscheiden. Dies liegt zum einen an den verschiedenen angelegten Abstraktionskriterien, zum anderen an den unterschiedlichen angewendeten Abstraktionsoperationen. Ausschließlich der Ansatz aus [4] erlaubt die Eliminierung von Startereignissen, in den anderen Ansätzen bleiben diese erhalten und müssen kombiniert werden. Dies steht nicht immer im Einklang mit dem Abstraktionsgrad des restlichen Modells, was besonders beim Modell aus [13] zu sehen ist. Da dort das Abstraktionskriterium ausschließlich Funktionen betrachtet, müssen die Ereignisse manuell angepasst werden. Die strenge Definition von Kriterium und Signifikanzfunktion bestimmt hier den Abstraktionsgrad, der daher nur schwer kontrolliert werden kann. Dieser ist in den beiden anderen Modellen niedriger gewählt worden, könnte aber weiter erhöht werden.

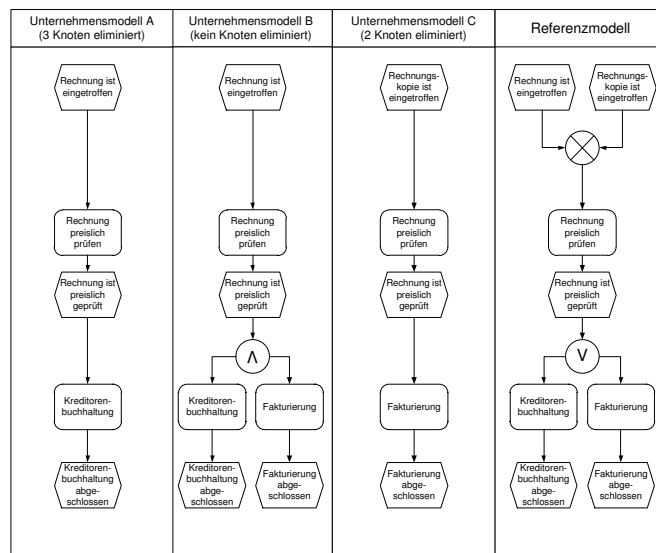


Abb. 7. Illustration des Ansatzes aus [4]

Dagegen decken nur die Modelle, die mithilfe der Konzepte aus [11] und [13] entstehen, den ganzen Prozess ab, was in der angewendeten Operation begründet ist. Hier wurde ausschließlich die Aggregation verwendet, sodass nicht die Granularität sondern eben der Grad der Abdeckung modifiziert wird. Betrachtet man die oben definierten Kriterien zur Bewertung der Abstraktionsansätze, so stellt man fest, dass die Identifizierung gemeinsamer Strukturen bei Polyvyanyy et. al. [11] vor allem manuell durch die Benennung der aggregierten Funktionen und die Identifizierung von Teilprozessen geschieht. Ansonsten werden die Knoten nur durch ihre Labels miteinander

identifiziert, hier könnte eine vorherige manuelle Zuordnung Abhilfe schaffen. Dies würde vielleicht auch die Benennung der Funktionen vereinfachen, da auf sprachliche Aspekte ansonsten nicht eingegangen wird. Dagegen können Unternehmensspezifika der einzelnen Modelle hier gezielt abstrahiert werden, da Kriterium und Operation relativ frei gewählt werden können.

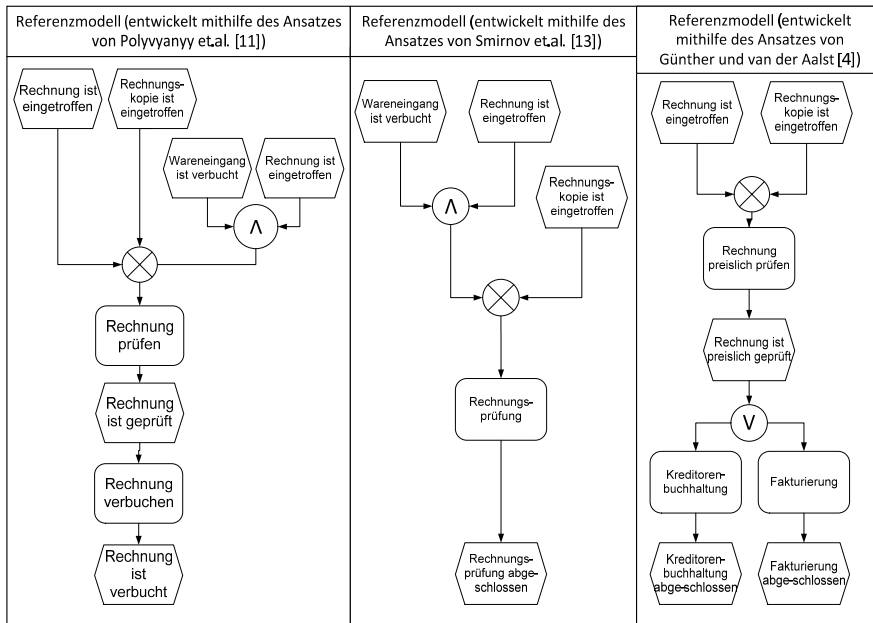


Abb. 8. Vergleich der verschiedenen Referenzmodelle

Bei Smirnov et. al. [13] dagegen ist die Identifikation von Gemeinsamkeiten durch das kontrollierte Vokabular (im Original *ontology*, hier realisiert als Funktionsbaum), die alle in den Modellen auftretenden Funktionen identifiziert und miteinander in Relation setzt, möglich. Diese ermöglicht auch den Umgang mit den entsprechenden Sprachkonstrukten. Insgesamt ist die Existenz eines solchen kontrollierten Vokabulars natürlich nicht immer zu garantieren und nimmt einen Teil der Arbeit auch schon vorweg.

Günther und van der Aalst beschreiben einen sehr allgemein gehaltenen Ansatz, wo viel über die Wahl der Abstraktionskriterien konfiguriert werden kann. Allerdings beziehen auch sie nicht mit ein, wie man mit der Verwendung natürlicher Sprache, wie etwa in Knotenlabels, umgeht. Durch die freie Wahl der Abstraktionskriterien können aber unternehmensspezifische Besonderheiten gezielt ausgeblendet werden. Die Identifizierung der Besonderheiten, die wie bei den anderen Ansätzen auch, insbesondere beim Zusammenfügen der Modelle wichtig ist, erfolgt auch hier nur über gleiche Knotenlabels.

Tabelle 1. Vergleich von acht Abstraktionsansätzen

Kriterien	Ansätze													
Prinzip		Syntaktisch	Syntaktisch	Strukturell	Strukturell	Semantisch	Semantisch	Semantisch	Semantisch	„Hybrid“				
Nutzero.		o	-	+	-	+	o	o	o	+				
Übergeordnetes Ziel		Komplexitätsreduzierung	Aufdeckung struktureller Konflikte	Verwaltung verschiedener Modelle	Generierung von Prozesssichten	Verwaltung verschiedener Modelle	-	Generierung von Prozesssichten	Verdichtung von Modellen					
Abstraktionsobjekte		Prozessfragmente	Prozessfragmente	SESE-Regionen	Knotenmengen	Knotencluster	Knotenmengen	Prozessfragmente	Knoten & Kanten					
Abstraktionskriterien		Transitions-wahrsch.	Strukturelle Korrektheit	-	-	Ähnlichkeit des Clusters	Grad der Abdeckung	-	Signifikanz & Korrelation					
Signifikanzfunktion		Schwellenwerte	Existenzfunktion	-	-	Schwellenwerte	Höchste Abdeckung	-	Schwellenwerte					
Abstr.-Oper.		Aggr. & Elim.	Elim.	Aggr.	Aggr.	Aggr.	Aggr.	Aggr. & Elim.	Aggr. & Elim.					
Anpassung		o	+	+	o	-	+	+	-					
Benötigte Informationen		Erweiterte EPKs	Korrekte Strukturen	Prozessstrukturbaum	Umsetzung der Regeln	Zusätzliche Attribute	kontrolliert. Vokabular	-	Relative Häufigkeiten					
Mögliche Ergänzungen		Semantische Kriterien	Kontrolle des Abstr.-grades	Abstraktionskriterien	Abstraktionskriterien	-	Konkrete Anwendung	Abstraktionskriterien	-					
Besond.		+	o	o	+	+	o	+	+					
Gemein.		-	-	o	-	o	+	o	-					
Sprache		-	-	-	-	o	o	o	-					

6 Schlussfolgerung

Insgesamt hat jeder der acht hier evaluierten Ansätze ein bezüglich des betrachteten Prozesses sinnvolles Referenzprozessmodell ergeben, sodass sie alle weiter verfolgt werden könnten. Allerdings wiesen einige Ansätze Schwächen wie mangelnde Flexibilität oder die nicht spezifizierte Benennung neu aggregierter Knoten auf. Gerade

letztere ist notwendige Voraussetzung um Abstraktion zu automatisieren, und kann Grundlage für weitere Betrachtungen sein. Weiterhin wäre eine Studie zur konkreten Anwendung der Algorithmen auf weit verbreitete Modellierungssprachen ebenso interessant wie hilfreich, um die praktische Umsetzbarkeit der Ansätze zu beurteilen.

Durch die im Rahmen dieser Evaluation gewonnenen Erkenntnisse können folgende allgemeine Anforderungen an Abstraktionsansätze, die zur Entwicklung von Referenzprozessmodellen verwendet werden sollen, formuliert werden:

- Ausnutzung der verfügbaren Daten: Häufig sind in Prozessmodellen zusätzliche Informationen gespeichert, beispielsweise in Form von Input- und Outputinformationen, die im Zuge der Abstraktion soweit wie möglich ausgenutzt werden sollten.
- Einbeziehung natürlicher Sprache: Zur Entwicklung von Referenzprozessmodellen ist es unabdingbar, natürliche Sprache zu betrachten.
- Unabhängigkeit vom Zusammenfügen: Hier wird das Referenzmodell im Zuge des Zusammenfügens, beispielsweise durch linguistische Abstraktion, verfälscht. Daher sollte ein Abstraktionsansatz möglichst unabhängig vom Zusammenfügen sein.
- Hierarchisierung: Insbesondere bei der Aggregation besteht ein Zusammenhang der Abstraktion mit hierarchisierten Prozessmodellen. Diese Beziehungen zwischen Knoten sollte ein solcher Ansatz ausnutzen.

Erkenntnistheoretisch ist eine induktive Herangehensweise kritisch zu betrachten. Wenn ausgehend von einzelnen Sachverhalten auf allgemeine Gesetzmäßigkeiten geschlossen wird, besteht das Risiko eines Fehlschlusses. Allerdings stellt sich die Frage, inwieweit dies bei der Modellanalyse zu einem Problem wird. Im Kontext der induktiven Referenzmodellierung werden die (automatisiert) gewonnen Modelle nicht vorbehaltlos übernommen und verwendet, sondern kritisch betrachtet und überarbeitet, um sinnvolle Referenzmodelle sicherzustellen.

Abschließend lässt sich die anfangs gestellte Frage zur Eignung von Abstraktionsansätzen für die induktive Entwicklung von Referenzmodellen grundsätzlich positiv beantworten, wobei keiner der Ansätze alle notwendigen Kriterien erfüllt. An dieser Stelle könnte die Entwicklung eines neuen Abstraktionsansatzes speziell zum Zwecke der induktiven Entwicklung von Referenzmodellen gemäß den oben identifizierten Kriterien eine Lösung sein.

Danksagung. Die hier dargestellten Ergebnisse stammen unter anderem aus dem Projekt „Konzeptionelle, methodische und technische Grundlagen zur induktiven Erstellung von Referenzmodellen (Reference Model Mining)“, gefördert durch die DFG (GZ LO 752/5-1).

Literatur

1. Becker, J., Delfmann, P., Knackstedt, R., Kuropka, D.: Konfigurative Referenzmodellierung. In: Becker, J., Knackstedt, R. (eds.): Wissensmanagement mit Referenzmodellen. Konzepte für die Anwendungssystem- und Organisationsgestaltung. Springer, Berlin et al. (2002)
2. Bobrik, R., Reichert, M., Bauer, T.: View-Based Process Visualization. Alonso, G., Dadam, P., Rosemann, M. (eds.): BPM 2007. LNCS, Vol. 4714, pp. 88-95. Springer, Berlin Heidelberg (2007)
3. Chiu, D.K.W., Cheung, S.C., Till, S., Karlapalem, K., Li, Q., Kafeza, E.: Workflow view driven crossorganizational interoperability in a web service environment. *Information Technology and Management* 5 (3-4), 221-250 (2004)
4. Günther, C.W., van der Aalst, W.M.P.: Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. In: Alonso, G., Dadam, P., Rosemann, M. (Hg.): BPM 2007. LNCS, Vol. 4714, pp. 328-343. Springer, Berlin, Heidelberg (2007)
5. Keller, G., Nüttgens, M., Scheer, A.: Semantische Prozessmodellierung auf der Grundlage "Ereignisgesteuerter Prozessketten (EPK)". Technical Report Heft 89, Veröffentlichungen des Instituts für Wirtschaftsinformatik Universität des Saarlandes (1992)
6. Li, C., Reichert, M., Wombacher, A.: Discovering Reference Models by Mining Process Variants Using a Heuristic Approach. In: BPM'09. LNCS, Vol. 5701, pp. 344-362. Springer (2009)
7. Liu, D., Shen, M.: Workflow modeling for virtual processes: an order-preserving process-view approach. *Information Systems* 28 (6), 505-532 (2003)
8. La Rosa, M., Dumas, M., Uba, R., Dijkman, R.: Merging Business Process Models. In: Meersman, R., Dillon, T.S., Herrero, P. (eds.): Proceedings of the International Conference on Cooperative Information Systems. LNCS, Vol. 6426, pp. 96-113. Springer (2010)
9. Polyvyanyy, A., Smirnov, S., Weske, M.: On Application of Structural Decomposition for Process Model Abstraction. In: Proceedings of the 2nd International Conference on Business Process and Services Computing, pp. 110-122. Leipzig (2009)
10. Polyvyanyy, A., Smirnov, S., Weske, M.: Process Model Abstraction: A Slider Approach. In: EDOC 2008, pp. 325-331 (2008)
11. Polyvyanyy, A., Smirnov, S., Weske, M.: Reducing Complexity of Large EPCs. In: EPK'08 GI-Workshop, Saarbrücken. LNI, Vol. P-141, pp. 195-208. GI, Bonn (2008)
12. Sadiq, W., Orłowska, M.E.: Analyzing Process Models Using Graph Reduction Techniques. *Information Systems* 25 (2), 117-134 (2000)
13. Smirnov, S., Dijkman, R., Mendling, J., Weske, M.: Meronymy-Based Aggregation of Activities in Business Process Models. In: Parsons, J. et al. (Hg.): ER 2010. LNCS, Vol. 6412, pp. 1-14. Springer, Berlin Heidelberg (2010)
14. Smirnov, S., Reijers, H., Nugteren, T., Weske, M.: Business Process Model Abstraction: Theory and Practice. Technical Report 35, Hasso-Plattner-Institut, <http://bpt.hpi.unipotsdam.de/pub/Public/SergeySmirnov/abstractionUseCases.pdf> (2010)
15. Smirnov, S., Reijers, H., Weske, M.: A Semantic Approach for Business Process Model Abstraction. In: Mouratidis, H., Rolland, C. (eds.): CAiSE 2011. LNCS, Vol. 6741, pp. 497-511, Springer, Berlin (2011)
16. Smirnov, S., Reijers, H., Weske, M., Nugteren, T.: Business process model abstraction: a definition, catalog, and survey. *Distributed and Parallel Databases* 30 (1), 63-99 (2012)

Spezifikation von funktionalen und nichtfunktionalen Systemanforderungen auf Basis von Geschäftsprozessmodellen

Beate Hartmann, Andree Teusch, und Matthias Wolf

Universität Bamberg, Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung und Datenbankanwendung, Bamberg, Germany
{beate.hartmann, andree.teusch, matthias.wolf}@uni-bamberg.de

Abstract. Anwendungssysteme unterstützen die Durchführung von Geschäftsprozessen. Daher ist eine enge Abstimmung von dem zu entwickelnden Anwendungssystem mit den unterstützten Geschäftsprozessen notwendig. In diesem Beitrag wird gezeigt, wie sich diese Abstimmung über die Ableitung von funktionalen und nichtfunktionalen Anforderungen an die Anwendungssysteme aus Geschäftsprozessmodellen realisieren lässt. Den gängigen Geschäftsprozessmodellierungssprachen liegt hierzu i. d. R. aufgrund ihres heterogenen Begriffs- und Modellverständnisses kein einheitliches Vorgehen zugrunde. Zur Beschreibung von Geschäftsprozessen wird daher ein allgemeingültiges Aufgabenkonzept verwendet, auf dessen Basis die Spezifikation von Anforderungen allgemein aus Geschäftsprozessmodellen erfolgen kann. Anschließend werden die Elemente von drei gängigen Geschäftsprozessmodellierungssprachen diesem Aufgabenkonzept zugeordnet, um damit konkret angegeben zu können, wie sich Anforderungen aus diesen Modellierungssprachen ableiten lassen.

Keywords: Anforderungsspezifikation, Geschäftsprozessmodellierung, Aufgabenmodell

1 Einleitung

Zur Unterstützung von Geschäftsprozessen werden Anwendungssysteme entwickelt. Ein bedeutender Erfolgsfaktor in der Systementwicklung und ein Schwerpunkt des IT/Business-Alignments ist dabei die enge Abstimmung von Zielen sowie zugehöriger Geschäftsprozesse eines Unternehmens mit dem zu realisierenden Anwendungssystem [1-2]. Neben der Dokumentation von Geschäftszielen und Aufgaben haben sich Geschäftsprozessmodelle bei der Spezifikation fachlicher Systemanforderungen (im Folgenden als Anforderungen bezeichnet), aber auch bei der Ableitung detaillierter Spezifikationen im Rahmen der modellgetriebenen Systementwicklung bewährt (z. B. [3]).

Ausgangspunkt für die Spezifikation von Anforderungen auf Basis von Geschäftsprozessmodellen bilden die dort modellierten Aufgaben. Eine Aufgabe beschreibt hierbei, was das System leisten soll. Dies wird in der Anforderungsspezifikation in

funktionalen Anforderungen formuliert. Neben diesen funktionalen Anforderungen werden in der Systementwicklung auch sogenannte nichtfunktionale Anforderungen an das zu realisierende Anwendungssystem gestellt, welche seine Funktionalität beschränken.

Bisher dienen Geschäftsprozessmodelle primär der Dokumentation und Analyse von funktionalen Anforderungen an das System. In den letzten Jahren wurden zur Modellierung nichtfunktionaler Anforderungen in Geschäftsprozessmodellen verschiedene Vorschläge gemacht (z. B. [4-6]). Dabei wurden individuelle Vorschläge für einzelne Modellierungssprachen vorgestellt. Eine Untersuchung der Spezifikation von Anforderungen auf Basis von Geschäftsprozessmodellen im Allgemeinen erfolgte bisher nicht. Für solch eine Untersuchung ist eine nähere Betrachtung des zentralen Elements eines Geschäftsprozesses, der Aufgabe, als Quelle der Systemanforderungsspezifikation notwendig. Dies soll in diesem Beitrag aufgegriffen werden. Ziel des vorliegenden Beitrags ist es daher, die Aufgabe als Kernelement von Geschäftsprozessen bezüglich ihrer Eignung als Basis für die Anforderungsspezifikation zu untersuchen und darauf aufbauend Anforderungen aus gängigen Geschäftsprozessmodellierungssprachen zu spezifizieren. Für diese Untersuchung wird ein Aufgabenkonzept herangezogen, mithilfe dessen die Spezifikation von Anforderungen aus Geschäftsprozessmodellen allgemein betrachtet werden kann. Anschließend erfolgt eine Übertragung des Konzepts auf gängige Geschäftsprozessmodellierungssprachen.

Im folgenden Kapitel 2 werden zunächst Grundlagen zu funktionalen und nicht-funktionalen Anforderungen gegeben, bevor das Konzept der betrieblichen Aufgabe und die auf diesem Konzept basierende Ableitung von Anforderungen beschrieben werden. In Kapitel 3 folgt die Übertragung auf gängige Modellierungssprachen für Geschäftsprozesse. Kapitel 4 fasst die Erkenntnisse in einer Diskussion zusammen und es wird ein Ausblick gegeben.

2 Spezifikation von Anforderungen auf Basis von Geschäftsprozessmodellen

Geschäftsprozesse beschreiben betriebliche Abläufe aus fachlicher Sicht. Anwendungssysteme führen die automatisierten Aufgaben der Geschäftsprozesse durch. Somit bilden Geschäftsprozessmodelle die Grundlage für die Definition von Anforderungen an diese Systeme.

2.1 Funktionale und nichtfunktionale Anforderungen

In der Systementwicklung beschreiben Anforderungen Dienste und Bedingungen, die ein System oder eine Systemkomponente leisten und erfüllen soll, um ein Problem zu lösen oder ein Ziel zu erreichen (nach [7-8]). Hierbei wird häufig in der Literatur zwischen funktionalen und nichtfunktionalen Anforderungen unterschieden ([8-9]).

Funktionale Anforderungen beschreiben Dienste oder Funktionalitäten, die ein System bereitstellen soll. Diese Beschreibung kann in allgemeiner Form Funktionen

des Systems definieren oder als Spezifikation detaillierte Aussagen zum gewünschten Verhalten, Datenstrukturen oder Funktionen des Systems festhalten.

Nichtfunktionale Anforderungen definieren dagegen Beschränkungen der Dienste oder Funktionalität des Systems und beziehen sich häufig nicht auf eine Funktionalität sondern auf Teile des Systems oder das gesamte System. Für nichtfunktionale Anforderungen werden i. d. R. drei Klassen unterschieden:

- **Qualitätsanforderungen** an die Dienstleistung, z. B. bezüglich Zuverlässigkeit oder Geschwindigkeit,
- **Rahmenbedingungen**, welche das System oder die Systementwicklung einschränken, z. B. rechtliche Vorgaben oder Unternehmensanforderungen an die Umgebung des Systems oder die einzusetzende Entwicklungsplattform,
- **unterspezifizierte funktionale Anforderungen**, die noch nicht detailliert spezifiziert wurden, z. B. ein sicheres, einfach bedienbares System.

Qualitätsanforderungen gelten für das gesamte System oder für einzelne Teile des Systems. Rahmenbedingungen gelten häufig für das gesamte Unternehmen und werden daher selten in Geschäftsprozessmodellen dokumentiert. Unterspezifizierte funktionale Anforderungen gelten zunächst allgemein für das gesamte System bis diese detailliert als funktionale und/oder nichtfunktionale Qualitätsanforderungen formuliert sind. Im Folgenden wird diese Unterteilung der nichtfunktionalen Anforderungen nur aufgegriffen, falls es explizit notwendig erscheint.

2.2 Ableitung von Anforderungen aus Geschäftsprozessmodellen

Der Begriff Geschäftsprozess ist in der Literatur nicht einheitlich definiert. In der Regel wird unter einem Geschäftsprozess eine zielgerichtete, zeitlich-logische Abfolge von Aufgaben, welche der Leistungserstellung dient, verstanden [10]. Geschäftsprozesse beschreiben demnach betriebliche Abläufe und werden im Allgemeinen von der Fachabteilung in Ausrichtung auf die Unternehmensziele modelliert. Geschäftsprozessmodelle sind abstrakte Repräsentationen von Geschäftsprozessen.

In Geschäftsprozessmodellierungssprachen werden Begriffe wie *Aufgabe*, *Funktion*, *Aktivität* - teilweise synonym - als zentrales Element eines Geschäftsprozesses verwendet. Da der Begriff Aufgabe ebenfalls ein zentrales Element der Betriebswirtschaftslehre darstellt und Geschäftsprozesse eine betriebswirtschaftliche Perspektive auf ein Unternehmen einnehmen, wird im Folgenden der Begriff *Aufgabe* verwendet. Der betriebswirtschaftliche Begriff Aufgabe wurde von Kosiol geprägt [11]. Er definiert eine Aufgabe als eine zielorientierte Verrichtung an einem Aufgabenobjekt. Ferstl und Sinz [12] definieren ein auf Kosiol aufbauendes Aufgabenkonzept, mithilfe dessen Geschäftsprozesse spezifiziert werden können. Auf Basis dieses Konzepts werden im weiteren Verlauf der Arbeit funktionale und nichtfunktionale Anforderungen an Anwendungssysteme abgeleitet. Nach diesem Konzept lässt sich eine betriebliche Aufgabe in einer Außensicht und einer Innensicht beschreiben. Das zugehörige Aufgabenmodell ist in Abbildung 1 dargestellt. Bei definierten Vorereignissen wird eine Verrichtung an dem Aufgabenobjekt der Aufgabe entsprechend der vorgegebenen Sach- und Formalziele durchgeführt und ein Nachereignis produziert.

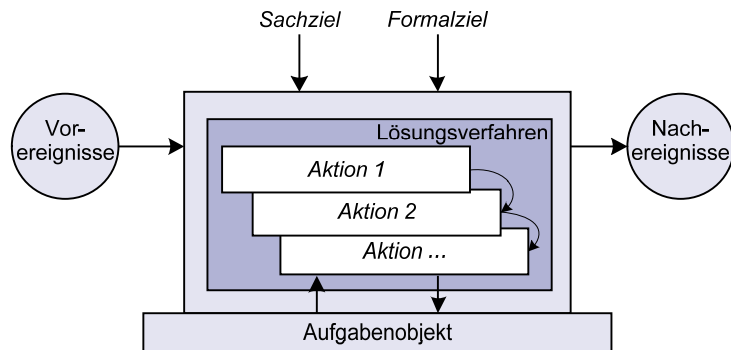


Abb. 1. Aufgabenmodell (nach [12])

Außensicht einer Aufgabe. Die Außensicht einer Aufgabe definiert das Aufgabenobjekt, die Ziele der Aufgabe sowie Vor- und Nachereignisse. Es wird noch kein Verfahren der Durchführung (Lösungsverfahren) angegeben. Durch das Sachziel der Aufgabe werden die Fragen „Was soll die Aufgabe tun?“ bzw. „Welche Nachzustände sollen erreicht werden?“ beantwortet. Dieses Verständnis trifft ebenso auf funktionale Anforderungen zu. Somit lassen sich funktionale Anforderungen aus den Sachzielen einer Aufgabe ableiten.

Wie die Aufgabe durchzuführen ist, d. h. welches Lösungsverfahren zu wählen ist, geben Formalziele vor. Sie beschränken somit die Aufgabenverrichtung. Damit bieten Formalziele die Grundlage um nichtfunktionale Anforderungen abzuleiten. Formalziele können sowohl ökonomische als auch soziale, technische und ökologische Ziele sein [13]. Im Gegensatz zu Sachzielen beziehen sich Formalziele häufig nicht auf genau eine Aufgabe sondern auf mehrere Aufgaben, die nicht zwingend hintereinander ausgeführt werden müssen, oder auf das gesamte Unternehmen (diese bilden dann Rahmenbedingungen). So muss z. B. das Formalziel „Höchster Sicherheitsstandard bei der Verschlüsselung von sensiblen Daten“ bei allen Aufgaben berücksichtigt werden, welche diese Daten manipulieren. Dieses Formalziel ist somit eine nichtfunktionale Anforderung an das gesamte System.

Innensicht einer Aufgabe. In der Innensicht einer Aufgabe beschreibt das Lösungsverfahren wie die Sachziele - unter Berücksichtigung der Formalziele - zu erreichen sind. Es besteht aus einer Menge von Aktionen, deren Reihenfolge durch eine Aktionssteuerung (in Abbildung 1 nicht dargestellt) bestimmt wird. Das Lösungsverfahren wird in Abhängigkeit des Aufgabenträgertyps (Mensch, Maschine) beschrieben. Eine Aktion, die funktional beschreibbar ist, kann von einem Anwendungssystem durchgeführt werden. Aus solchen Aktionen können ebenfalls direkt funktionale Anforderungen abgeleitet werden. Diese sind somit als Detaillierung der aus dem Sachziel der Aufgabe abgeleiteten funktionalen Anforderung zu verstehen. Personell durchzuführen sind Aktionen, die nicht funktional beschreibbar sind. Aus diesen Aktionen können Stellenbeschreibungen abgeleitet werden. Nichtfunktionale Aktionen werden im Rahmen dieser Arbeit nicht weiter betrachtet.

Zerlegung von Aufgaben. Geschäftsprozessmodelle können sukzessive verfeinert werden. Dies geht mit einer Aufgabenzerlegung einher und dient der Komplexitätsreduktion sowie der zielspezifischen Detaillierungsmöglichkeit. Die Zerlegungstiefe wird dabei vom Modellierer in Abhängigkeit des Modellierungszwecks bestimmt. Geschäftsprozessmodelle zur Beschreibung von fachlichen Abläufen werden in der Regel weniger tief zerlegt, als solche Modelle, die als Grundlage für eine Workflow-Ausführung¹ vorgesehen sind. Bei der Aufgabenzerlegung können gemäß dem Verichtungs- oder Objektprinzip das Aufgabenobjekt oder das Lösungsverfahren zerlegt werden. Sachziele werden folglich ebenso zerlegt und weiter detailliert. Damit werden mit jeder Aufgabenzerlegung die funktionalen Anforderungen spezifischer. Im Gegensatz dazu entspricht die Zerlegung von Formalzielen i. d. R. nicht der Zerlegung von Aufgaben, da diese zumeist für mehrere Aufgaben gelten. Auch nichtfunktionale Anforderungen beziehen sich häufig auf das gesamte System und sind nicht einzelnen funktionalen Anforderungen zuzuordnen. Formalziele können in einem Zielsystemmodell [14] visualisiert werden. Dabei wird ein übergeordnetes Ziel sukzessive in Teilziele zerlegt. Anschließend kann eine Zuordnung der (Teil-) Formalziele zu korrespondierenden (Teil-) Aufgaben vorgenommen werden. Ist eine Zuordnung zu konkreten Aufgaben nicht möglich, so lassen sich allgemein zu beachtende nichtfunktionale Anforderungen, z. B. Rahmenbedingungen, ableiten.

Die folgende Tabelle 1 fasst zusammen, welche Elemente des Aufgabenmodells und damit eines Geschäftsprozessmodells für die Spezifikation von Systemanforderungen herangezogen werden können. Es sei ausdrücklich betont, dass die Ableitung von Anforderungen aus (formalen) Geschäftsprozessmodellen einen wichtigen Beitrag für die Anforderungsspezifikation von Systemen leisten kann. Dennoch müssen diese Anforderungen ggf. weiter präzisiert bzw. ergänzt werden, da sicherlich nicht alle Anforderungen an ein zu entwickelndes System Geschäftsprozessmodellen zu entnehmen ist, z. B. detailliertes Ausnahmeverhalten. Außerdem werden einige nichtfunktionale Anforderungen, z. B. an die Usability, selten als Formalziele in Geschäftsprozessmodellen beschrieben.

¹ Auf die Abgrenzung von Geschäftsprozessen und Workflows wird in Kapitel 3 eingegangen.

Tabelle 1. Spezifikation von Anforderungen auf Grundlage des Aufgabenmodells

Art der Anforderung	Spezifikation aus dem Aufgabenmodell
Funktional	<ul style="list-style-type: none"> • Sachziel: gibt direkt eine funktionale Anforderung an ein System vor. • Aktion: Bezeichnung der Aktion ist zumeist eine funktionale Anforderung. Diese Anforderungen konkretisieren die aus dem Sachziel abgeleiteten Anforderungen. • Ereignisse: stellen definierte Zustände des zu entwickelnden Systems dar, auf die innerhalb der Anforderungsspezifikation Bezug genommen werden kann. • Aufgabenobjekt: Grundlage der Datenspezifikation, sofern diese Teil der Anforderungsspezifikation sein soll.
Nichtfunktional	<ul style="list-style-type: none"> • Formalziel: kann nichtfunktionale Anforderungen vorgeben. Gilt häufig für das gesamte System.

Die bisherigen Erkenntnisse sollen anhand eines konkreten Fallbeispiels veranschaulicht werden. Die Grundlage hierfür bildet ein Geschäftsprozess eines Sicherheitsdienstleisters, welcher automatisierte Einlasskontrollprüfungen von Veranstaltungen für seine Kunden abwickelt. Abbildung 2 zeigt die zu erbringende Gesamtaufgabe.

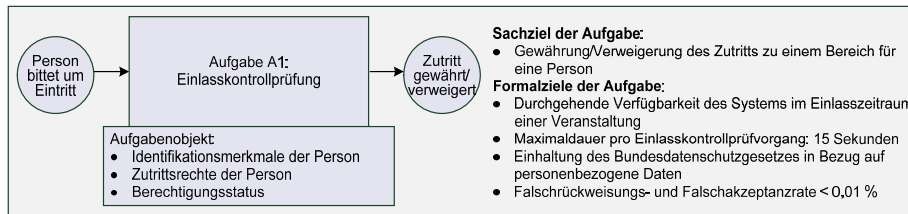


Abb. 2. Aufgabe „Einlasskontrollprüfung“

Die Zerlegung der Aufgabe Einlasskontrollprüfung in zwei Teilaufgaben mit Angabe des Lösungsverfahrens ist in Abbildung 3 dargestellt.

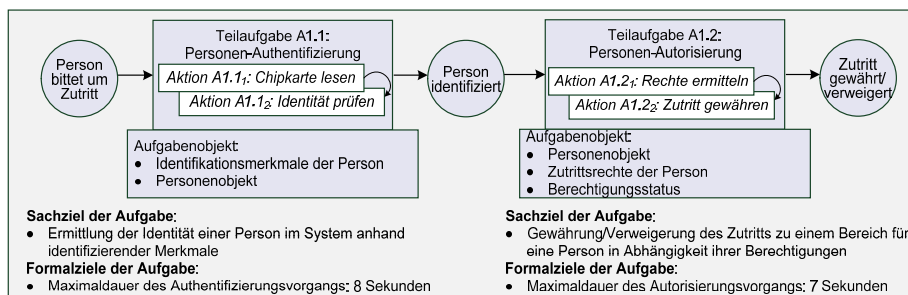


Abb. 3. Zerlegung der Aufgabe „Einlasskontrollprüfung“ und Angabe des Lösungsverfahrens

Für ein System zur Durchführung der Einlasskontrollprüfung lässt sich gemäß Tabelle 1 folgende initiale Anforderungsspezifikation anhand des Aufgabenmodells ableiten:

Tabelle 2. Initiale Anforderungsspezifikation für ein Anwendungssystem zur Unterstützung der Aufgabe „Einlasskontrollprüfung“

Funktionale Anforderungen (F)

/F10/	Das System muss die Identität einer Person anhand vorgegebener Merkmale ermitteln.	Sachziel A1.1
/F11/	Das System muss eine Chipkarte einlesen können.	Aktion A1.1 ₁
/F12/	Das System muss die Identität im System prüfen können.	Aktion A1.1 ₂
/F20/	Das System muss den Zugriff zu einem Bereich in Abhängigkeit von den einer Person zugewiesenen Berechtigungen gewähren bzw. verweigern.	Sachziel A1.2
/F21/	Das System muss die Rechte einer Person ermitteln können.	Aktion A1.2 ₁
/F22/	/F22/ Das System muss den Zutritt freischalten können.	Aktion A1.2 ₂
/F30/	Das System muss folgende Personendaten permanent speichern: Personen-ID, Vorname, Nachname, Geburtsdatum, Adresse	Aufgabenobjekt A1
/F40/	Das System muss folgende Berechtigungszuordnungsdaten permanent speichern: Personen-ID, Bereichs-ID, Bereichs-Name	Aufgabenobjekt A1

Nichtfunktionale Anforderungen (NF)

/NF10/	Das System muss durchgehend im Einlasszeitraum einer Veranstaltung zur Verfügung stehen.	Formalziel A1
/NF20/	Das Bundesdatenschutzgesetz in Bezug auf personenbezogene Daten ist einzuhalten.	Formalziel A1
/NF30/	Die Falschrückweisungs- und Falschakzeptanzrate beträgt < 0,01 %.	Formalziel A1
/NF40/	Der Autorisierungsvorgang darf maximal 7 Sekunden dauern.	Formalziel A1.2
/NF50/	Der Authentifizierungsvorgang darf maximal 8 Sekunden dauern.	Formalziel A1.1

3 Spezifikation von Anforderungen am Beispiel ausgewählter Geschäftsprozessmodellierungssprachen

Das vorgestellte Aufgabenkonzept dient im Folgenden als Grundlage zur Spezifikation von Anforderungen am Beispiel von drei Geschäftsprozessmodellierungssprachen. Es wird untersucht, inwiefern die Elemente des Aufgabenmodells in den Modellierungssprachen wiederzufinden sind und somit bestimmt werden kann, wie Anforder-

rungen aus den Geschäftsprozessmodellen abgeleitet werden können. Dabei werden auch in der Literatur genannte Erweiterungen der Sprachen berücksichtigt.

Untersucht werden die weit verbreiteten Geschäftsprozessmodellierungssprachen Business Process Model and Notation (BPMN) und ereignisgesteuerte Prozesskette (EPK). Ebenso wird die Unified Modeling Language (UML) betrachtet, obwohl deren Fokus nicht auf der Modellierung von Geschäftsprozessen liegt. Dennoch wird die UML häufig für die Modellierung von Geschäftsprozessen eingesetzt.

Für die Untersuchung der einzelnen Modellierungssprachen ist eine Abgrenzung der Begriffe Workflow und Geschäftsprozess voneinander hilfreich. Während ein Geschäftsprozess die auf die Unternehmensziele ausgerichteten betrieblichen Abläufe beschreibt, ist ein Workflow Grundlage für einen automatisiert ausführbaren Prozessablauf [15]. Die Workflowmodellierung folgt somit der Geschäftsprozessmodellierung [10]. Geschäftsprozesse werden aus der strategischen Ebene abgeleitet und auf der fachlich-konzeptuellen Ebene modelliert. Workflows erweitern Geschäftsprozesse um für die automatisierte Aufgabendurchführung notwendigen Spezifikationen. Die Anwendungssystemgestaltung sowie die Organisationsgestaltung, also die Zuordnung von Aufgabenträgern zu den Aufgaben, erfolgt auf Basis von Workflowmodellen. Betrachtet man das Aufgabenmodell unter diesem Blickwinkel, so kann die Außen-sicht einer Aufgabe zur Beschreibung von Geschäftsprozessen verwendet werden, während die Spezifikation des Lösungsverfahrens in der Innensicht Grundlage für eine Workflowmodellierung ist. Das bedeutet insbesondere, dass für die Spezifikation von funktionalen und nichtfunktionalen Anforderungen sowohl Workflow- als auch Geschäftsprozessmodelle herangezogen werden können.

3.1 Business Process Model and Notation

Die Business Process Model and Notation (BPMN) entstand als Prozessmodellierungssprache zur Ausführung in Business Process Management-Systemen [16]. Inzwischen wird die BPMN auch zur Modellierung von fachlich ausgerichteten Modellen verwendet und entwickelt sich dabei in der Praxis zum de-facto Standard der Geschäftsprozessmodellierung im Kontext der Prozessautomatisierung [17]. Dennoch ist die BPMN primär auf die Modellierung von Workflows ausgerichtet, da der Fokus auf der Spezifikation von Lösungsverfahren für die einzelnen Aufgaben eines Geschäftsprozesses liegt [18].

Die BPMN wird von der Object Management Group verwaltet und liegt gegenwärtig in der Version 2.0² vor. Für das Untersuchungsziel dieser Arbeit sind nur wenige der sehr zahlreichen Modellierungselemente der BPMN von Bedeutung (um Mehrdeutigkeiten bei der Übersetzung von bzw. Verwendung der deutschen Begriffe zu vermeiden, wird die englische Bezeichnung angegeben).

² <http://www.omg.org/spec/BPMN/2.0/>

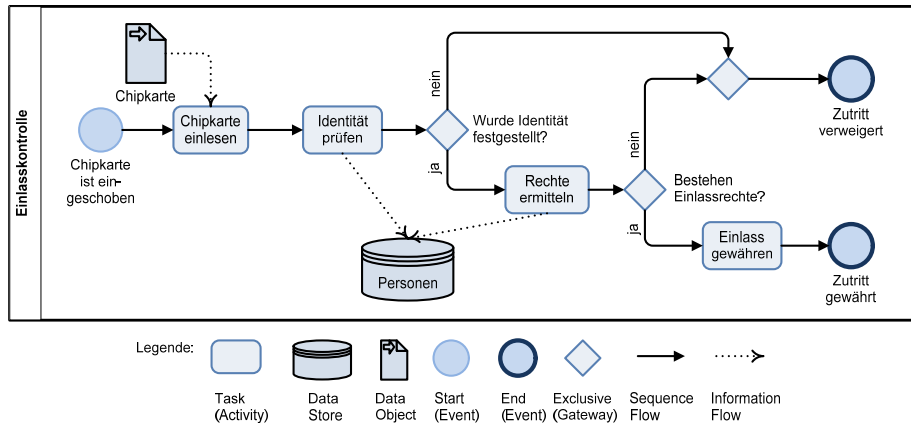


Abb. 4. Geschäftsprozess „Einlasskontrollprüfung“ in BPMN-Notation

Das Fallbeispiel „Einlasskontrollprüfung“ ist in Abbildung 4 in BPMN-Notation dargestellt. Zentrales Element eines BPMN-Modells sind Activities. Diese können entweder Teilprozesse (Sub-Process - hier nicht verwendet) oder elementar sein (Task). Activities können solange zerlegt werden, bis es sich nur noch um Tasks handelt. Tasks entsprechen somit Aktionen gemäß dem Aufgabenmodell. Ereignisse werden als Start- und Endereignis modelliert, Zwischenereignisse können ebenso angegeben werden. Für die Modellierung von Zielen ist in der BPMN kein Notationselement vorgesehen. Das Sachziel wird häufig durch die Bezeichnung der Activity angegeben, das Aufgabenobjekt kann durch Data Stores modelliert werden. Formalziele können in der BPMN nicht modelliert werden. Somit ist eine Spezifikation von nicht-funktionalen Anforderungen in BPMN nicht möglich [19]. In der Literatur gibt es Vorschläge, wie nichtfunktionale Anforderungen in BPMN modelliert werden können. Charles und Hollunder schlagen bspw. vor, nichtfunktionale Anforderungen als Text-Artefakte abzubilden [4], während Pavlovski und Zou als zusätzliche Modellierungselemente Operation Conditions and Control Case diskutieren [5]. Tabelle 3 fasst die Zuordnung von Elementen des Aufgabenmodells zu den Modellierungselementen der BPMN zusammen.

Tabelle 3. Zuordnung Aufgabenmodell - BPMN

Elemente des Aufgabenmodells	Modellierungselemente BPMN
Aktion	Task (Activity)
Sachziel	Das Sachziel spiegelt sich meist in der Bezeichnung einer Activity wieder.
Formalziel	Kein Modellierungselement vorhanden
Aufgabenobjekt	Informationsflüsse zu Data Stores
Vor-/Nachereignisse	Events, modellierbar als Start-, Zwischen- und Endereignisse

3.2 Ereignisgesteuerte Prozessketten

Die Ereignisgesteuerte Prozesskette (EPK) wurde erstmals 1992 von Keller, Nüttgens und Scheer [20] vorgestellt und ist die zentrale Beschreibungssprache für Prozesse in der Steuerungssicht der Architektur integrierter Informationssysteme (ARIS) [21]. Die EPK stellt eine betriebswirtschaftliche Sichtweise in den Vordergrund der Prozessmodellierung [20-21]. Sie wird insbesondere im Rahmen des Geschäftsprozessmanagements zur Modellierung von Geschäftsprozessen eingesetzt [22-23].

Die EPK umfasst in ihrer Grundform die Elemente Funktion, Ereignis, Konnektor, Prozessschnittstelle und Kante. In der Regel wird mit dem Begriff EPK jedoch die erweiterte Form der EPK bezeichnet, welche Erweiterungsobjekttypen wie Ziel, Organisationseinheit oder Informationsobjekt bereitstellt [24]. Im weiteren Verlauf der Arbeit bezeichnet der Begriff EPK die Erweiterungsform der Ereignisgesteuerten Prozesskette.

Zentraler Baustein eines EPK-Geschäftsprozessmodells ist die Funktion. Mit ihr werden aktive Elemente modelliert, die einen Input in einen Output transformieren. Synonyme Bezeichnungen für Funktion sind in der Literatur Tätigkeit, Aufgabe oder Aktivität. Die Funktionsbenennung beschreibt den Zweck und somit das Sachziel der Aufgabe. Ereignisse als passive Elemente stehen für Vor- (Auslöser) und/oder Nachereignisse (Ergebnis) einer Funktionsausführung im Ablauf des Geschäftsprozesses und steuern den Kontrollfluss. Jeder EPK-Prozess beginnt und endet mit einem Ereignis. Als Erweiterungen der EPK werden für das Beispiel der Einlasskontrollprüfung nur Ziele und Informationsobjekte betrachtet. Funktionen werden von Zielen gesteuert, welche somit Formalzielen entsprechen. Weiter greifen Funktionen lesend und schreibend auf Informationsobjekte, welche die Beziehungen zum Aufgabenobjekt beschreiben, zu.

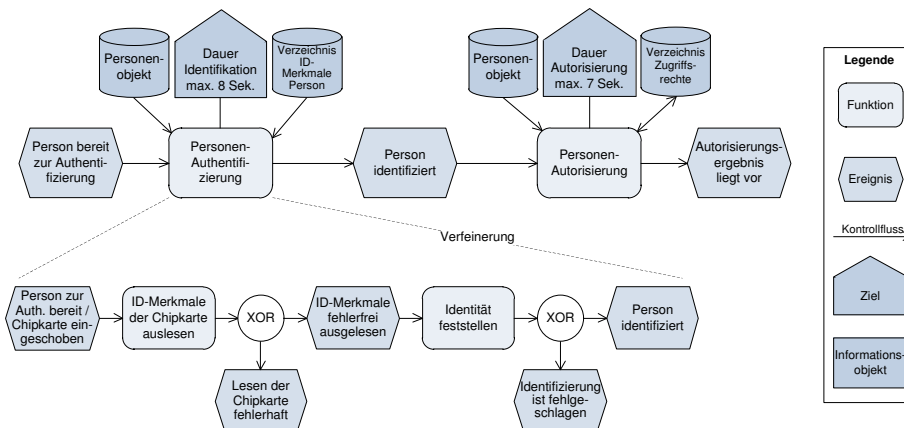


Abb 5. EPK-Geschäftsprozess der Einlasskontrollprüfung

Jede Funktion einer EPK kann, in einer sogenannten vertikalen Zerlegung, wiederum als eigenständiger Prozess verfeinert werden (vgl. Abbildung 5). Somit ist es möglich, Funktionen eines Geschäftsprozesses auf einem tieferen Abstraktionsniveau zu ver-

feinern und als Aktionenfolge zu interpretieren. Der Einsatz von EPKs zur Beschreibung von Lösungsverfahren ist somit grundsätzlich möglich. Dennoch wird aufgrund der fachlichen Ausrichtung der Sprache diese eher als Einstiegspunkt oder Grundgerüst für eine detaillierte Beschreibung von Lösungsverfahren gesehen [25].

Die folgende Tabelle 4 fasst die Zuordnung von Elementen des Aufgabenmodells zu den Modellierungselementen der EPK zusammen.

Tabelle 4. Zuordnung Aufgabenmodell - EPK

Elemente des Aufgabenmodells	Modellierungselemente EPK
Aktion	Funktion (abhängig von Interpretation der vertikalen Zerlegung)
Sachziel	Benennung der Funktion beschreibt ihren Zweck und spiegelt damit das Sachziel wieder
Formalziel	Ziele steuern die Ausführung von Funktionen und beschreiben somit Formalziele, häufig Qualitätsziele
Aufgabenobjekt	Referenzen auf Informationsobjekte
Vor-/Nachereignis	Ereignisse, welche eine Funktion auslösen oder Ergebnis der Funktionsausführung sind

3.3 Unified Modeling Language

Die Unified Modeling Language (UML) dient zur Modellierung, Dokumentation, Spezifikation und Visualisierung komplexer Systeme, unabhängig von deren Fach- und Realisierungsgebiet [6]. Da die UML vorrangig zur Spezifikation von Anwendungssystemen dient, liegt in der Literatur bezüglich der Nutzung von UML und ihrer Diagramme zur Geschäftsprozessmodellierung kein einheitliches Verständnis vor. Oftmals wird vorgeschlagen, Aktivitätsdiagramme in den Mittelpunkt der Betrachtung zu stellen (vgl. [6], [26]). Aktivitätsdiagramme zielen darauf, das Verhalten eines Systems und dabei insbesondere alternative Abläufe, Reihenfolgen von Aktivitäten, parallele und verschachtelte Aktivitäten sowie Ausnahmen und deren Behandlung abzubilden [27]. Entsprechend dem hiermit gegebenen Detaillierungsgrad der Betrachtung kann eine Aktion, die im Aktivitätsdiagramm die fundamentale Einheit ausführbarer Funktionalität darstellt [27], dem gleichnamigen Konzept des Aufgabenmodells zugeordnet werden. Für die aggregierte Betrachtung mehrerer Aktionen im Konzept einer Aufgabe hingegen findet sich im Aktivitätsdiagramm kein korrespondierendes Konzept. Hierfür bietet sich die Betrachtung des strukturorientierten Anwendungsfalldiagramms an. Ein Anwendungsfall kapselt eine in sich geschlossene Sammlung von Aktionen, die in einer spezifischen Reihenfolge ablaufen und die im Zusammenspiel eine Dienstleistung an der Umwelt des Anwendungsfalls erbringen [6]. Abbildung 6 demonstriert anhand des Fallbeispiels der Einlasskontrollprüfung, wie die genannten Diagramme in Beziehung zueinander stehen.

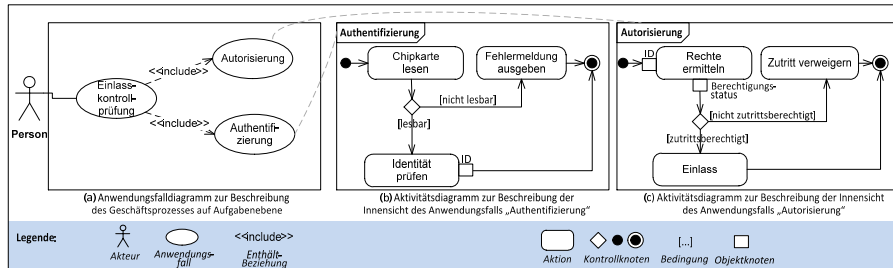


Abb. 6. Geschäftsprozess „Einlasskontrollprüfung“ in UML-Notation

Die Gesamtaufgabe und ihre Teilaufgaben führen jeweils zu separaten Anwendungsfällen, die über eine von der Gesamtaufgabe ausgehende Aggregationsbeziehung verknüpft sind. Jede Teilaufgabe wird in Form eines Aktivitätsdiagramms beschrieben. In Abbildung 6 nicht berücksichtigt ist die Darstellung der Innensicht der Gesamtaufgabe Einlasskontrollprüfung in Form eines Aktivitätsdiagramms. Aus diesem würden die Reihenfolgebeziehungen der gezeigten Teilaufgaben ersichtlich werden. Die folgende Tabelle 5 fasst die Zuordnung von Elementen des Aufgabenmodells zu den Modellierungselementen der UML zusammen.

Tabelle 5. Zuordnung Aufgabenmodell - UML

Elemente des Aufgabenmodells	Modellierungselemente UML
Aktion	Aktion (Aktivitätsdiagramm)
Sachziel	Das Sachziel spiegelt sich meist in der Bezeichnung eines Anwendungsfalls/einer Aktion wieder.
Formalziel	Kein Modellierungselement vorhanden
Aufgabenobjekt	Datenfluss und ggf. erforderliche Datenpuffer können durch Objektknoten abgebildet werden (Aktivitätsdiagramm). Eine Detailbetrachtung des Aufgabenobjekts kann zudem in einem Klassendiagramm erfolgen (hier nicht dargestellt).
Vor-/Nachereignisse	Kommentarfelder mit UML-Schlüsselwörtern <i>localPrecondition</i> bzw. <i>localPostcondition</i> (Aktivitätsdiagramm).

Eine Modellabbildung und -zuordnung von Zielen und damit indirekt von (nicht-) funktionalen Anforderungen ist in der UML nicht vorgesehen. Diese Lücke kann durch die System Modeling Language (SysML) geschlossen werden. SysML wird ebenso wie UML von der Object Management Group herausgegeben und stellt eine Erweiterung zur UML um eigene Modellierungsprimitive dar [6]. Die SysML sieht zwar keine Zielmodellierung vor, sie ergänzt die UML aber um eine Anforderungsdiagrammsicht, mit der die aus Zielen abgeleiteten Anforderungen an Aufgaben und zugehörigen Aktionen direkt abgebildet werden können. Abbildung 7 zeigt die beispielhafte Zuordnung von Anforderungen zu einem Anwendungsfall.

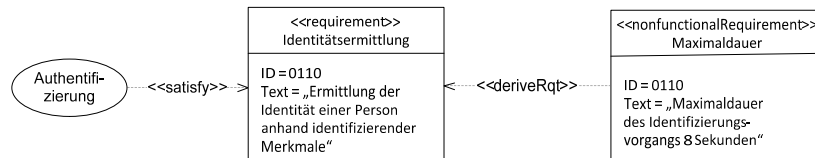


Abb. 7. Modellierung von Anforderungen mit der SysML

4 Zusammenfassung, Diskussion und Ausblick

Der vorliegende Beitrag beschäftigt sich mit der Frage, inwiefern funktionale und nichtfunktionale Anforderungen als Bestandteil der Spezifikation von Anwendungssystemen auf Basis zugrundeliegender Geschäftsprozessmodelle spezifiziert werden können. Eine dem Geschäftsprozessverständnis zugrundeliegende Annahme ist, dass alles betriebliche Handeln in Aufgaben stattfindet [11]. Während über diese Annahme hinaus Geschäftsprozessmodellierungssprachen i. d. R. ein heterogenes Modell- und Begriffsverständnis aufweisen, ermöglicht das Modell der betrieblichen Aufgabe ein gemeinsames Verständnis zur Einordnung von Modellelementen verschiedener Geschäftsprozessmodellierungssprachen. Dieses Verständnis kann für die Untersuchung der Ableitung von Anforderungen aus den einzelnen Elementen verschiedener Geschäftsprozessmodelle genutzt werden. In dem vorliegenden Beitrag wurde hierzu zunächst analysiert, wie sich funktionale und nichtfunktionale Anforderungen jeweils anhand der einzelnen Bestandteile des Aufgabenmodells begründen lassen. Dieses wurde anschließend an einer Fallstudie veranschaulicht. Zur Absicherung der gewonnenen Erkenntnisse wurde in einem Folgeschritt das Aufgabenmodell auf drei wesentliche Geschäftsprozessmodellierungssprachen übertragen und es wurde untersucht, inwiefern sich das Vorgehen zur Spezifikation von Anforderungen in den jeweiligen Sprachen - sofern berücksichtigt - mit dem Verständnis des vorliegenden Beitrags deckt. Im Detail konnten hierbei folgende Erkenntnisse gewonnen werden:

Zu den Konzepten der betrieblichen Aufgabe sowie den darin enthaltenen Aktionen konnten in jeder der drei untersuchten Modellierungssprachen korrespondierende Konzepte gefunden werden.

Entsprechend dem Aufgabenmodell bilden Sachziele einer Aufgabe die Basis zur Ableitung funktionaler Anforderungen, während Aktionen die identifizierten Anforderungen weiter konkretisieren. Sachziele werden in keiner der untersuchten Modellierungssprachen direkt berücksichtigt. Indirekt spiegelt aber in allen drei Fällen die Bezeichnung des mit einer Aktion korrespondierenden Elements das zugrundeliegende Sachziel wider.

Formalziele im Aufgabenmodell bilden die Grundlage zur Zuordnung nichtfunktionaler Eigenschaften. Bzgl. der untersuchten Modellierungssprachen werden Formalziele lediglich von EPK in Form des Ziel-Artefakts berücksichtigt, was erklärt, dass in BPMN und UML zur Zuordnung nichtfunktionaler Anforderungen auf Spracherweiterungen zurückgegriffen werden muss.

Ein Referenzieren des einer Aufgabe zugrundeliegenden Aufgabenobjekts ist in allen untersuchten Modellierungssprachen möglich, was die diesbezügliche

Zuordenbarkeit von Anforderungen zum Aufgabenobjekt aus Außensicht sicherstellt. Eine detaillierte Betrachtung der Struktur des Aufgabenobjekts (Innensicht) findet sich nur in der UML in Form des Klassendiagramms.

Zusammenfassend kann festgestellt werden, dass Geschäftsprozessmodelle eine geeignete und sinnvolle Ableitungsquelle für Systemanforderungen darstellen. Die zu erfüllenden Voraussetzungen einer Geschäftsprozessmodellierungssprache zur Ableitbarkeit von Anforderungen können anhand des Aufgabenmodells bestimmt und überprüft werden. Sofern eine Modellierungssprache einzelne Bestandteile des Aufgabenmodells nicht berücksichtigt, sind ggf. Spracherweiterungen in Betracht zu ziehen, um eine vollständige Ableitbarkeit von Anforderungen zu ermöglichen.

Aufbauend auf den gewonnenen Ergebnissen kann weiter untersucht werden, ob sich das Aufgabenmodell grundsätzlich für ein einheitliches Geschäftsprozessverständnis heranziehen lässt.

Weiterer Forschungsbedarf wird zudem in einer Übertragung und Weiterentwicklung der Erkenntnisse auf den Bereich des Model-Driven-Requirements-Engineerings (MDRE) gesehen. Hierfür kann untersucht werden, inwieweit das Aufgabenmodell einen geeigneten Ausgangspunkt bildet, um aus Geschäftsprozessmodellen automatisiert einen wesentlichen Teil der Anforderungsspezifikation eines Pflichtenhefts ableiten zu können.

Literatur

1. Pastor, Ó., España, S.: Full Model-Driven Practice: From Requirements to Code Generation. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B. et al. (eds.): CAiSE 2012. LNCS, Vol. 7328, pp. 701–702. Springer, Berlin Heidelberg (2012)
2. Mili, H., Tremblay, G., Jaoude, G.B., Lefebvre, É., Elabed, L., Boussaidi, G.E.: Business process modeling languages. *ACM Comput. Surv* 43, 1–56 (2010)
3. Bernhard, R., Jahn, B.U.: Modellgetriebene Entwicklung von serviceorientierten Architekturen. In: Hansen, H.R., Karagiannis, D., Fill, H.-G. (eds.): Business Services: Konzepte, Technologien, Anwendungen - Band 1. WI2009, Wien, Österreich, 25.–27.02.2009, pp. 89–98. Österreichische Computer Gesellschaft, Wien (2009)
4. Charles, O., Hollunder, B.: Non-Functional Requirements for Business Processes in the Context of Service-Oriented Architectures. In: ICSEA 2011: The Sixth International Conference on Software Engineering Advances, pp. 112–117 (2011)
5. Pavlovski, C.J., Zou, J.: Non-functional requirements in business process modeling. In: Hinze, A., Kirchberg, M. (eds.): Proceedings of the Fifth Asian-Pacific Conference on Conceptual Modelling (APCCM 2008). Wollongong, NSW, Australia, January 2008. Association for Computing Machinery, New York, N.Y (2008)
6. Rupp, C., Queins, S.: UML 2 glasklar. Praxiswissen für die UML-Modellierung. Hanser, Carl, München (2012)
7. IEEE: IEEE Standard Glossary of Software Engineering Terminology. Institute of Electrical and Electronics Engineers, New York, N.Y., USA (1990)
8. Sommerville, I.: Software Engineering. Pearson, München (2012)
9. Pohl, K.: Requirements Engineering. Grundlagen, Prinzipien, Techniken. Dpunkt, Heidelberg (2008)

10. Gadatsch, A.: Grundkurs Geschäftsprozess-Management. Methoden und Werkzeuge für die IT-Praxis. Vieweg + Teubner, Wiesbaden (2010)
11. Kosiol, E.: Organisation der Unternehmung. Gabler, Wiesbaden (1976)
12. Ferstl, O.K., Sinz, E.J.: Grundlagen der Wirtschaftsinformatik. Oldenbourg, München (2008)
13. Bea, F.X., Schweitzer, M.: Allgemeine Betriebswirtschaftslehre. Bd. 1: Grundfragen. UTB GmbH, Stuttgart (2009)
14. Becker, J., Probandt, W., Vering, O.: Grundsätze ordnungsmäßiger Modellierung. Konzeption und Praxisbeispiel für ein effizientes Prozessmanagement. Springer, Berlin Heidelberg, (2012)
15. Krallmann, H., Trier, M.: Workflow-Modellierung. In: Kurbel, K., Becker, J., Gronau, N., Sinz, E., Suhl, L. (eds.): Enzyklopädie der Wirtschaftsinformatik: Online-Lexikon. Oldenbourg, München (2008)
16. Allweyer, T.: BPMN 2.0 - Business Process Modeling Notation. Einführung in den Standard für die Geschäftsprozessmodellierung. Books on Demand, Norderstedt (2009)
17. Freund, J., Rücker, B.: Praxisbuch BPMN 2.0. Hanser, München [u.a.] (2010)
18. Pütz, C., Sinz, E.J.: Modellgetriebene Ableitung von BPMN-Workflowschemata aus SOM-Geschäftsprozessmodellen. In: Engels, G., Karagiannis, D., Mayr H.C. (eds.): Modellierung 2010. 24. - 26. März 2010, Klagenfurt, Österreich. LNI, Vol. 161, pp. 253–268. GI, Bonn (2010)
19. Gorton, S., Reiff-Marganec, S.: Towards a Task-Oriented, Policy-Driven Business Requirements Specification for Web Services. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B. et al. (eds.): LNCS, Vol. 4102, pp. 465–470. Springer, Berlin Heidelberg (2006)
20. Keller, G., Nüttgens, M., Scheer, A.-W.: Semantische Prozeßmodellierung auf der Grundlage "Ereignisgesteuerter Prozeßketten (EPK). In: Scheer, A.-W. (ed.): Veröffentlichungen des Instituts für Wirtschaftsinformatik. Heft 89, Saarbrücken (1992)
21. Scheer, A.-W.: ARIS - vom Geschäftsprozess zum Anwendungssystem. Springer, Berlin Heidelberg (2002)
22. Rump, F.J.: Geschäftsprozessmanagement auf der Basis ereignisgesteuerter Prozessketten. Formalisierung, Analyse und Ausführung von EPKs Teubner, Stuttgart (1999)
23. Becker, J.: Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. Springer, Berlin [u.a.] (2008)
24. Lehmann, F.R.: Integrierte Prozessmodellierung mit ARIS. Dpunkt, Heidelberg (2008)
25. Freund, J., Götzer, K.: Vom Geschäftsprozess zum Workflow. Ein Leitfaden für die Praxis Hanser, Carl, München (2008)
26. Oestereich, B.: Objektorientierte Geschäftsmodellierung mit der UML. Dpunkt, Heidelberg (2003)
27. Kecher, C.: UML 2.0. Das umfassende Handbuch. Galileo Press, Bonn (2007)

SIMchronization: Eine Modellierungsmethode zur Synchronisation von Material- und Informationsflüssen in der Instandhaltung

Christoph Prackwieser

Universität Wien, Forschungsgruppe Knowledge Engineering,
Währinger Straße 29, 1090 Wien, Österreich
prackw@dke.univie.ac.at

Abstract. Effiziente und reaktionsschnelle Supply Chain Networks bedingen eine enge Verknüpfung und verzögerungsfreie Kopplung von Informations- und Materialflüssen. Besonders wichtig ist das synchrone Zusammenspiel in der Domäne Instandhaltung, die sog. Maintenance Supply Chains (MSC) müssen regelmäßige Wartungsaufgaben kostengünstig durchführen und bei unplanbaren Ereignissen, wie Maschinenausfällen, kurzfristig, flexibel und zielgerichtet reagieren. In einer MSC werden Informationen über Ausfälle oder Ersatzteilbestände zur Steuerung herangezogen und erzeugen als Aufträge wiederum Informationsflüsse, welche auf die Materialbewegung einwirken. Diese komplexen und hochdynamischen Wechselwirkungen sollen mit der hier vorgestellten Methode SIMchronization transparent gemacht werden.

SIMchronization ist ein Instrument zur Dokumentation, Analyse und Verbesserung von Maintenance Supply Chains. Die statische Struktur der Supply Chain wird graphisch modelliert und das aktive und reaktive Verhalten ihrer Objekte mittels Regeln beschrieben. Eine Simulation dynamisiert das Modell und generiert sogenannte Zustandsfolgemodelle. Diese machen die chronologischen Wirkungszusammenhänge von Steuerungsinformationen und resultierenden Materialbewegungen im logistischen Netzwerk transparent.

Keywords: Modellierungsmethode, Metamodell, Maintenance Supply Chain, Instandhaltung, Simulation

1 Einleitung

Instandhaltungsmanager sehen sich zwei zentralen Herausforderungen gegenüber. Auf der einen Seite sollen im normalen Betriebszustand einer Anlage alle instandhaltungsspezifischen Tätigkeiten, wie Inspektion oder Wartung so ressourcenschonend und kostengünstig wie möglich durchgeführt werden. Konsequenz daraus ist, dass Lagerbestände für Ersatzteile reduziert und Instandhaltungsdienstleistungen und damit verbundenes Know-how externalisiert werden [1]. Auf der anderen Seite muss aber, im Falle einer ungeplanten Betriebsunterbrechung, beispielsweise in Folge eines Anlagendefekts, die Instandhaltungs-Organisation schnell und flexibel reagieren und

die benötigten Mittel kurzfristig zur Verfügung stellen. Dabei werden die verursachten Ausfallkosten maßgeblich von der Dauer der Betriebsunterbrechung bestimmt [1], die wiederum von der zeitgerechten Bereitstellung benötigter Informationen, Ersatzteile, Spezialisten und Werkzeuge am Einsatzort der fehlerhaften Anlage abhängt [2]. Dieses zur Durchführung aller Instandhaltungstätigkeiten notwendige Netzwerk an Informations-, Materialflüssen und Prozessschritten, das vom Lieferanten von Ersatzteilen und Dienstleistungen bis zum Abnehmer, der instand zu haltenden Anlage, reicht, wird als *Maintenance Supply Chain (MSC)* [3] bezeichnet. Maßgebliche Faktoren für die Effizienz einer MSC sind die wirksame Koordination übergreifender logistischer Abläufe wie Material- und Informationsflüsse und die reibungsfreie Schnittstellengestaltung [4]. Viele Schritte im Materialfluss werden erst durch Steuerungsinformationen ausgelöst, die wiederum Resultat vorhergehender Materialbewegungen sind. Je größer ein Netzwerk und je mehr Teile im logistischen Prozess involviert sind, desto komplexer wird dieses Zusammenspiel. Um hierbei mögliche Deadlocks auszuschließen, Verzögerungen in der Instandsetzung zu verringern und somit verursachte Lageraufwände und Kosten zu reduzieren, gilt es, die Wechselwirkungen zwischen Steuerungsinformation und physischen Teilen transparent zu machen und somit Optimierungspotentiale aufzudecken.

Die vorgestellte Methode *SIMchronization* hat zum Ziel die Analyse und Optimierung von Maintenance Supply Chains und die operative Bewertung von Instandhaltungsmethoden zu unterstützen. Insbesondere beantwortet sie die zugrundeliegende Forschungsfrage nach einer Methode zur Ermittlung der minimalen Bevorratung von Ersatzteilen unter der Prämisse einer anforderungsgerechten Anlagenverfügbarkeit.

Zunächst wird im Kapitel 2 anhand eines Literaturüberblicks eine Forschungslücke identifiziert und der gewählte Ansatz motiviert. Anschließend werden Anforderungen an die Modellierungsmethode und darauf anzuwendende Mechanismen abgeleitet. In Kapitel 4 wird das Ergebnis der Konzeptualisierung der Anforderungen vorgestellt und in Kapitel 5 die Dynamisierung des Modells mittels Regeln und Simulation erläutert. Den Abschluss dieser Arbeit bilden die Vorstellung eines praxisnahen Beispiels das zur initialen Evaluierung des Ansatzes dient und eine Zusammenfassung mit Ausblick auf kommende Forschungsaufgaben.

2 Motivation und verwandte Arbeiten

Eine strukturierte Literaturrecherche zum Thema Instandhaltungsoptimierung und MSC ergab, dass in vielen Arbeiten Ansätze zur Entscheidungsunterstützung für spezifische Fragestellungen der Instandhaltung vorgestellt werden. Als Input für das Lagermanagement werden ein- oder mehrdimensionale Klassifikationssysteme zur Ermittlung der Kritikalität von Ersatzteilen präsentiert [5] und zur Auswahl der passenden Instandhaltungsmethode mittels mathematischer Modellierung analytische Optimierungsansätze zur Minimierung der Kosten oder Erhöhung der Verfügbarkeit und Zuverlässigkeit entwickelt [6]. Allen diesen Modellen ist gemein, dass sie jeweils für die Lösung einer ganz spezifischen Problemsituation konzipiert sind und dem Anwender nur sehr wenige Freiheitsgrade bei der Anpassung an seine individuelle

Problemstellung lassen [6]. Wang [7] zeigt in seinem Aufsatz im Weiteren, dass die Optimierung der meisten Modelle nur nach genau einem Kriterium erfolgt; um reale Fragestellungen der Instandhaltung abzubilden, ist dies aber unzureichend. Dekker [8] schlussfolgert, bestätigt durch Van Horenbeek [6], dass die bestehenden mathematischen Modelle für den Instandhaltungspraktiker sehr schwer zu verstehen sind und die Anwendbarkeit durch ihre Starrheit leidet.

Andere Ansätze stellen mittels Petri-Netzen (bspw. [9]) oder Warteschlangenmodellen (bspw. [10]) den Abnutzungsprozess von Teilen dar oder verbessern die Integration der Instandhaltung in den Produktionsprozess mittels Simulation. Kontrollflussorientierte Modellierungssprachen wie Ereignisgesteuerte Prozessketten (EPK) oder die Business Process Modeling Notation (BPMN) (bspw. [11]) werden verwendet, um den Geschäftsprozess der Instandhaltung zu optimieren. Diese Ansätze können aber keine Aussagen über die Bevorratungsmenge von Ersatzteilen treffen.

Zusammenfassend kann als Ergebnis des Literaturstudiums festgestellt werden, dass aktuell kein Ansatz existiert, mit dem die gesamte MSC End-to-End analysiert und optimiert werden kann. Die bestehenden Optimierungsmodelle, die jeweils nur Fragmente der MSC betrachten, sind wenig flexibel und schwer verständlich. Durch die Vielzahl von Einflussfaktoren, das dynamische Verhalten der MSC und die stochastische Charakteristik der Entwicklung des Anlagenzustands ist ein mathematisches Modell zur Optimierung nicht geeignet. Zur Umsetzung der Methode *SIMchronization* wird daher das Instrument der Simulation verwendet, um die mittels graphischer Modelle erstellten MSCs zu analysieren und zu bewerten.

3 Anforderungen an eine Modellierungsmethode für die Domäne Instandhaltung

In diesem Kapitel werden Anforderungen an die Modellierungsmethode definiert und daraus abgeleitet, ob eine bestehende Methode, gegeben falls mit Modifikationen, verwendet werden kann oder eine neue Methode entwickelt werden muss. Heym [12] bezeichnet die daraus folgenden Schritte als „Method Engineering“. Zu den wesentlichen Komponenten einer Methode zählen für ihn und Gutzwiller [13] Aktivitäten, Rollen, Techniken, Ergebnisse und das Metamodell. Für Holten [14] beispielsweise, besteht eine Modellierungstechnik aus einer Sprache und Handlungsanweisungen. Beide dargestellten Ansätze sehen Algorithmen, die auf die erstellten Modelle angewendet werden, nicht als zentralen Bestandteil einer Methode an. Für die Konzeption der Methode *SIMchronization* ist jedoch neben der Modellierungssprache und dem Vorgehensmodell, der zur Analyse verwendete Simulations- und Animationsalgorithmus von gleichrangiger Bedeutung. Ein Ansatz der alle drei Komponenten, also *Modellierungstechnik, Auswertungsmechanismen und Algorithmen* und das *Vorgehensmodell* umfasst und deren Operationalisierung auf Basis einer metamodellierungsfähigen Plattform berücksichtigt, wurde von Karagiannis und Kühn [15] vorgestellt und als Bezugsrahmen für die Ableitung der Anforderungen in dieser Arbeit gewählt.

3.1 Anforderungen zur Modellierungstechnik

Zur Beschreibung der Modellierungstechnik gehören die Festlegung der Modellierungssprache und deren Handhabung. Formale Vorgaben ermöglichen dem Benutzer die Erstellung semantisch und syntaktisch korrekter Modelle auf denen die Auswertungsalgorithmen angewendet werden können. Frank [16] stellt darüber hinaus anwender- und anwendungsbezogene Anforderungen an eine Modellierungssprache.

Anwenderbezogene Anforderungen. In der Domäne Instandhaltung und besonders bei der Betrachtung der unternehmensübergreifenden MSC sind verschiedene Anwendergruppen, wie interne Instandhaltungs-Manager, externe Dienstleister oder Produzenten von Anlagen involviert. Sie alle sollten die Methode anwenden und deren Ergebnisse nutzen können. Somit ist eine Anforderung die Verwendung einer klar verständlichen, im Idealfall schon bekannten Notation und Begrifflichkeit. Die Methode muss unterschiedliche Detaillierungsgrade unterstützen, sich zur Abstimmung unternehmensübergreifender logistischer Prozesse eignen und den Spezialisten die notwendige Transparenz und Instrumente zur Optimierung und Synchronisation der Material- und Informationsflüsse geben.

Anwendungsbezogene Anforderungen. Ihrer Struktur nach handelt es sich bei einer MSC um eine ‚normale‘ Supply Chain deren Bedarf aber nicht von einer Kundennachfrage determiniert wird [1], sondern von der geforderten Reaktion auf die Zustandsentwicklung einer bestimmten Entität, dem Instandhaltungsobjekt. Inhalt einer MSC kann somit die Feststellung des aktuellen Zustands des Instandhaltungsobjekts, dessen Wartung zur Beibehaltung des betriebsfähigen Zustands oder die Instandsetzung mittels Austausch durch ein neuwertiges Ersatzteil sein. Aus diesen Anwendungsfällen ergeben sich die zentralen Anforderungen an die Abbildung von Instandhaltungsobjekten und der zugrundeliegenden Supply Chain.

Instandhaltungsobjekte und Ersatzteile. Müssen als eigenständige Objekte im Modell abbildbar und auswertbar sein. Sie werden als temporäre Objekte bezeichnet, da während der Betrachtungsdauer einer Analyse neue Ersatzteile im Modell generiert werden können, Instandhaltungsobjekte durch Ersatzteile getauscht werden, beziehungsweise verbrauchte Teile aus dem Modell entfernt werden.

Zur Ableitung von Anforderungen an die Modellierungsmethode der Supply Chain wurden verschiedene Quellen analysiert. Um die anwenderbezogenen Anforderungen bezüglich Bekanntheit zu erfüllen, wurde als Ausgangsbasis das Metamodell des SCOR-Modells [17] herangezogen. Im Rahmen einer Delta-Analyse wurden die Konzepte des SCOR-Modells mit den Basisklassen bekannter Supply Chain Ontologien verglichen und unter Berücksichtigung instandhaltungsbezogener Erfolgsfaktoren [1] weitere MSC-relevante Anforderungen identifiziert. Folgende Ontologien wurden für diese Analyse berücksichtigt: Onto-SCM [18], SCOntology [19], TOVE Ontology [20], IDEON Ontology [21], Ontologie mobiler Betriebsmittel [22]. Es ergaben sich folgende relevante Anforderungen:

Supply Chain. Die Ressourcen und Ersatzteile müssen am Einsatzort der Anlage bereitgestellt werden. Die dafür notwendigen Prozessschritte der MSC sind u.a. Beschaffung, Lagerung, Transport, Inspektion, Instandsetzung.

Ressourcen. Zur Inspektion oder Instandhaltung werden u.a. Werkzeuge, Sensoren, Handbücher und Mitarbeiter mit speziellen Kenntnissen oder Fähigkeiten benötigt.

Supply Chain Steuerung. Die Steuerung erfolgt über den Austausch von Informationen. Um die Synchronisation zwischen Material- und Informationsfluss modellhaft abbildbar und somit auch optimierbar zu machen, muss der Informationsaustausch zwischen Elementen der Supply Chain darstellbar sein. Die Steuerung und Auslösung der MSC kann zentral, dezentral oder autonom durch das Instandhaltungsobjekt erfolgen.

Um zu entscheiden, ob eine existierende graphische Modellierungsmethode verwendet werden kann, wurden 17 Modellierungsmethoden daraufhin untersucht, ob sie die zentrale Forderung nach Abbildung einzelner temporärer Objekte, also in der MSC bewegte Teile und Informationen, unterstützen. Die untersuchten Methoden umfassten sowohl kontrollflussorientierte Geschäftsprozess-Modellierungssprachen als auch Notationen, die im Produktions- oder Supply Chain Umfeld Verwendung finden, wie beispielsweise das Wertstrom-Mapping [23] oder die Prozess-Orientierte Analyse [24].

Nur Warteschlangenmodelle und gefärbte Petri Netze erfüllen dieses Kriterium. Beide Methoden lassen sich analytisch auswerten und es existieren bereits viele Auswertungsalgorithmen. Allerdings erfolgt die Auftragssteuerung in Warteschlangenmodellen ausschließlich aufgrund stochastischer Einflüsse und ist somit zur Abbildung eines Steuerungsprogramms ungeeignet. Die Marken eines gefärbten Petri Netzes stellen den jeweiligen Systemzustand modellhaft gut dar. Größere Petri-Netze können aber leicht unübersichtlich werden, da durch die Beschränkung auf zwei Arten von Knoten, den Stellen und Transitionen, die Abbildung der selben fachlichen Inhalte mehr Objekte benötigt als zum Beispiel bei Modellierungssprachen mit vielen Objekttypen wie BPMN. Diese zunächst subjektive Aussage kann durch Auswertung einer Arbeit von Dijkman et al. [25] objektiviert werden. Sie beschreibt eine automatische Transformation von BPMN-Modellen in Petri Netze. Während zur Modellierung der 13 Petri Netz-Modelle 546 Objekte erforderlich sind, so werden in BPMN nur 216 Objekte, also ca. 40%, benötigt, um denselben fachlichen Inhalt abzubilden.

Da somit keine der untersuchten Methoden alle Anforderungen unterstützt, wurde für die Methode *SIMchronization* eine neue Modellierungsmethode entwickelt.

3.2 Anforderungen zu Auswertungsmechanismen und Algorithmen

Die Konzeption von Auswertungsmechanismen für die Methode *SIMchronization* steht noch am Anfang und wird maßgeblich beeinflusst von der anstehenden Evaluation und endgültigen Ausgestaltung der Modellierungstechnik. Grundlegende Anforderung ist eine transparente Darstellung der chronologischen Abläufe von Materialbewegungen und Informationsflüssen in Kombination mit einer Darstellung des Zustands des Gesamtsystems. Damit die operativen Umsetzungsalternativen quantitativ

vergleichbar werden, sind Ressourcenkosten und Zeitinformationen, wie Durchlaufzeiten und Bearbeitungszeiten, erforderlich.

3.3 Anforderungen zur Vorgehensweise

Das Vorgehensmodell enthält Anleitungen, wie durch Anwendung der Methode die gewünschten Effekte in der MSC erreicht werden können. Eine Simulation liefert per se noch keine optimalen Ergebnisse, sondern ermöglicht eine Bewertung eines parametrisierten Modells. Um dennoch zu verbesserten Ergebnissen zu kommen, muss die Simulation im Rahmen einer Meta-Heuristik angewendet werden. So verändern beispielsweise Genetische Algorithmen [6] einzelne Modellparameter und reagieren auf die Ergebnisveränderung einer Simulation. Da in einer MSC sehr viele mögliche Parameter, wie Lagerpolitiken, Instandhaltungsstrategien, Ressourcenplanung oder auch Bearbeitungsreihenfolgen verändert werden können sieht der aktuelle Stand des Vorgehensmodells vor, dass die zielgerichtete Veränderung des Modells durch den Anwender erfolgt. Der Experte wird durch die graphische Modellierung und Transparentmachung der Wirkungszusammenhänge im Zeitverlauf mittels Animation in die Lage versetzt, Schwachstellen des Modells und deren Ursachen zu erkennen und durch Veränderung entsprechender Parameter diese zu beheben.

4 Metamodell

Die Konzeptualisierung der Anforderungen führt zu dem, in Abbildung 1 dargestellten Metamodell zur Beschreibung der MSC. Die Bildung der einzelnen Modelltypen folgte der Überlegung, dass Ressourcen und Teile/Komponenten in vielen verschiedenen MSCs eines Unternehmens Verwendung finden können und durch die explizite Abbildung in eigenen Modellen deren Wiederverwendung erleichtert wird:

- Supply Chain Network Modell zur Abbildung der Material- und Informationsflüsse
- Ressourcen Modell, enthält alle Ressourcen, die in einem oder mehreren Supply Chain Network Modellen genutzt werden
- Komponenten Modell, enthält die beweglichen Objekte des Materialflusses (Teile) und deren Struktur. In der Domäne Instandhaltung sind dies die in einer Instandhaltung Anlage verbauten Bauteile und die Ersatzteile, die zur Instandsetzung benötigt werden.

Für die Klassennamen des Supply Chain Network Modells wurden die Begriffe „Marke“, „Plan“, „Source“ und „Deliver“ des SCOR-Modells [17] übernommen.

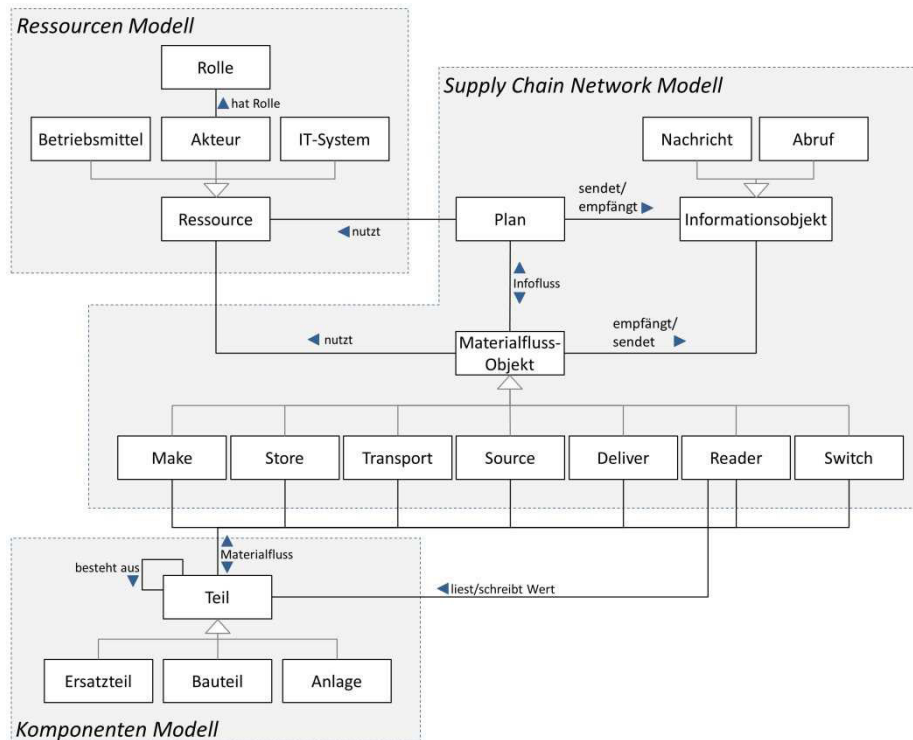


Abb. 1. Ausschnitt aus dem Metamodell der SIMchronization-Methode

Im Folgenden werden die wichtigsten Modellierungsklassen des Supply Chain Network Modells erläutert. Auf Attribute wird nur insofern eingegangen, als dass sie direkt in Zusammenhang mit der Dynamisierung des Modells stehen.

4.1 Modellierungsklasse „Make“

Objekte der Klasse „Make“ beschreiben die Tätigkeiten im Materialfluss, die mit der Gewinnung, Be- und Verarbeitung von stofflichen Gütern befasst sind [26]. Darunter fallen bspw. Produktionsschritte, Inspektionen, Verlade- oder Entladevorgänge und Instandsetzungstätigkeiten.

Attribut Produktionsprogramm. Ein Make-Prozessschritt benötigt als Input eine Kombination verschiedener Teiletypen in jeweils definierter Menge. Nach einer Bearbeitungszeit, liefert der Produktionsschritt als Output eine bestimmte Anzahl von Teilen pro Teiletyp. Diese Festlegung der benötigten Input-Kombination, der gelieferten Output-Kombination und der Bearbeitungszeit wird durch die Angabe eines *Produktionsprogramms* vorgegeben.

Attribut Auftragshöhe. Die Anzahl der in einer Periode zu startenden Aufträge wird als Integer-Zahl ≥ 0 angegeben. Aufträge können über einen Informationsfluss-Kanal per Nachricht erteilt werden, oder sich als Ergebnis einer in diesem Attribut auszuwertenden Regel (siehe Kapitel 5.1) ergeben.

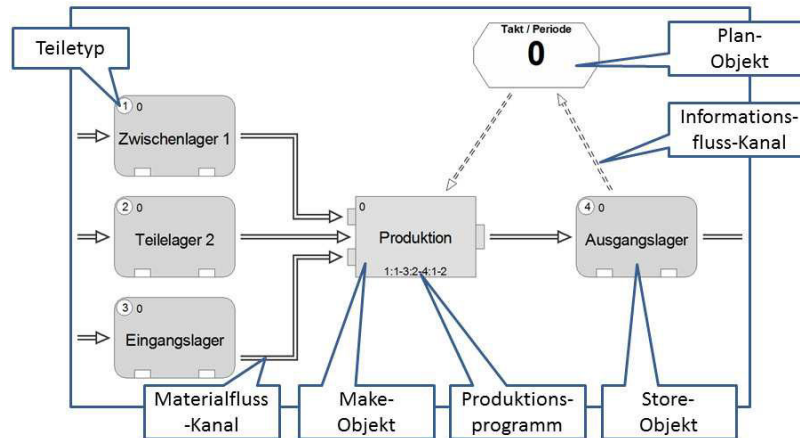


Abb. 2. Verwendung von Modellierungsklassen in einem Beispielmodell

4.2 Modellierungsklasse „Store“

Ein Objekt der Klasse „Store“ bildet einen Liegeplatz ab. Das Lagern und Puffern von Teilen dient der Zeitüberbrückung. Vorgängerobjekte im Materialfluss lagern Teile ein und nachfolgende Objekte rufen benötigte Teile aus einem Store ab.

Attribut Teiletyp. In einem Store können eine beliebige Anzahl von Teile genau eines Teiletyps gelagert werden. Die Typisierung eines Stores definiert somit eindeutig den Teiletyp der Teile, die über die ein- und ausgehenden Materialfluss-Kanäle des Stores bewegt werden können.

Attribut Wertminderung. Während der Lagerung können die Teile einer Wertminderung unterliegen. Diese Wertminderung kann in diesem Attribut als Regel hinterlegt werden. In der Domäne Instandhaltung wird dieses Attribut verwendet, um die Abnutzungs- oder Zustandsfunktion eines Teils abzubilden.

Attribut Lagerstand. Aktuelle Anzahl von Teilen des definierten Teiletyps die abrufbar im Lager liegen. Die Simulation (siehe Kapitel 5.2) aktualisiert dieses Attribut nach jedem Lagerab- oder zugang.

In der Domäne Instandhaltung wird der Liegeplatz neben dem klassischen Einsatz als Puffer auch zur Abbildung der instand zu haltenden Anlage verwendet. Für jede Komponente einer Anlage wird ein Store-Objekt modelliert. Das darin „gelagerte“ Teil ist somit das Instandhaltungsgut, das an seinem Einsatzort einer Abnutzung unterliegt.

4.3 Modellierungsklasse „Plan“

Zentrale Steuerungsaufgaben können mit Objekten der Klasse „Plan“ abgebildet werden. Plan ist Teil des Informationsflusses. Die Methode Wertstromdesign [27] verwendet ein ähnliches Konstrukt zur Abbildung von Steuerungsaufgaben, worunter Rother [27] den Geschäftsprozess Produktionsplanung und -steuerung (PPS) versteht.

Attribut Aktives Steuerungsprogramm. Kann ein Set von Regeln enthalten, die während der Simulation zu Beginn jeder Periode ausgewertet werden.

Attribut Reaktives Steuerungsprogramm. Kann ein Set von Regeln enthalten, die immer beim Eingang einer Nachricht ausgewertet werden.

4.4 Weitere Modellierungsklassen

Der Modelltyp *Supply Chain Network Modell* enthält zusätzlich folgende Klassen.

Relationsklasse „Informationsfluss-Kanal“. Zwischen den beiden verbundenen Objekten besteht ein Kanal zur Weitergabe von Informationen.

Relationsklasse „Materialfluss-Kanal“. Zwischen den beiden verbundenen Objekten besteht ein Kanal zur Bewegung von Teilen.

Klasse „Reader“. Liest Daten aus einem beweglichen Teil aus, bzw. beschreibt den Datenspeicher des Teils.

Klasse „Switch“. Verzweigungen im Materialfluss-Kanal. Die Steuerung, welchen Weg ein Teil nimmt kann dezentral oder zentral erfolgen.

Klasse „Source“. Bildet den Beschaffungsprozess ab. Generiert Teile und speist sie in das Materialflusssystem ein.

Klasse „Deliver“. Bildet den Absatzprozess ab. Konsumiert Teile und entnimmt sie dem Materialflusssystem.

Klasse „Transport“. Wie Klasse „Store“, nur mit einer festgelegten minimalen Liegezeit, der Transportzeit.

5 Dynamisierung des Modells

Voraussetzung für die Dynamisierung des Modells ist die Festlegung einer Periodendauer bzw. Taktzeit. Die Periodendauer wird maßgeblich von der Granularität der Modellierung der Supply Chain bestimmt und ist mindestens so lange wie die kürzeste Bearbeitungszeit aller Make-Prozesse des Modells. Perioden können bspw. für Minuten, Tage oder Schichtdauern stehen.

5.1 Regeln

Durch Instanziierung der Klassen des vorgestellten Metamodells wird ein statisches Strukturmodell der Supply Chain erstellt. Mittels zusätzlicher Beschreibung von Objekteigenschaften durch Regeln kann das aktive bzw. reaktive Verhalten von Objekten

im Zeitverlauf definiert werden. Die Regeln, gemäß dem OMG-Standard [28] sogenannte *production rules*, werden in dafür vorgesehene Attribute der Modellierungsobjekte eingetragen und beeinflussen damit während der Simulation das Verhalten dieser Objekte, beispielsweise indem sie:

- in Make-Objekten die Auftragshöhe ermitteln
- in Plan-Objekten durch Reaktion auf abgerufene Objektzustände und Versand von resultierenden Aufträgen mittels Nachrichten die Supply Chain steuern
- in Reader-Objekten Daten aus Teilen auslesen und weitersenden bzw. Daten darauf speichern (zur Abbildung von AutoID-Technologien, wie RFID).

Beispiele für Regeln.

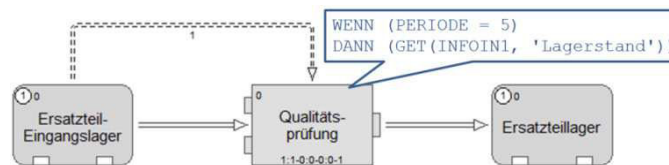


Abb. 3. Regel im Make-Objekt

Durchläuft die Simulation die 5. Periode, so feuert die Regel (siehe Abbildung 3) im Make-Objekt und liest den Lagerstand des Objekts am anderen Ende der Relation „INFOIN1“ ein. Dieser Wert wird der Simulation als Auftragshöhe zurückgegeben.

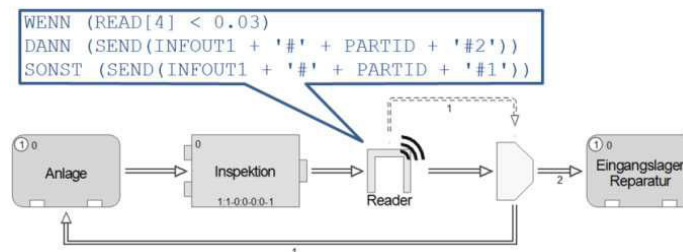


Abb. 4. Regel im Reader-Objekt

Diese Regel (siehe Abbildung 4) wird für jedes Teil ausgewertet, das am Reader vorbei geführt wird. Ist der Abnutzungsvorrat [1] eines Teils (hier ausgelesen aus Speicherplatz 4) kleiner als 0.03, so muss dieses Teil zur Reparatur. Der Reader sendet über ausgehenden Informationsfluss-Kanal „INFOOUT1“ eine Nachricht an den Switch, dass das Teil mit der PARTID den Materialfluss-Kanal mit Nummer 2 nehmen muss.

5.2 Simulation

Ein diskreter Simulationsalgorithmus durchläuft für eine vorzugebende Anzahl von Perioden wiederholt das Modell und bewegt Teile und Nachrichten durch die Supply Chain. Dafür werden in jeder Periode die Regeln in Kombination mit der Struktur gemäß einem Prioritätenschema ausgewertet.

Die Simulation liest für alle Objekte deren Vorbedingungen erfüllt sind, die verhaltensbeschreibenden Regeln aus und wertet diese mittels eines Regelauswertungsmechanismus (Rules Engine) unter Parametrisierung mit Input-, aktuellen Modell- und Simulationsdaten kontextspezifisch aus. Aus den zurückgegebenen Ergebnissen werden Nachrichten generiert und diese über Informationskanäle an Empfängerobjekte gesendet, Lagerbestände per Abruf ermittelt oder Produktionsaufträge für Make-Objekte generiert, die in weiterer Folge zu Teilebewegungen im Materialfluss führen.

5.3 Animation und Zustandsfolgemodelle

Die Animation visualisiert die Teile, Nachrichten und Abrufe, die durch die Informations- und Materialfluss-Kanäle bewegt werden. Dabei kann entweder der Weg eines Teils über mehrere Perioden hinweg durch die MSC betrachtet werden oder der Zustand der gesamten MSC während einer Periode dargestellt werden. Diese Darstellung einer Periode bezeichnen wir als *Zustandsfolgemodell*.

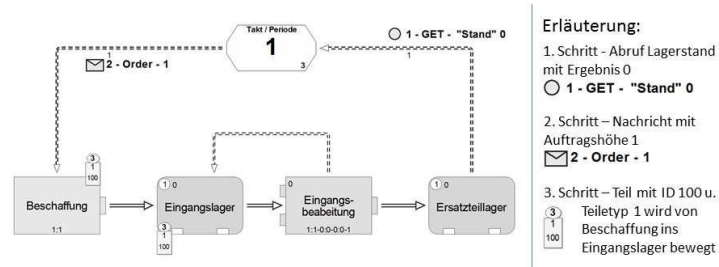


Abb. 5. Einfaches Beispiel eines generierten Zustandsfolgemodells für die Periode 1

Um die für die Analyse der Synchronisation erforderliche Reihenfolge der Vorgänge darzustellen, werden alle Teile, Nachrichten und Abrufe mit einer Nummerierung versehen. Teile, die in dieser Periode gelagert sind, sich also nicht bewegen, werden ebenfalls visualisiert. Man erkennt somit beispielsweise, welche Aktionen durch eine Nachricht auslöst werden und kann die resultierende Abfolge analysieren. Vergleicht man die Zustandsfolgemodelle mehrere aufeinanderfolgenden Perioden miteinander, kann man die Entwicklung von Lagermengen beobachten und somit Wartezeiten identifizieren.

5.4 Toolimplementierung und Evaluation mit Anwendungsbeispiel

Die Umsetzung von *SIMchronization* auf der Metamodellierungs-Plattform *ADOxx*[®] leitete die Evaluation der Modellierungstechnik und der Auswertungsalgorithmen ein (*ADOxx*[®] ist eine registrierte Marke der BOC Information Technologies Consulting AG). Es wurden sämtliche beschriebene Klassen und deren Attribute erstellt und der Simulations- und Animationsalgorithmus auf Skriptbasis prototypisch implementiert.

Anwendungsbeispiel. Im in Abbildung 6 dargestellten praxisnahen Beispiel soll die Anlagenausfallszeit durch eine effiziente Inspektionsplanung verringert werden und gleichzeitig die notwendige Ersatzteil-Lagermenge reduziert werden. Dafür wurde eine MSC erstellt, die das Instandhaltungsobjekt und dessen Zustandsverlaufs-Funktion (dargestellt im Diagramm rechts unten), die Inspektion, Instandsetzung und eine 2-stufige Beschaffungs-Supply Chain vom Ersatzteil-Lieferanten bis ins hausinterne Lager enthält. Die Steuerung der Ersatzteilbeschaffung erfolgt durch ein zentrales System, der Transport vom Lieferanten ins Haus dauert 5 Perioden, ein Instandsetzungsvorgang 2 Perioden. Das Modell ist über Regeln so parametrisiert, dass alle 5 Perioden eine Inspektion des Zustands des Instandhaltungsobjekts erfolgt. Liegt der Zustand unter einer definierten Grenze wird das Teil angesteuert und in Folge die Anlage mit einem Ersatzteil aus dem hausinternen Lager instandgesetzt. Die zentrale Steuerung bestellt für jedes verbaute Ersatzteil beim Lieferanten ein neues und ruft zusätzlich, jede 6. Periode den Lagerstand des internen Ersatzteil-Lagers ab. Ist dieser kleiner als 2 wird 1 Ersatzteil beschafft.

Der Simulationsalgorithmus wurde für 100 Perioden angewendet und liefert folgende Ergebnisse. Das Balkendiagramm „Anlagenverfügbarkeit“ stellt die Betriebsfähigkeit der Anlage durch Abbildung der Werte „0“ und „1“ dar. Aus dem Diagramm ist ersichtlich, dass die oben angeführte Parametrisierung des Modells während jedem Instandsetzungsvorgang zu einem 3 Perioden langen Anlagenausfall führt. Dieser ist durch die Dauer des Instandsetzungsvorgangs und eine verzögerte Beauftragung bedingt. Das Balkendiagramm „Lagermenge hausinterne Ersatzteile“ zeigt den Verlauf des hausinternen Ersatzteil-Lagerstands, der nach einer Einschwingphase 3 Stück beträgt und nach einer Instandsetzung kurz auf 2 Stück abfällt.

Eine Analyse des Modells zeigt, dass der nicht benötigte Lagerstand hauptsächlich durch ein Synchronisationsproblem von Abfragezeitpunkten des internen Lagerstands und der durch den Transport verursachten Zeitverzögerung zwischen Bestellung und Lagereingang entsteht. Passt man die Abfragefrequenz an die Lieferzeit an so ergibt eine erneute Simulation, dass sich bei gleichbleibender Anlagenverfügbarkeit die durchschnittliche Ersatzteil-Lagermenge dieses Typs um 1 Stück reduziert.

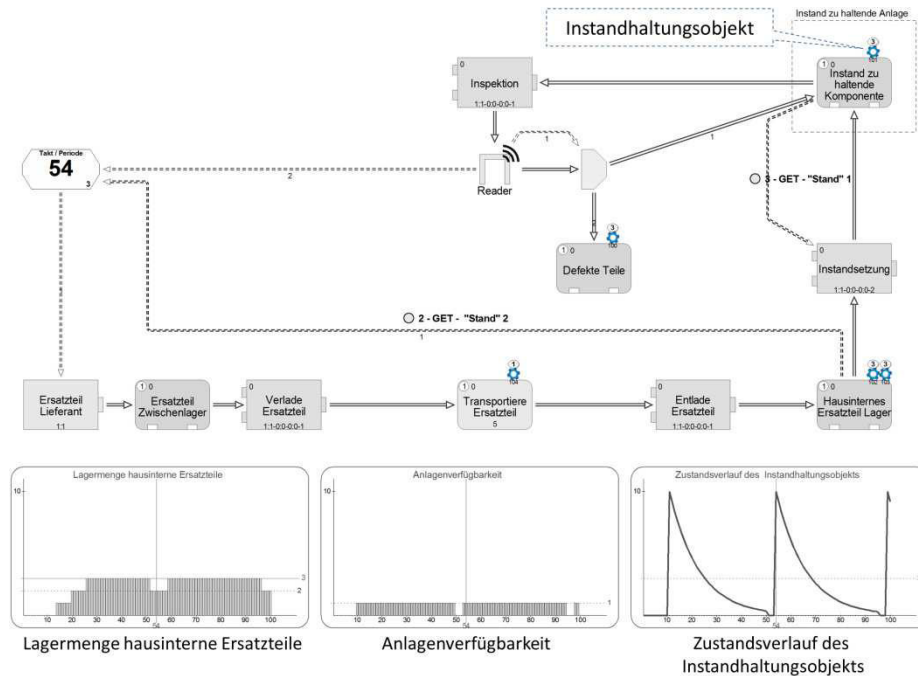


Abb. 6. Anwendungsbeispiel der Methode *SIMchronization*

6 Zusammenfassung und Ausblick

Die in dieser Arbeit vorgestellte Methode *SIMchronization* stellt ein Instrument zur Dokumentation, Analyse und Verbesserung von Maintenance Supply Chains dar. Die im Zuge der Methodenanwendung für jede Periode automatisch generierten Zustandsfolgemodelle geben Aufschluss über die chronologischen Wirkungszusammenhänge von Steuerungsinformationen und resultierenden Materialbewegungen im logistischen Netzwerk.

Die Anforderungen an die Modellierungstechnik wurden aus der wissenschaftlichen Literatur der Domänen Instandhaltungsmanagement und Supply Chain Management abgeleitet. Eine erste modellhafte Abbildung praxisnahe Fallbeispiele ist gelungen, muss aber noch mit Domänenexperten diskutiert werden, um gegebenenfalls weitere, bisher noch nicht berücksichtigte Anforderungen mit aufzunehmen.

Die technische Umsetzbarkeit der Methode auf einer Metamodellierungs-Plattform wurde anhand eines Prototyps gezeigt. Die in der Forschungsfrage formulierte Zielsetzung nach Unterstützung bei der Minimierung der Ersatzteil Bevorratung und Erhöhung der Anlagenverfügbarkeit wurde mittels eines praxisnahen Anwendungseispiels dargestellt. Weiterer Forschungsbedarf besteht in der Evaluation der Methode bezüglich ihrer Anwendbarkeit, Flexibilität und Akzeptanz in der Praxis. Des Weiteren muss mittels einer empirischen Untersuchung ein Vergleich dieser Methode mit

bestehenden, beispielsweise mathematisch-analytischen Methoden, angestellt werden um die Sinnhaftigkeit ihrer Anwendung zu belegen.

Zentral für die Anwendbarkeit und die Ergebnisqualität der Methode ist das noch detailliert zu beschreibende Vorgehensmodell. Es enthält neben den Handlungsanweisungen und Richtlinien zur Erfassung der benötigten Daten und Erstellung der Modelle auch Anleitungen zur zielgerichteten Anwendung der Simulation im Rahmen einer Meta-Heuristik und Interpretation der resultierenden Ergebnisse. Da ein entscheidender Erfolgsfaktor für die Praktikabilität und den Einsatz der Methode die zur Verfügung stehenden Auswertungsmechanismen sein werden, wird in deren Konzeption und Implementierung weiterer Forschungsbedarf gesehen.

Literatur

1. Alcalde Rasch, A.: Erfolgspotential Instandhaltung: theoretische Untersuchung und Entwurf eines ganzheitlichen Instandhaltungsmanagements. Erich Schmidt, Berlin (2000)
2. Pérès, F., Noyes, D.: Envisioning e-logistics developments: Making spare parts in situ and on demand: State of the art and guidelines for future developments. *Computers in Industry* 57, 490-503 (2006)
3. Fabry, C., Schmitz-Urban, A.: Maintenance Supply Chain Optimisation within an IT-Platform: Network Service Science and Management. In: 2010 International Conference on Management and Service Science (MASS), pp. 1-4 (2010)
4. Fettke, P.: Supply Chain Management: Stand der empirischen Forschung. *Zeitschrift für Betriebswirtschaft* 77, 417-461 (2007)
5. Kennedy, W.J., Wayne Patterson, J., Fredendall, L.D.: An overview of recent literature on spare parts inventories. *International Journal of Production Economics* 76, 201-215 (2002)
6. Van Horenbeek, A., Pintelon, L., Muchiri, P.: Maintenance optimization models and criteria. *International Journal of Systems Assurance Engineering and Management* 1, 189-200 (2010)
7. Wang, H.: A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research* 139, 469-489 (2002)
8. Dekker, R.: Applications of maintenance optimization models: a review and analysis. *Reliability Engineering & System Safety* 51 (3), 229-240 (1996)
9. Volovoi, V.: Modeling of system reliability Petri nets with aging tokens. *Reliability Engineering & System Safety* 84 (2), 149-161 (2004)
10. Park, C.-W., Lee, H.-S.: A multi-class closed queueing maintenance network model with a parts inventory system. *Computers & Operations Research* 38, 1584-1595 (2011)
11. Campos, M.A.L., Márquez, A.C.: Modelling a maintenance management framework based on PAS 55 standard. *Quality and Reliability Engineering International* 27, 805-820 (2011)
12. Heym, M.: Methoden-Engineering: Spezifikation und Integration von Entwicklungsmethoden für Informationssysteme. Universität St. Gallen, Hallstadt (1993)
13. Gutzwiller, T.A.: Das CC-RIM Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen. *Physica*, Heidelberg (1994)
14. Holten, R.: Konstruktion domänenspezifischer Modellierungstechniken für die Modellierung von Fachkonzepten. Inst. für Wirtschaftsinformatik, Münster (2001)
15. Karagiannis, D., Kühn, H.: Metamodelling platforms. In: Bauknecht, K., Tjoa, A.M., Quirchmayr, G. (eds.): *E-Commerce and Web Technologies*. LNCS, Vol. 2455, pp. 182-182 (2002)

16. Frank, U., Bodo, v.L.: Anforderungen an Sprachen zur Modellierung von Geschäftsprozessen (2003)
17. Supply Chain Council Inc.: Supply Chain Operations Reference (SCOR®) model. Overview - Version 10.0. (2010)
18. Ye, Y., Yang, D., Jiang, Z., Tong, L.: Ontology-based semantic models for supply chain management. *International Journal of Advanced Manufacturing Technology* 37, 1250-1260 (2008)
19. Vegetti, M., Gonnet, S., Henning, G., Leone, H.: Towards a supply chain ontology of information logistics within process industry environments. In: 4th Mercosur Congress on Process Systems Engineering, Rio de Janeiro (2005)
20. Fox, M.S., Barbuceanu, M., Gruninger, M.: An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour. *Comput. Ind.* 29, 123-134 (1996)
21. Madni, A.M., Lin, W., Madni, C.C.: IDEONTM: An extensible ontology for designing, integrating, and managing collaborative distributed enterprises. *Systems Engineering* 4, 35-48 (2001)
22. Jendoubi, L.: Management mobiler Betriebsmittel unter Einsatz ubiquitärer Computersysteme in der Produktion. Universitätsbibliothek der Universität Stuttgart, Stuttgart (2007)
23. Erlach, K.: Wertstromdesign: der Weg zur schlanken Fabrik. Springer, Berlin Heidelberg (2010)
24. Meyer, U.B., Creux, S.E.M., Weber Marin Silva, A.K.: Grafische Methoden der Prozessanalyse : für Design und Optimierung von Produktionssystemen. Hanser, München [u.a.] (2005)
25. Dijkman, R.M., Dumas, M., Ouyang, C.: Semantics and analysis of business process models in BPMN. *Information and Software Technology* 50, 1281-1294 (2008)
26. VDI: Begriffe und Erläuterungen im Förderwesen. VDI-Richtlinie 2411 (1970)
27. Rother, M., Shook, J.: Sehen lernen: mit Wertstromdesign die Wertschöpfung erhöhen und Verschwendung beseitigen. Lean Management Inst., Aachen (2004)
28. Production Rule Representation (PRR). Version 1.0, Vol. formal/2009-12-01 (2009)

Modeling Clinical Pathways - Design and Application of a Domain-Specific Modeling Language

Martin Burwitz, Hannes Schlieter, and Werner Esswein

Technische Universität Dresden, Chair of Information Systems, esp. Systems Engineering,
Dresden, Germany
{martin.burwitz,hannes.schlieter,werner.esswein}@tu-dresden.de

Abstract. Networking and collaboration in clinical care are increasingly entailing new requirements on supporting medical processes. The information technology (IT) in public health accordingly earns strategic relevance and encounters new potentials as well as challenging demands. The application of conceptual models in health care domain is almost entirely restricted to documentation tasks. Approaches like Model-Driven-Architectures or Workflow Management Systems have shown that the application of models, e.g. transformation, execution and formal interpretation, has huge potential. This article presents a modeling language for modeling clinical pathways. Three scenarios show the potential of conceptual models in health care domain and provide foundations for language requirements. Presenting a state-of-the-art of modeling languages for clinical domain and evaluating existing approaches to the requirements provide the gap to develop a domain-specific language. The potentials of the language and the use of corresponding models in medical treatment are demonstrated exemplarily including a discussion on model-driven management.

Keywords: Domain-specific Modeling Language, Medical Treatment Process, Clinical Pathway, Workflow Model

1 Introduction

1.1 Current Challenges of Hospital Organizations

The introduction of the Diagnosis-related Groups (DRG)-system initiated a paradigm shift in public health care. The new system of payment forced the hospitals to organize a process-based alignment of their care structures. The overall social development (demographics, multimorbidity) drives the system to a high degree of specialization and division of labor along the treatment chain.

For this reason hospitals are facing two essential challenges. There is a need for instruments and description facilities on the one hand, to instruct and to control the internal communication. On the other hand, breaking up traditional structures raises new demands pertaining the IT-support of the process-aligned health care. Conceptual models therefore can be used to face these challenges providing a knowledge base with clinical, medical and administrative knowledge, as well as formal process de-

scriptions as the foundation for future IT of hospital information systems (HIS), e.g. workflow management systems [1-2].

To further delineate the context of model usage in health care, we outline three different scenarios in section 1.2. Those scenarios are very challenging for hospital organization and show the potential for the application of conceptual models within health care domain. Therefore, we apply a framework that reflects the link of medical treatment instructions to the context of hospital organizations (Figure 1). Professional societies set common standards for medical treatments by publishing clinical practice guidelines (CPGs) that reflect current findings of medical science [3] (upper layer). These lead to clinical pathways (CP) within the organizational introduction (middle layer) in combination with the local clinical structure. The CPs can further be adapted to the HIS-Layer, where medical treatments are supported by CP-based IT systems (bottom layer). To provide the necessary information of the treatment to the medicating physician as the primary stakeholder, conceptual models of CPs should focus the middle layer. Therefore, CP-models represent the current medical scientific knowledge in combination with institution-specific facts and prepare it as best practices, e.g. a handbook for treatment in a human understandable form. The transformation for an IT-support of treatments is obviously possible but only a subordinate objective.

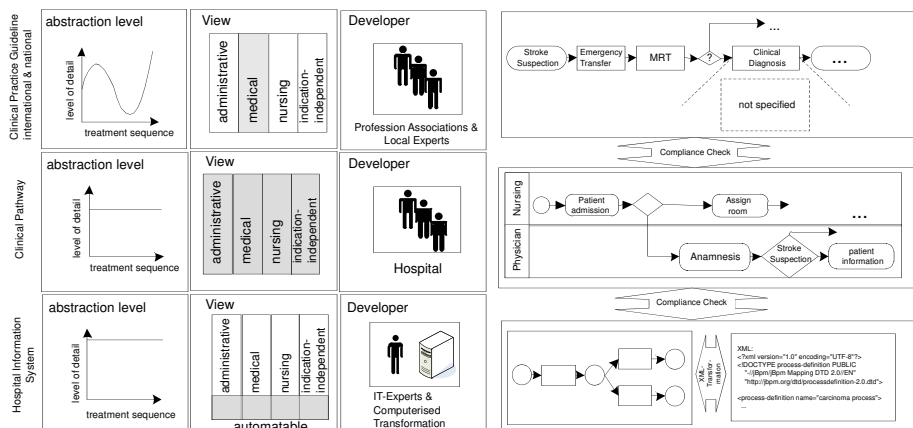


Fig. 1. Framework of CPG and CP in context of hospital organizations [4]

1.2 Scenarios for the Use of Conceptual Models in Health Care

Scenario 1: Implementing and Utilization of Clinical Pathways. Actually, hospitals increasingly need instruments and description approaches to guide and check the internal communication as well as to set standards for treatment processes. On organizational level, CPs have been established in recent years as a promising tool to ensure an adequate description of hospitals processes and to manage the balance between economic efficiency and the best medical treatment of patients [1], [5-6]. They are specific, standardized descriptions of clinical processes for defined combinations of

symptoms that are adapted to clinical conditions [7-8]. They are the integration medium for actual operating procedures, checklists, decision trees, and national as well as international suggestions as well as tool for resource management and continuous process improvement. One challenge of building CPs is the selection of modeling method. On the one hand the modeler should not be limited describing clinical treatment processes. On the other hand for model user, the physicians, the models have to be intuitively understandable.

Scenario 2: Integration of Clinical Practice Guidelines into Clinical Pathways. CPGs are defined as “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [9]. They provide decision guidance for health care provider that is based on evident practice for specific indications (diseases). Hospitals adapt recommendations of CPGs into CPs, but physicians require transparent information, where does a recommendation of CP come from or what is source of evidence. Firstly, between the layer of CPG and CP, there is the challenge of adapting a CP according to a CPG. Secondly, CPG information has to be integrated into CPs. This link can be used, e.g. if a CPG is under revision and some recommendations are changed.

Scenario 3: Clinical Workflows. Last but not least, the integration of standardization and daily business operations will only be reached, if CP-positioned process knowledge is integrated into the HIS and furthermore supports the medical treatment and if perceptions, gained on the operational level, flow back (bottom layer, Figure 1). The current application of CP usually remains on the organizational level. The pathway-driven IT-integration predominantly fails due to an excessive separation of process design and process execution. The break-up of traditional structures comes along with new requirements for the health care IT-support. The development of hospital information systems has been shaped by an evolutionary development of modules in the field of laboratory, radiology, nursing, image archiving systems and administrative, accounting-related systems. The actual supervision of patients instead, using evidence-based processes, was neglected since the systems rather tended to dealing with accountings than to an appropriate process support. HIS typically including a heterogeneous network of program systems from different manufacturers is the consequence of this development [10].

1.3 Research Method and Structure of Work

The essential objective of our research is the development of a modeling language, which covers the requirements that result from the previous illustrated challenges of the scenarios. Thus, this work is assigned to the design-oriented branch of management information science, the design science [11]. The principle of design science is based on the engineering paradigm in development of information systems. It focuses on solving practical and theoretical problems by creating new and innovative artifacts [11]. Core of design science is to eliminate arbitrary design results. Moreover, the design process should follow strict rules [11]. IT artifacts are the result of design sci-

ence. MARCH ET AL. differ four types of artifacts: constructs, methods, models or instantiations [12]. PEFFERS ET AL. propose a procedure for design science research [13]. The main steps of procedure are problem-centered initiation, definition of objective of a solution, design and development of the artifact as well as the demonstration and evaluation.

According to these steps, the paper is structured as follows: After introducing current challenges in health care and describing scenarios for using conceptual models, related work showing the state-of-the-art of modeling CPs is presented in section 2.1 followed by a presentation of requirements for a model-based building of pathways in section 2.2. This paragraph ends up in a comparison of existing modeling approaches with reference to the formulated requirements. In section 3, we present a domain-specific modeling language, which fulfills the requirements that are derived from the scenarios. The application of a pathway model is exemplarily illustrated by the treatment of a wisdom tooth in section 4. The paper closes with a discussion of research implications, and opportunities for further research in section 5.

2 Modeling Clinical Pathways

2.1 Related Work

The diagrammatical description of treatments and esp. treatment logics is encountered in various scientific disciplines with different objectives. Thus, a comparison of approaches for modeling CPs is difficult without revealing the intended goals of them. The most important approaches come on the one hand from the field of IT, esp. Health Informatics, and on the other hand from the field of Management Information System (MIS). While the former approaches are focusing on formalization of treatment procedures in a computer interpretable format, the latter ones aim to describe treatments in a human understandable form. In IT the models serve as a bridge between human and machine. In MIS, they are primarily an instrument for a standardized communication between humans, e.g. between medical specialists and medical assistants or between medical societies and hospitals. However, they can also be used to derive a more formal treatment description, which can be applied by a decision support system. This differentiation is also illustrated in Fig. 1. Physicians (middle layer) work with CPs (scenario 1 & 2). Therefore, they have to be in a human understandable form, which also means that it is mostly not suitable for a computer-based application (scenario 3).

In IT science, there are a variety of projects that investigate the area of construction and application of computer-interpretable CPs, which are namely: Asbru [14], EON [15], GLIF [16], GUIDE [17], PROforma, Prodigy [17-18]. They often use CPG as templates to demonstrate the applicability of the methods. The goal of all these approaches is to transfer the algorithmic structure of a clinical treatment into a formal shape. These models are in turn transformed into a computer readable form, e.g. proprietary XML dialects or the Arden-Syntax [19].

The MIS discipline concentrates on abstraction of Information Systems, their elements and relations. Therefore a variety of modeling language has been developed.

They can be divided into the branches of general-purpose languages (GPL) and the Domain Specific Modeling Languages (DSML) [20].

GPLs that are typically used for modeling CPs are the flowchart notation, Business Process Modeling Notation (BPMN) [21] as well as Activity Charts of the Unified Modeling Language (UML) [22].

In contrast, DSML are also applied for CP modeling. The Clinical Algorithm as one of the most popular notations goes back to the Society for Medical Decision Making, who promotes this language for the description of fundamental treatment logic in a diagrammatical way [23]. The notation is oriented to particular concepts of the flow chart. The purpose is to find a solution to the problem using a finite sequence of decisions and if-then-conditions, based on a clearly defined initial state [3], [18]. Without considering organizational aspects and individual patient data, the modeling of clinical algorithms takes place exclusively on the level of clinical process guidelines. This notation is also the basis for projects like GLIF [16] or GUIDE [24].

2.2 Requirements for Pathway Modeling Languages

The following set of requirements is derived by the analysis of the specified scenarios and typical treatment procedures supported by common experiences in process modeling. While IT-oriented approaches have developed several requirements under formalization aspects [25-26], the basis for the requirement analysis are MIS-approaches [27-28] focusing the organizational use of CP-models (middle layer in Figure 1).

Considering the key characteristics, the processes of clinical service provision can be perceived as medical business processes. Focusing the patient, these are defined by goal-oriented flows of activities (treatment steps) and decisions being initiated by patient states. The process model supports the task of controlling and optimizing these flows considering the available resources and including the illustration of input-output-relations between process steps and organizational units as well as information objects.

Requirement R1. *A language for modeling clinical pathways should provide the basic concepts of the medical business process modeling (patient state, treatment step, decision, process flow) and the ability to integrate information objects and responsibilities.*

Due to the properties of a treatment process as a service to the patient, the individual treatment steps are usually carried out sequentially. However different process analysis of existing use cases revealed that the specific order may vary from one process instance to another (variable flow). Also some treatments are recommended as optional or as a set of actions, where each, some or at least one of these has to be done.

Additionally, there are medical dependencies between several treatment steps, compulsory requiring a simultaneous and parallel execution (parallel flow). Furthermore, treatment process have often activities which are repeated permanently until the patient's condition is changing (iterative flow).

Requirement R2. *A language for modeling clinical pathways should provide concepts for describing indefinite order relations as well as compulsory parallel relations between treatment steps and iterating treatment steps.*

CPs are developed on the basis of CPGs and evidence-based knowledge (see scenario 1). However, compared to common process modeling techniques, there are no concepts to define the class of evidence of a recommendation given by a model element. This has to be considered within a modeling language. It would help to link the recommendations of CPGs more closely to the organizational standards of CPs.

Evidence knowledge is also important for decisions. In context of workflow execution (scenario 3), on the basis of patient-customized values of decision criteria, the further course of treatment is decided on during the pathway's execution. Therefore rules, generating decision-supporting recommendations resting upon the value combinations, can be defined already during modeling and planning the pathway. Additionally a medical decision against the system-generated proposal should naturally be given the higher priority along with an appropriate documentation and justification. The documentation later facilitates the traceability of occurred pathway deviations and their analysis [29].

Requirement R3. *A language for modeling clinical pathways should provide concepts for describing evidence-class of any recommendation and linking the source of evidence. Additionally, a concept to describe evidence-based decision is required.*

In medical treatment, especially in acute care, the compliance of waiting times as well as time limits is an important fact. For example, some medications can only be given exclusively within the specified time frame.

Requirement R4. *A language for modeling clinical pathways should provide concepts to describe temporal dependencies and explicit time events.*

Table 1 summarizes the compliance of the requirements of existing language approaches, categorizing them into the fields of Management Information Systems and Health Informatics. In the latter field, only languages that use diagrammatic concepts were investigated, since this is assumed as a critical need for human understandability. The detailed description of an approach that lead to this evaluation result can be found in the particular literature mentioned in section 2.1 on related work. In the comparison, it becomes obvious, that the researched model-based description approaches suffer from deficits, especially in the field of domain-specific requirements. Hence, the need for the development of a comprehensive domain-specific modeling language resp. at least for a language extension arises, improving the existing concepts and eliminating deficits by language extensions.

Table 1. Compliance of requirements of the modeling approaches

Requirement		R1				R2			R3			R4
		Clinical state	Treatment step	Decision	Process flow	Resources & responsibilities	Variable flow	Parallel flow	Iteration	Evidence indicator	Evidence-based decision	Time event
Approach												
MIS	BPMN	●	●	●	●	●	—	○	○	—	—	○
	Activity Chart	●	●	●	●	●	—	○	○	—	—	○
	Clinical Algorithm	●	●	●	●	—	—	○	○	—	—	—
IT	GLIF	●	●	●	●	○	○	○	○	○	—	○
	Guide	●	●	●	●	○	○	○	○	—	—	—
	PROforma	—	●	●	●	○	—	○	○	○	—	○
	Prodigy	●	●	●	●	—	—	—	—	○	—	○
	EON	●	●	●	●	○	—	—	○	○	—	○

● Requirement met ○ Requirement partly met — Requirement not met

3 Modeling Clinical Pathways – CP-Mod

The Clinical Algorithm provides the basic concepts of the below presented modeling language for CPs (CP-Mod). It is a common approach for the organization-focused modeling of CPGs and part of the “Leitlinienmanual”, a guide for a standardized development of CPGs in Germany, that concepts are often used for modeling CPs [17], [26–28]. However, it is lacking from several deficits regarding the requirements as Table 1 shows. Therefore it will be extended by additional language concepts that are especially used for modeling complex health care processes and reduce the existing deficits by addressing the requirements that are only partly or even not met.

3.1 Basic Concepts

The Clinical Algorithm already provides the basic process modeling concepts for treatment processes. *Clinical states* describe different constitutions of the patient type. The treatment according to the instructions of the pathway then hopefully improves the constitution of the patient and turns into another state at the end. A clinical state can also represent intermediates as milestones for the treatment. Clinical states can be further specified by a characterizing property, e.g. to describe the progress of the treatment or the indication. Clinical States at the beginning of a process can either occur exclusively (XOR) or simultaneously (AND).

Clinical states are transformed by actions, representing elementary activities or treatment steps. To improve the domain-specific semantic, an action can be catego-

rized by a type attribute, differentiating between diagnoses, therapy and supporting activities (see Figure 3). Only for these actions, which are not parts of diagnosis or therapy (e.g. prevention, after-care or rehabilitation) or which cannot be explicitly assigned, the generic action without specified type should be used. Supporting activities are rarely used, since CPGs and CPs focus medical processes, however it could be used e.g. for nursing activities.

Further basic concepts are decisions for modeling alternative process flows and information objects representing needed and resulting information. The sequence of treatment steps is defined by a control flow, a directed edge, which can be split using decisions (see section 3.2) or parallelization nodes. The latter is needed for parallel activities where the process is split into concurrent strands. The execution of these strands is enabled simultaneously and finally synchronized into a joint process flow. Moreover, loops and regions are used to describe iterative flows respectively pools of activities where an explicit execution order is not definable and individual due to the disposability of needed resources or where some treatment steps are compulsory or optional. Information objects can be integrated using the information flow type, which is also a directed edge, but using another concrete syntax to visualize the semantic difference. This could be used to integrate the cycle of necessary documents (e.g. lab reports) to support the obligatory documentation requirement. For modeling explicitly defined responsibilities, we used the lane concept of the BPMN 2.0.

The concrete syntax for each of the basic modeling concepts is designed under the premise of human understandability and can exemplarily be seen in section 4. Concerning the remaining requirements, the following section provides improvements of existing concepts and further extensions of the modeling language.

3.2 Extended Concepts

Evidence-based Medicine and Decision Support. Clinical practice guidelines, developed without a systematic research and analysis of the evidence of recommendations, do not meet the international quality standards. Accordingly, the demand for documenting to what extent recommendations draw upon results of clinical studies or expert opinions exists among others. Thus, various classification systems provide a quality assessment of particular sources of knowledge and information and so the quality assessment of the CPG. To evaluate process components within a treatment regarding its evidence and to comply the demand for documentation, the concept of the evidence indicator can be used (see the excerpt of the meta model in Figure 2). An evidence indicator can be attached to every process concept, such as action or decision, to allow the classification of the attached component according to the system of evidence levels of the AWMF [3]. The evidence indicator can contain further information to provide the relevant source of evidence, e.g. a reference to the corresponding section or page of the clinical practice guideline as depicted in scenario 2 in section 1.2.

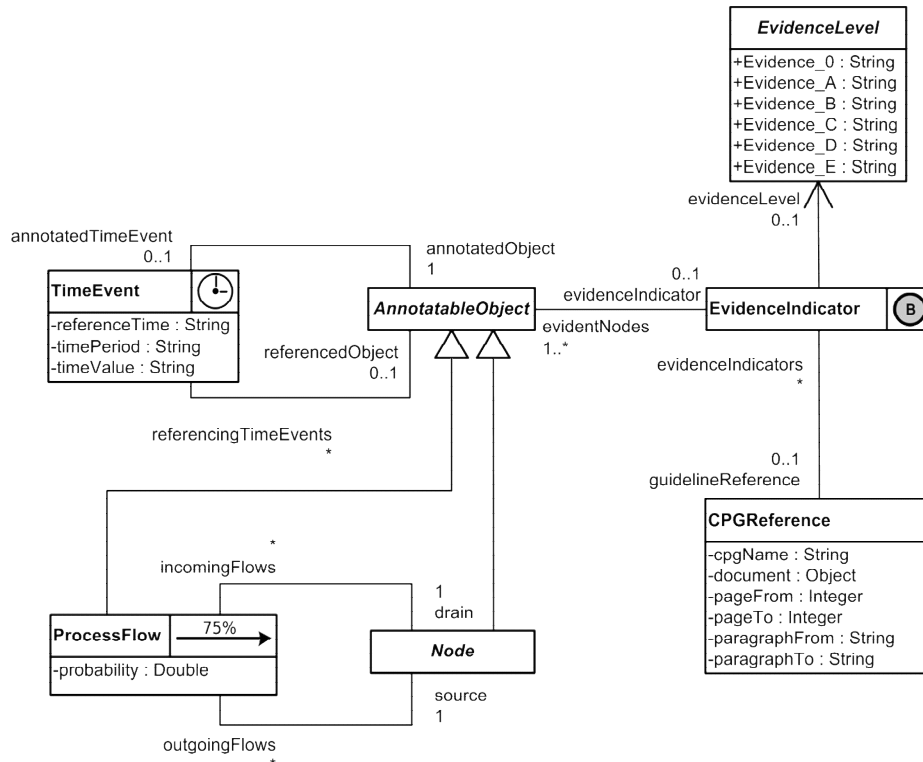


Fig. 2. CP-Mod meta model excerpt – Evidence Indicator and Time Event

To provide for decision support using evidence-based knowledge, the decision has to be extended furthermore by some constructs. Hence, it should be possible to define different decision criteria. Additionally, the description of decision logic provides a recommendation based on these criteria during the execution of the pathway. The attributes and dependencies of those concepts are illustrated in the excerpt of the meta model in Figure 3.

To illustrate an example, the decision for a medication of a patient could be affected by its blood sugar level. Hence, the model should contain the blood sugar level as criterion including a usual interval of values and allocate this to the decision of medication. The particular medication depending on the blood sugar level can then be described using logical expressions, allocated to the decision and concerning the defined criteria. During the execution, a model-based IT-System as part of scenario 3 (section 1.2) could use the logic and apply it to the current treatment, to suggest the recommended medication.

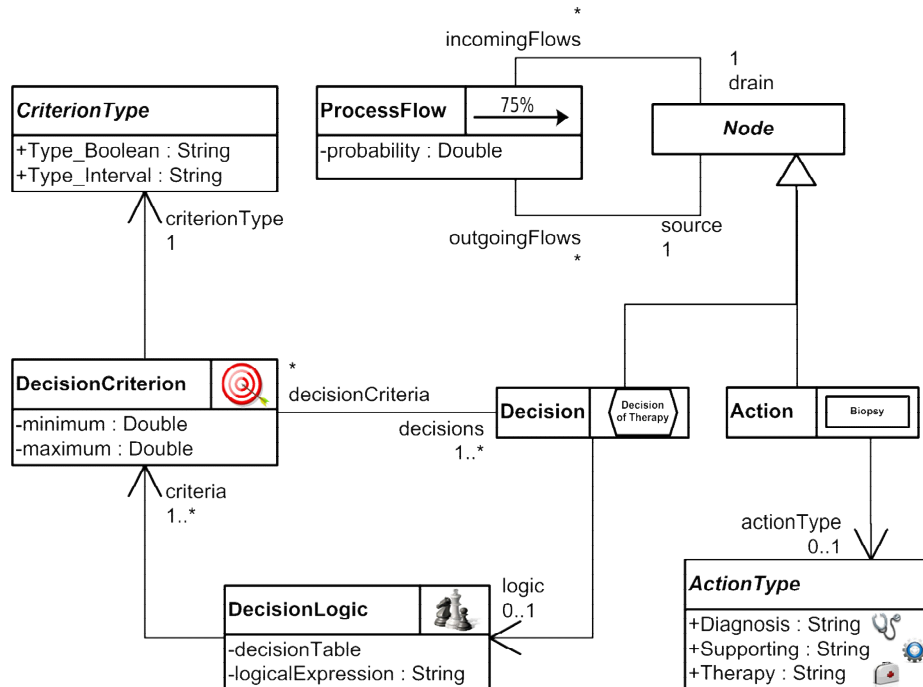


Fig. 3. CP-Mod meta model excerpt – Evidence-based Decision and categorized Action

Classification of Different Treatment Alternatives. Depending on the condition and progress of the patient resp. its convalescence, the decisions within a clinical pathway lead to different pathway progressions. For an optimal planning of a treatment it should be possible to forecast the pathway variant the patient will supposedly follow. Hence, process flows coming out of a decision (alternatives) can be attributed with an according *probability* (see Figure 3). The combination of all decisions shows the pathway’s total alternatives and their total probabilities as usual in a decision tree. As part of a continuous model improvement, these statistics are typically derived on an institutional level by a clinic board through analyzing past treatments. Due to the formal aspects of a CP model and its application for treatment execution, this analysis can be supported by IT based on the execution model (CP instance). Hence, the integration of probabilities could deserve different challenges of each above-mentioned scenario.

Time Events and Waiting Periods. Especially for process modeling in acute care, where intervals between individual treatments are relevant, appropriate approaches are needed to represent critical time slots, waiting periods or temporal dependencies between process steps. Therefore, the *time event* is added to the modeling language to adequately fulfill that requirement. In addition to the specification of a time value or a time period (e.g. within 4 hours), the relation to other time events and hence to other treatment steps can be drawn using a temporal reference point (e.g. within 2 hours

after surgery), so that a suitable representation according to the problem domain can be effected.

4 Demonstration

The developed modeling language for CPs (CP-Mod) was implemented in a meta-CASE-tool to demonstrate the modeling of a CP according to the CPG of a wisdom tooth treatment as an example. The resulting model is shown in Figure 4. Furthermore we implemented a prototypical workflow engine, which uses the created model to support the execution of a pathway during the treatment of an individual patient [27], [30].

The overall process is aligned to the responsible actors using lanes. The treatment starts with the obligatory actions of anamnesis and a radiographic test. Since the order of these steps is not explicitly determined, we can use an AND-region containing both. The radiograph created is necessary for the later decision of therapy, which is expressed by the information flow. Computer tomography and biopsy are additional but optional explorations expressed by an OR*-region. All these actions are diagnostic, which is illustrated by the corresponding icon. Since the dentist decides for a surgical excision, this action has to be supported concurrently by an anesthesia and therefore modeled using a parallelization. The following hemogram creation needs to be done within the first 3 hours, expressed by the time event. Since there is a need for further medication, this action has to be repeated until the hematocrit value fits a specific interval, shown by the loop. Evidence indicators are attached to each action, decision and information object, expressing the evidence level. The radiographic test and the anamnesis e.g. are highly approved by the current medical science, which is why we used the A-level classification. Finally, there are issues, which are not visible in the model presentation but integrated as attributes. The decision of therapy has several decision criteria (e.g. radix aberration [true/false], carious defect [true/false]) and decision logic, expressing that if a patient has a radix aberration and no carious defect, the antibiotics therapy is usually preferred.

For a treatment of an individual patient with indication for wisdom tooth troubles, the modeled treatment steps can be scheduled and illustrated in a pathway overview using a workflow engine. A system of traffic lights indicates the patient's current position within the pathway and controls the execution of the particular steps corresponding to the modeled workflow description. A recommendation is generated for each decision, based on the evaluation of patient-customized data concerning the modeled decision criteria and logic.

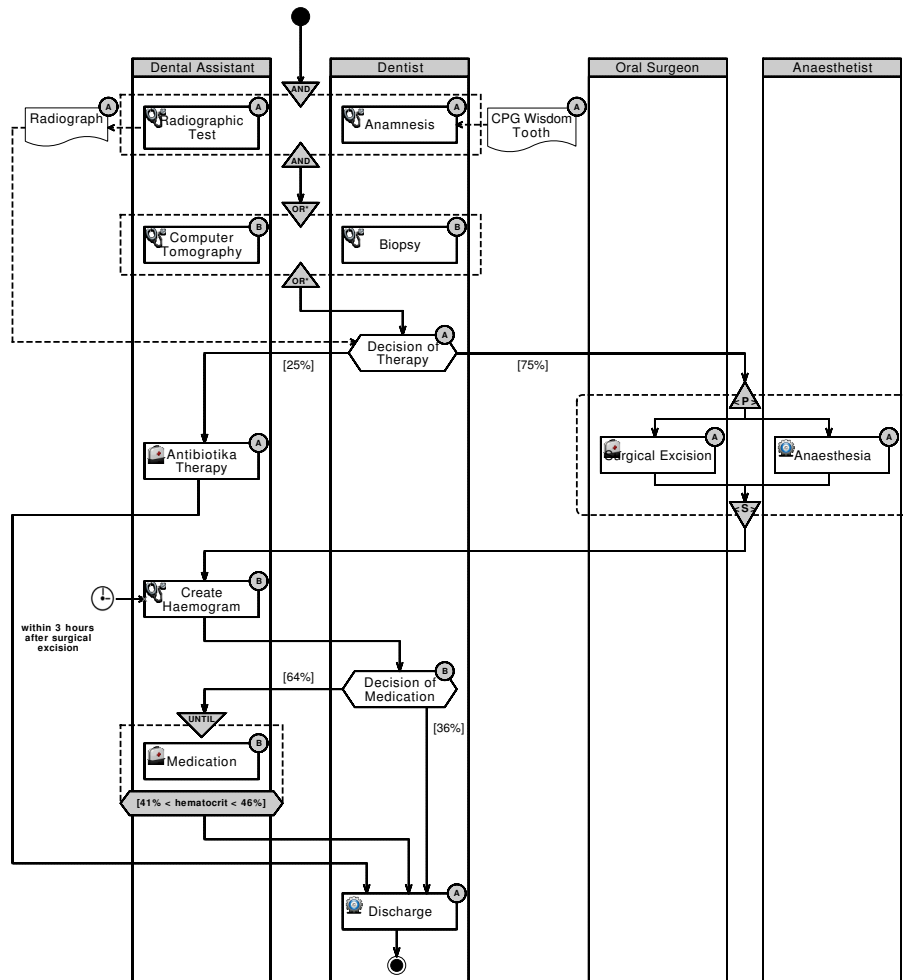


Fig. 4. CP model for a wisdom tooth treatment using CP-Mod

5 Conclusion

In the paper, we present an innovative approach for modeling CPs. Therefore, we outline three different scenarios in section 1.2, which show the potential range of model usage in health care. On the basis of these scenarios, requirements have been derived for an adequate modeling of CPs, which further have been applied to related work to evaluate existing approaches. It was shown that there are deficits in the area of integrating evidence information, i.e. of CPGs, modeling variable processes or time restrictions. Hence, a modeling language has been developed, which considers concepts for these aspects (see evaluation result in Table 2). The application of the designed language was demonstrated on the example of wisdom tooth treatment in

chapter 4. Therefore, we implemented the language in a meta-CASE-tool and used it in a prototypically developed workflow engine. Thus, we can show how the introduced modeling language fosters the scenarios.

Table 2. Compliance of requirements of the developed modeling language

Requirement	R1				R2			R3		R4	
	Clinical state	Treatment step	Decision	Process flow	Resources & responsibilities	Variable flow	Parallel flow	Iteration	Evidence indicator	Evidence-based decision	Time event
Approach											
CP-Mod	●	●	●	●	●	●	●	●	●	●	●

● Requirement met ○ Requirement partly met – Requirement not met

The recent explanation of the present work mainly focused on the adequate modeling of CPs (Scenario 1) considering the aspects of Evidence-based medicine (Scenario 2) and execution on a workflow management system (Scenario 3). However, many perceptions incur during the process execution in day-to-day business, which are usually hard to identify, because a multitude of decisions and process parameters are not documented or are even not aware to the hospital.

However, these perceptions should be transferred back into the conceptual level of CPs. Hence, further research should focus on information of the pathway instances, esp. in case of ad-hoc modification on instance level can flow back systematically and permanently into an improvement of the hospital-customized CPs, in terms of a management cycle.

Furthermore, information would be gained, serving the enterprise controlling (activity-based costing) as well as the top management on the strategic design of the hospital's process landscape. So the model-driven management cycle is closing. Extended by the possibility of executing agile processes based on CPGs and CPs, the presented approach now serves the continuous improvement process and feedback of information.

Additionally, further research should focus on the integration of resources, which are necessary for the treatment of patient regarding CPs. These considerations (not regarding the general data protection) for example would enable the creation of work and load profiles for the medical staff with reference to a particular treatment, analyzing processing times and costs for particular patients or patient types, identifying usual treatment patterns or evaluating the actual resource allocation.

The specified modeling language, CP-Mod, therefore, provides a conceptual foundation. Furthermore, a modeling language that covers all domain-concepts fosters building adequate process knowledge for hospitals, which not at least create a clear market advantage in the competitive health care market. With respect to the use of GPLs, another promising investigation would be the domain-specific extension of the BPMN meta model. The BPMN also tries to integrate human understandability and IT

formalization for business processes by providing an intuitive concrete syntax as well as specifications for transforming models into computer-interpretable workflow descriptions. The research task therefore is to evaluate the extensibility and improvements of BPMN and to find solutions for integrating concepts for health care specific problems, e.g. evidence-based aspects. This could be a great contribution in the field of pathway modeling and IT-support for medical treatments and could advance to an internationally wide accepted management instrument.

References

1. Rotter, T., Kugler, J., Koch, R., Gothe, H., Twork, S., van Oostrum, J., Steyerberg, E.: A systematic review and meta-analysis of the effects of clinical pathways on length of stay, hospital costs and patient outcomes. *BMC Health Services Research* 8, 265 (2008)
2. Juhirsch, M., Schlieter, H., Dietz, G.: Konzeptuelle Modellierung im klinischen Umfeld. *HMD - Praxis der Wirtschaftsinformatik* 281, 99–108 (2011)
3. ÄZQ, AWMF: Manual for Clinical Practice Guidelines (in German). *ZaeFQ* 95, 1–84 (2001)
4. Schlieter, H., Juhirsch, M., Dietz, G., Esswein, W.: Governance in Hospitals - The Case of Business Process Alignment. In: *AMCIS 2012 Proceedings*, Paper 30 (2012)
5. De Bleser, L., Depreitere, R., De Waele, K., Vanhaecht, K., Vlayen, J., Sermeus, W.: Defining pathways. *Journal of Nursing Management* 14, 553–563 (2006)
6. Panella, M., Marchisio, S., Di Stanislao, F.: Reducing clinical variations with clinical pathways: do pathways work?. *International Journal for Quality in Health Care* 15, 509–521 (2003)
7. Campbell, H., Hotchkiss, R., Bradshaw, N., Porteous, M.: Integrated care pathways. *BMJ* 316, 133–137 (1998)
8. Vanhaecht, K., Bollmann, M., Bower, K., Gallagher, C., Gardini, A., Guezo, J., Jansen, U., Massoud, R., Moody, K., Sermeus, W., et al.: Prevalence and use of clinical pathways in 23 countries – an international survey by the European Pathway Association. *Int. Journal of Care Pathways* 10 (1), 28–34 (2006)
9. Grossman, J.H., Field, M.J., Lohr, K.N.: *Clinical practice guidelines: directions for a new program*. National Academies Press (1990)
10. Winter, A., Zimmerling, R., Bott, O., Gräber, S., Hasselbring, W., Haux, R., Heinrich, A., Jaeger, R., Kock, I., Moeller, D.P.F.: Management of Hospital Information Systems: A definition (in German). In: *Proceedings of the 41th GMDS-Annual Conference*, pp. 34–38. Bonn (1996)
11. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. *MIS Quarterly* 28, 75–106 (2004)
12. March, S.T., Smith, G.F.: Design and natural science research on information technology. *Decision Support Systems* 15 (4), 251–266 (1995)
13. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24, 45–77 (2007)
14. Kaiser, K., Akkaya, C., Miksch, S.: How can information extraction ease formalizing treatment processes in clinical practice guidelines?. *Artif. Intell. Med* 39 (2), 151–163 (2007)
15. Tu, S.W., Musen, M.A.: Modeling data and knowledge in the EON guideline architecture. *Studies in Health Technology and Informatics* 84, 280–284 (2001)

16. Boxwala, A.A., Peleg, M., Tu, S., Ogunyemi, O., Zeng, Q.T., Wang, D., Patel, V.L., Greenes, R.A., Shortliffe, E.H.: GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *Journal of Biomedical Informatics* 37, 147–161 (2004)
17. Sarshar, K., Loos, P.: Klassifikation von Sprachen zur Modellierung medizinischer Behandlungspfade. In: Rebstock, M. (ed.): *Modellierung betrieblicher Informationssysteme – MobIS 2004*. LNI, Vol. P-40, pp. 43–59. GI, Bonn (2004)
18. Meiler, C.: *Modellierung, Planung und Ausführung Klinischer Pfade*. ibidem, Stuttgart (2005)
19. HL7: Arden V2.7, <http://www.hl7.org/implement/standards/ardensyntax.cfm>.
20. van Deursen, A., Klint, P.: Domain-specific language design requires feature descriptions. *Journal of Computing and Information Technology* 10, 1–17 (2004)
21. OMG: Business Process Model and Notation - Version 2.0, <http://www.omg.org/spec/BPMN/2.0/PDF>
22. OMG: Unified Modeling Language Specification Version 2.3, <http://www.omg.org/spec/UML/2.3/Infrastructure/PDF>
23. Barak, N., Margolis, C., Gottlieb, L.: Proposal for Clinical Algorithm Standards: Society for Medical Decision Making Committee on Standardization of Clinical Algorithms. *Medical Decision Making* 12, 149–154 (1992)
24. Johnson, P.D., Tu, S., Booth, N., Sugden, B., Purves, I.N.: Using scenarios in chronic disease management guidelines for primary care. In: *Proceedings of the AMIA Symposium*, pp. 389–393 (2000)
25. de Clercq, P.A., Blom, J.A., Korsten, H.H., Hasman, A.: Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial Intelligence in Medicine* 31, 1–27 (2004)
26. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D.: Comparing Computer-interpretable Guideline Models: A Case-study Approach. *Journal of the American Medical Informatics Association* 10, 52–68 (2003)
27. Burwitz, M., Schlieter, H., Esswein, W.: Modellgestütztes Management in Krankenhausinformationssystemen am Beispiel der Klinischen Prozesssteuerung. In: *Tagungsband Informatiktagung 2011*, Berlin (2011)
28. Heise, D., Heß, M., Strecker, S., Frank, U.: Rekonstruktion eines klinischen Behandlungspfad mithilfe domänenspezifischer Erweiterungen einer Geschäftsprozessmodellierungssprache: Anwendungsfall und Sprachkonzepte. In: Thomas, O., Nüttgens, M. (eds.): *Dienstleistungsmodellierung 2010*. Physica HD, Heidelberg (2010)
29. Schlieter, H.: *Ableitung von Klinischen Pfaden aus Medizinischen Leitlinien – Ein Modellbasierter Ansatz*, Univ.-Dissertation, <http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-95488> (2012)
30. Burwitz, M., Schlieter, H., Esswein, W.: Agility in medical treatment processes – A model-based approach. In: *Proceedings of Modellierung 2012*, pp. 267–279 (2012)

Integration and Architectures

Witold Abramowicz¹ und Robert Tolksdorf²

¹ Poznan University of Economics, Poznan, Poland

² Freie Universität Berlin, Berlin, Germany

Moderne Informationssysteme stützen sich nicht mehr auf einen singulären Datenbestand, sondern erzeugen Mehrwert durch die Integration heterogener Informationen aus unterschiedlichen verteilten Quellen. Die dafür notwendigen Architekturen sind zunehmend komplex, da in ihnen eine Fülle von Komponenten untereinander und mit externen Systemen interagieren muss. Entsprechend vielfältig sind die Gestaltungsmöglichkeiten und grundlegenden Paradigmen dieser Architekturen. Technisch sind neben der zunehmenden Auswahl von dienstorientierten Paradigmen verschiedene neuere Ansätze wie Semantische Technologien oder Anwendungen von Prinzipien der Selbstorganisation relevant.

Gleichzeitig stehen Informationssysteme in einem Kontext, der durch die Rahmenbedingungen ihrer Entstehung und Weiterentwicklung sowie durch umliegende Ökosysteme während ihrer Nutzung gebildet wird. Hier reichen die Gestaltungsmöglichkeiten von offenen Ansätzen wie Open Source bis zu geschlossenen Umgebungen wie Apples iOS. Dabei wird mit der Gestaltungsentscheidung auch immer eine Vorgabe für zugrundeliegende Geschäftsmodelle vorgenommen.

Eine Bewertung von Informationssystemen ist nur im Zusammenhang mit ihrem Kontext möglich. So kann der wirtschaftliche Vorzug eines Geschäftsmodells eine bestimmte aufwändigere technologische Grundlage in einer Gesamtschau rechtfertigen. Umfassende Kosten- und Nutzenmodelle sind notwendig, um diese Gesamtbewertung zu ermöglichen.

In diesem breiten Kontext sind in dem Track Beiträge ausgewählt, die besonders interessante und aktuelle Themen präsentieren: die Standardisierungsbemühungen im Cloud Computing, die Bestimmung der QoS für komponierten Web Services, die Governanceaspekte von mobilen Plattformen, Bewertungsmöglichkeiten eines App-Store Modells und einen Anwendungsansatz zur Integration von Planungssystemen in der Logistik auf Basis der Modellintegration.

Programmkomitee

Martin Bichler, Technische Universität München

Christian Bizer, Freie Universität Berlin

Hans Ulrich Buhl, Universität Augsburg

Agata Filipowska, Poznan University of Economics

Martin Gersch, Freie Universität Berlin

Norbert Gronau, Universität Potsdam

Volker Gruhn, Universität Duisburg-Essen

Martin Hepp, Universität der Bundeswehr München

Monika Kacmarek, Poznan University of Economics
Helmut Krcmar, Technische Universität München
Alexander Löser, Technische Universität Berlin
Günter Müller, Albert-Ludwigs-Universität Freiburg
Steffen Staab, Universität Koblenz-Landau
Andreas Speck, Universität Kiel

Ontology-Based QoS Aggregation for Composite Web Services

Paul Karaenke^{1,2}, Joerg Leukel¹, and Vijayan Sugumaran^{3,4}

¹ University of Hohenheim, Department of Information Systems 2, Stuttgart, Germany
{paul.karaenke, joerg.leukel}@uni-hohenheim.de

² FZI Forschungszentrum Informatik, Karlsruhe, Germany

³ Oakland University, School of Business Administration, Rochester, MI, USA
sugumara@oakland.edu

⁴ Sogang University, Department of Global Service Management, Seoul, Republic of Korea

Abstract. Determining the QoS (quality of service) of composite Web services is of high importance for both service providers and service consumers. Heterogeneity of service descriptions, however, often hinders the aggregation of QoS parameters. We propose ontology-based QoS aggregation that integrates the semantics of QoS parameters and their aggregation into the overall aggregation process. The contribution is a QoS aggregation ontology and a QoS aggregation method that uses this ontology. We demonstrate the usefulness of our proposal for designers of composite services and assess its computational efficiency.

Keywords: Ontology, QoS, Service Composition, SOA, Web Service

1 Introduction

Composite Web services play an important role in services computing, since they allow realizing business processes by composing elementary services under a shared workflow and providing this business process “as a service” [1]. An essential task within service composition [2] is determining the quality of service (QoS) of composite services. This task is called QoS aggregation.

QoS aggregation is non-trivial due to the heterogeneity of service descriptions, and in particular its QoS parameters. Heterogeneity is concerned with syntax, i.e., the lexicon for representing QoS, and semantics, i.e., the meaning of a QoS. Existing specifications for service descriptions (e.g., WSDL [3]) and service level agreements (SLAs) (e.g., WS-Agreement [4]) provide constructs for QoS parameters by a small set of attributes including name, data type, and unit of measurement. Service providers may use these constructs to define specific parameters. Whereas these specifications provide a common syntax, the problem of semantic heterogeneity still remains.

Any endeavor to resolve semantic heterogeneity faces the trade-off between standardization and flexibility. Standardization would restrict the use of QoS parameters to a particular set of standard parameters. Whereas aggregating these parameters is easy, the providers must revise their custom service descriptions according to the standard.

The main disadvantage of this approach is that it assumes the ontological commitment of all the service providers to that standard. This assumption, however, contradicts the idea of loosely coupled Web services in services computing [5]. Flexibility is achieved by giving the service providers full control over their service descriptions. The problem of heterogeneity must then be solved by the services users, who compose services from different providers under a shared workflow.

We aim at balancing the trade-off between standardization and flexibility by maintaining the custom service description and aligning them to an intermediate semantic layer. Unlike current semantic approaches to QoS standardization, we restrict the ontological commitment to one annotation per QoS parameter. Thus, the objectives of this research are to: (1) develop an ontology-based QoS aggregation method for composite Web services with well-formed workflows, and (2) apply this artifact to composite Web service of different complexity to demonstrate its usefulness and assess its computational efficiency. These objectives constitute design science research [6], because it designs and evaluates an artifact (method) that is informed by prior artifacts (composition patterns [14]) and theories (workflow [15], description logic [21]). We evaluate the proposed artifact by rigorously demonstrating its usefulness via a two-level evaluation, which consists of a detailed scenario (descriptive evaluation method) and a simulation using a prototype implementation (experimental evaluation method). The contributions of this research are the QoS aggregation ontology and aggregation method for composite Web services. In our previous work [7], we have presented a preliminary ontology. In this paper, we (1) propose and formalize the aggregation method, (2) revise the ontology, and (3) report the two-level evaluation.

The remainder of this paper is structured as follows. In the next section, we discuss the approaches for QoS aggregation. Subsequently, we provide the basic notions for composite service and QoS. Then, we present the QoS aggregation ontology and the method. We evaluate our proposal through a set of experiments and draw conclusions.

2 Related Work

QoS aggregation has been the subject of much research in services computing. We examine two aspects, heterogeneity of QoS parameters and aggregation methods.

Standardization on the syntactical level provides formats for representing parameters in service descriptions and SLAs. WS-Agreement [4] defines a format for SLAs. The QoS parameters are regarded as domain-dependent; hence WS-Agreement does not define specific parameters and does not solve semantic heterogeneity. Harmonization on the semantic level could overcome this problem, by defining common QoS parameters [8]. Ontologies have been proposed that specify the conceptualization of QoS formally [9-10], though they do not provide information for QoS aggregation.

For QoS aggregation, it would be sufficient to amend the service description with information about aggregation procedures. Haq et al. propose to state aggregation functions for SLA parameters explicitly [11]. These functions are stored in a specific attribute that extends the WS-Agreement specification. The drawback of this approach is service providers need to determine the function for all parameters correctly.

Aggregation functions for QoS parameters are used in [12-13], which stress that the specific parameter sets used are extensible without fundamentally altering the overall approach. This fact sheds light on the important insight to abstracting from diverse and domain-specific QoS parameters. A significant contribution stems from Jaeger et al. [14] who ground QoS aggregation on workflow patterns [15] and deduce so-called composition patterns and respective aggregation functions. Similarly, Cardoso et al. [12] propose a graph reduction algorithm for QoS aggregation.

The rationale for ontology-based QoS aggregation is to integrate constructs from Web services research, in particular, the composition patterns and aggregation functions, into an ontology and an aggregation method. This approach could minimize the additional effort for service providers in annotating their service descriptions as well as increase the flexibility for service consumers to explore a wider range of services.

3 Basic Model

We consider composite service as a collection of services that are governed under a shared workflow (directed acyclic graph, DAG). Figure 1 shows an example workflow, which contains six tasks and four gateways (represented by diamond shapes).

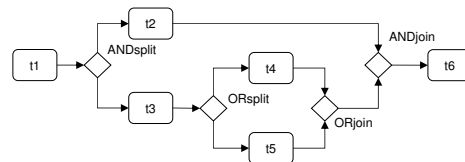


Fig. 1. Example workflow

To instantiate the workflow for different services, the tasks are separated from actual services by means of a binding, which is part of the succeeding definition.

Definition 1 (Composite Service). A composite service is a tuple $CS = (W, B)$ of workflow W and binding B . W is a directed acyclic graph $W = (T, G, A, CG)$, where

- T is a finite non-empty set of tasks $t \in T$, G is a finite set of gateways $g \in G$,
- $N = T \cup G$ is a finite set of nodes, with $T \cap G = \emptyset$,
- A is a set of arcs connecting tasks with tasks, tasks with gateways, and gateways with tasks, thus $A \subseteq (T \times T) \cup (T \times G) \cup (G \times T)$,
- CG is a function which assigns a type to each gateway; $CG: G \rightarrow \{XORsplit, XORjoin, ANDsplit, ANDjoin, ORsplit, ORjoin, Loop\}$.
- B is a binding function, which assigns a service $s \in S$ to each task, i.e., $B: T \rightarrow S$.

The model is restricted to well-formed (structured) workflows, where W has exactly one start and end node and is weakly connected. The correct usage of the different gateway types, thus how to connect nodes by arcs, is given as follows:

- Each task $t \in T$ has exactly one input arc and one output arc, i.e., $|\bullet t| = |t \bullet| = 1$.
- Each gateway $g' \in G$ with $CG(g') = \{XORsplit, ANDsplit, ORsplit\}$ has exactly one input arc and at least two output arcs, i.e., $|\bullet g'| = 1$ and $|g' \bullet| \geq 2$.
- Each gateway $g' \in G$ with $CG(g') = \{XORjoin, ANDjoin, ORjoin\}$ has at least two input arcs and exactly one output arc, i.e., $|\bullet g'| \geq 2$ and $|g' \bullet| = 1$.
- Each gateway $g' \in G$ with $CG(g') = \{Loop\}$ has one input arc and two output arcs, i.e., $|\bullet g'| = 1$ and $|g' \bullet| = 2$. Let $a_1 = (n_1, g')$ be the input arc, then one output arc a_2 connects back to node n_1 with $a_2 = (g', n_1)$, and one output arc a_3 connects to another node with $a_3 = (g', n_2)$ and $n_1 \neq n_2$.

The CS model is not bound to a particular workflow language such as the Business Process Model and Notation (BPMN) and the Web Services Business Process Execution Language (WS-BPEL), but conveys common constructs of workflow languages.

QoS is described in a SLA, which contains QoS information by means of guarantees on parameters; the most common guarantee is that the parameter fits into a given domain of minimum or maximum values, e.g., maximum execution time. The basic model of SLA is made of service, parameters, and parameter values (definition 3). The classification C is important for QoS aggregation, because otherwise service parameters of different services cannot be identified as those to be aggregated.

Definition 2 (Composite QoS). $QoS(CS)$ is a function that determines the composite QoS for the services S of a composite service CS .

Definition 3 (SLA). A service level agreement is a tuple $SLA_s = (P, V, C)$. P is a set of parameters p_1, \dots, p_j , V is a set of values v_1, \dots, v_j for these parameters, and C is a function which assigns a type to each parameter.

4 Ontology-based QoS Aggregation

We present our ontology-based QoS aggregation method by (1) deducing the aggregation ontology from composition patterns and (2) specifying the aggregation algorithm.

4.1 QoS Aggregation Ontology

Rationale. The composite QoS depends on two determinants: QoS parameters and workflow. QoS parameters are diverse with regard to number, name, data type, and conceptualization and rarely adhere to any standard. Instead of contributing to their harmonization, we propose a classification exclusively with regard to their aggregation. This classification is built upon the following principle: If two parameters share the same aggregation function, then they belong to the same parameter type, regardless of other characteristics. Applying this principle results in five parameter types:

- Type 1: Parameters that are always summed up along all deterministic paths of the workflow (e.g., cost of service execution). For non-deterministic paths, the aggregation depends on whether the lower or upper bound is calculated.

- Type 2: Parameters for which the critical path is determined by the maximal values in parallel executions (e.g., duration of execution time).
- Type 3: Parameters that denote a capacity (e.g., throughput).
- Type 4: Parameters that denote a probability (e.g., uptime probability).
- Type 5: Parameters for which the critical path is determined by the minimal values in parallel executions (e.g., key length of the service's encryption algorithm).

The second determinant workflow is analyzed by means of workflow patterns [14]; this analysis arrives at a set of composition patterns $CP = \{Sequence, Loop, XORXOR, ANDAND, ANDDISC, OROR, ORDISC\}$. For instance, the *OROR* pattern describes an OR-split followed by OR-join. *ANDDISC* and *ORDISC* describe AND-split and OR-split followed by a m-out-of-n join (discriminator). The latter is used when an activity is triggered after m out of n branches have been completed (e.g., to improve response time, two databases are queried and the first result is processed, the second is ignored) [15]. Both determinants span a matrix of cases, with each cell giving the respective aggregation function. We define the set of all aggregation functions as:

$$X = \{x_{cp,p} \mid cp \in CP \wedge p \in P_s, x_{cp,p} : R^n \rightarrow \mathbb{R}, n \in N^+\} \quad (1)$$

P_s is the set of parameters of service s . n denotes the number of parameters to be aggregated; the domain of $x_{cp,p}$ consists of n -tuples with QoS parameters of the constituent services. Since the aggregation function depends on (p, cp) , there exist $5 \cdot 7 = 35$ aggregation functions. These assume that the parameters share the same unit of measurement (UoM). The literature provides mature methods for converting UoM [19]. In addition, an OWL ontology for these UoM is available [20].

Several pairs (p, cp) share the same aggregation function. Thus, we reduce the number of functions to only seven. We define generic aggregation functions $x \in X$ as shown in Table 1, with x_1, \dots, x_n denoting the parameter values of the services to be aggregated, and k for the number of iterations in a *Loop* pattern. To each pair (p, cp) , we assign the respective aggregation function. Following [14], we distinguish upper and lower bound (as shown in Table 2). This distinction is necessary to assess whether a particular parameter value meets the guarantee defined in the SLA (see section 3). Calculating expected or average parameter values is only possible, if the distribution of the non-deterministic control flows (i.e., XORXOR, ANDDISC, OROR, and ORDISC) is known [14]; however, this information is not available.

Table 1. Generic aggregation functions.

Aggregation function	Definition	Aggregation function	Definition
x_{sum}	$x_{sum}(x_1, \dots, x_n) = \sum_{i=1}^n x_i$	x_{power}	$x_{power}(x) = x^k$
$x_{product}$	$x_{product}(x_1, \dots, x_n) = \prod_{i=1}^n x_i$	x_{linear}	$x_{linear}(x) = kx$
x_{max}	$x_{max}(x_1, \dots, x_n) = \max(x_1, \dots, x_n)$	$x_{identity}$	$x_{identity}(x) = x$
x_{min}	$x_{min}(x_1, \dots, x_n) = \min(x_1, \dots, x_n)$		

Table 2. Aggregation functions for upper and lower bounds of five QoS parameter types.

Type	Bound	Sequence	Loop	XORXOR	ANDAND	ANDDISC	OROR	ORDISC
1	Upper	x_{sum}	x_{linear}	x_{max}	x_{sum}	x_{sum}	x_{sum}	x_{sum}
	Lower	x_{sum}	x_{linear}	x_{min}	x_{sum}	x_{sum}	x_{min}	x_{min}
2	Upper	x_{sum}	x_{linear}	x_{max}	x_{max}	x_{max}	x_{max}	x_{max}
	Lower	x_{sum}	x_{linear}	x_{min}	x_{max}	x_{min}	x_{min}	x_{min}
3	Upper	x_{min}	$x_{identity}$	x_{max}	x_{min}	x_{min}	x_{sum}	x_{sum}
	Lower	x_{min}	$x_{identity}$	x_{min}	x_{min}	x_{min}	x_{min}	x_{min}
4	Upper	$x_{product}$	x_{power}	x_{max}	$x_{product}$	$x_{product}$	x_{max}	x_{max}
	Lower	$x_{product}$	x_{power}	x_{min}	$x_{product}$	$x_{product}$	$x_{product}$	$x_{product}$
5	Upper	x_{min}	$x_{identity}$	x_{max}	x_{min}	x_{min}	x_{max}	x_{max}
	Lower	x_{min}	$x_{identity}$	x_{min}	x_{min}	x_{min}	x_{min}	x_{min}

4.2 Ontology Specification

The ontology formally defines the conceptualization of composition patterns, parameter types, aggregation functions, and their relations. We specify the ontology using description logics (DL), which is a family of formalisms for representing knowledge within a domain. DL provides high expressiveness, while it retains computational completeness and decidability [21]. This logic is also the basis of the Web Ontology Language OWL DL [22]. The methodology for creating the ontology is a rigorous deduction process from the content of table 1 and 2. Syntactical correctness is checked by implementing the ontology in OWL DL and performing standard checks that are built in the OWL editor used. Completeness is guaranteed by covering all the content of table 2.

Figure 2 gives an overview of the concepts and roles. The ontology consists of three concept hierarchies for parameter types P , composition patterns CP , and aggregation functions AF . The actual formula term is represented by F and the functional role $hasF$. For each parameter type concept, we add one sub-concept for commonly used parameters (i.e., cost, execution time, throughput, uptime probability, and encryption). These parameters constitute examples only.

The dependencies are expressed by restrictions over the two roles $forP$ and $forCP$. Table 3 summarizes these restrictions. For instance, a sequence of $Type1$ parameters is aggregated by sum; hence the concept Sum is extended by $Sum \equiv \exists forP.Type1 \sqcap \forall forCP.Sequence$. Since Sum is valid for more than one pair, the concept definition consists of several pairs of restrictions being concatenated by a logical OR (in DL signified by \sqcup); i.e., for $Type1$ it is valid for $Sequence$, $ANDAND$, and $ANDDISC$, whereas for $Type2$ it is valid for $Sequence$ only.

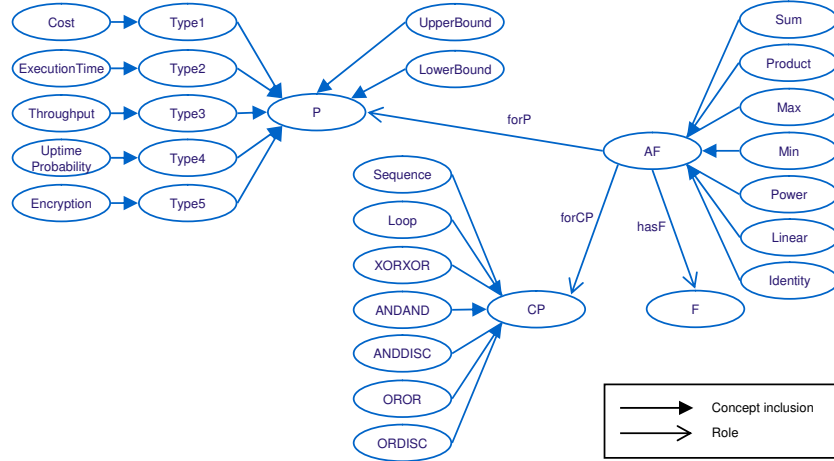


Fig. 2. QoS aggregation ontology (overview)

Table 3. Concept definitions for aggregation functions.

Concept	Definition
Sum	$\exists forP.Type1 \sqcap (\forall forCP.(Sequence \sqcup ANDAND \sqcup ANDDISC))$ $\sqcup (\exists forP.Type2 \sqcap \forall forCP.Sequence) \sqcup (\exists forP.(Type1 \sqcup Type3) \sqcap$ $\forall forCP.(OROR \sqcup ORDISC) \sqcap (\forall forP.UpperBound))$
Product	$\exists forP.Type4 \sqcap ((\forall forCP.(Sequence \sqcup ANDAND \sqcup ANDDISC))$ $\sqcup (\forall forCP.(OROR \sqcup ORDISC))) \sqcap (\forall forP.LowerBound)$
Min	$(\exists forP.(Type1 \sqcup Type2 \sqcup Type3 \sqcup Type5) \sqcap (\forall forCP.(XORXOR \sqcup OROR \sqcup ORDISC) \sqcap$ $(\forall forP.LowerBound))) \sqcup (\exists forP.(Type3 \sqcup Type5) \sqcap (\forall forCP.(Sequence \sqcup ANDAND \sqcup$ $ANDDISC))) \sqcup \exists forP.Type4 \sqcap \forall forCP.XORXOR \sqcap \forall forP.LowerBound) \sqcup (\exists forP.Type2 \sqcap$ $\forall forCP.ANDDISC \sqcap \forall forP.LowerBound)$
Max	$(\forall forCP.XORXOR \sqcap \forall forP.UpperBound) \sqcup (\exists forP.Type2 \sqcap \forall forCP.ANDAND) \sqcup$ $(\exists forP.Type2 \sqcap \forall forCP.(ANDDISC \sqcup OROR \sqcup ORDISC) \sqcap \forall forP.UpperBound) \sqcup$ $(\exists forP.(Type4 \sqcup Type5) \sqcap \forall forCP.(OROR \sqcup ORDISC) \sqcap \forall forP.UpperBound)$
Power	$\exists forP.Type4 \sqcap \forall forCP.Loop$
Linear	$\exists forP.(Type1 \sqcup Type2) \sqcap \forall forCP.Loop$
Identity	$\exists forP.(Type3 \sqcup Type5) \sqcap \forall forCP.Loop$

4.3 QoS Aggregation Algorithm

The aggregation is performed on workflow W . Algorithm 1 (Figure 3) is executed for W 's start node ns . If the workflow begins with an *XOR/AND/ORsplit*, all branches are processed recursively. The aggregated nodes are collected in the set N' (lines 01-05). The actual aggregation is performed based on the respective pattern using algorithm 2 (Figure 3) and the result is stored in the join nodes (lines 06-13). These nodes are used by algorithm 1 to detect subsequent *Sequence* and *Loop* patterns (line 14).

<p>Algorithm 1. $QoS(B(n)) = aggregateQoS(n)$ inputs: DAG node n outputs: aggregated QoS for node n 01 if $n \in \{ORsplit, ANDsplit, ORsplit\}$ then 02 $N' = \emptyset$ //nodes to be aggregated 03 foreach $n' \in N : (n, n') \in A :$ $n' \in T \cup \{XORsplit, ANDsplit, ORsplit\}$ 04 $QoS(B(n')) = aggregateQoS(n')$ 05 $N' = N' \cup \{n'\}$ 06 if $n = XORsplit$ then 07 $QoS(B(n)) = aggregateValues(XORXOR, N')$ 08 else if $n = ANDsplit$ then 09 $QoS(B(n)) = aggregateValues(ANDAND, N')$ 10 else // $n = ORsplit$ 11 $QoS(B(n)) = aggregateValues(OROR, N')$ 12 $n' = getJoinNode(n)$ 13 $QoS(B(n')) = QoS(B(n))$ // QoS for join node 14 $n = n'$ 15 if $\exists n' \in N : (n, n') \in A :$ $n' \in T \cup \{XORsplit, ANDsplit, ORsplit\}$ then 16 $QoS(B(n)) = aggregateValues(Sequence,$ $getSequenceNodes(n))$ 17 if $\exists n' \in N : (n, n') \in A \wedge n' = Loop$ then 18 $QoS(B(n)) = aggregateValues(Loop, \{n\})$ 19 return $QoS(B(n))$</p>	<p>Algorithm 2. $QoS(B(N')) = aggregateValues(cp, N')$ inputs: composition pattern $cp \in CP,$ set of DAG nodes N' outputs: aggregated QoS values of N' 01 foreach $p \in P$ 02 $myP = createIndividual(p)$ 03 $myCP = createIndividual(cp)$ 04 $myAF = createIndividual(AF)$ 05 $myAF.setProperty(forP, myP)$ 06 $myAF.setProperty(forCP, myCP)$ 07 foreach $f \sqsupseteq myAF$ 08 if $f \sqsubset AF$ then 09 break //end inner for loop 10 $SV(QoS(B(N')), p) = f(SV(QoS(n_i \in N'), p),$ $\dots, SV(QoS(n_{i+1} \in N'), p))$ 11 return $QoS(B(N'))$</p>
---	--

Fig. 3. Aggregation algorithms

The existence of a subsequent node in *DAG* results in the detection of *Sequence* patterns (line 15). However, determining the nodes that belong to a sequence is a non-trivial process, as sequence patterns are not made explicit in the *DAG*. Thus, we introduce the function *getSequenceNodes(n)* that performs this task. The function works as follows: For task nodes, non-nested *XORsplit*, *ANDsplit*, *ORsplit*, and subsequent task nodes are collected, whereas *XOR/AND/ORsplit* nodes are recursively aggregated using algorithm 1. For *XOR/AND/ORsplit* nodes, i.e., for the detection of nested *Sequence* patterns in *XORXOR*, *ANDAND*, as well as *XORXOR* patterns, *DAG* is traversed until the corresponding join node is reached for each branch. The nested patterns are also recursively aggregated using algorithm 1. *Loop* nodes indicate multiple executions of the preceding node. Algorithm 1 detects these patterns and performs the aggregation using algorithm 2 for the preceding node (algorithm 1, lines 17-18).

The concrete aggregation (i.e., calculation of the aggregated values) is performed by algorithm 2, which uses the aggregation ontology. For inferring the correct aggregation function, it first creates individuals for the parameter (line 02), composition pattern (line 03), and aggregation function (line 04). *myAF*, the instance of the latter, belongs to the general concept *AF* only, since we do not know the aggregation function yet. Next, we relate these three individuals by the roles *forP* and *forCP* (lines 05-06). Then, the knowledge base is queried for all concept memberships of *myAF*. DL reasoning returns three memberships (loop in line 07): *T* (the top concept to which all individuals belong), *AF* (asserted in line 4), and the specific one denoted by *f*. We use *f* for calculating the parameter value by considering all nodes of the input set *N'* (line 10). The QoS parameter aggregation is performed separately for each parameter (line

01). The function $SV: V \times P \rightarrow V$ reduces a vector of values V to a subvector of parameter values described by P . Finally, we return the aggregated QoS values.

5 Evaluation

We demonstrate the usefulness in a scenario of service composition. We assess the efficiency in a set of experiments using a prototype implementation.

5.1 Usefulness

We consider the following scenario: A designer has defined the workflow as shown in Figure 1. He has selected for each task one service and the workflow has been executed several times successfully. After some time, the execution fails due to the provider of the service for task t2, who is unable to guarantee the agreed service level. Therefore, this service must be replaced by a suitable service. By using a service discovery functionality (which is not subject of our proposal), three candidate services are found. These services have been described using custom parameters as shown in Table 4.

The problem with these services is that their QoS parameters are not aligned to a shared conceptualization. If a particular service is selected for t2, its parameters cannot be aggregated with the parallel service of task t3 and all subsequent tasks in the workflow. A potential solution for this problem would be that the designer rejects the service and asks the service providers to submit a revised service description, which conforms to the standard syntax of the designer; for example, replacing “Price” by “Cost”. The disadvantage of this approach is that the service composition process is broken and the designer must wait for the new service description to arrive.

Table 4. Example service offers.

Service provider	Service ID	Parameter	Value
SP1	S1	Cost	0.85
		ExecutionTime	500
SP2	S2	Price	0.75
		ResponseTime	600
	S3	Price	0.70
		ResponseTime	700

By using the proposed aggregation method, there are two options for solving the problem of heterogeneous QoS parameters. First, the designer could take the current service descriptions and add a semantic annotation to each parameter of service S1 through S3. Figure 4 illustrates such an annotation by showing the SLA for S2 specified in the WS-Agreement format. The original SLA offer is then amended by a reference to a concept from the ontology as follows: the attribute *sawsdl:modelReference* contains a mapping of the literal *ResponseTime* to the concept *ExecutionTime* of the ontology which is identified by the namespace *qosns*). Having made this annotation, the service parameters are aligned to the conceptualization that is also used in the

description of the other services in the workflow, in particular the services that are bound to task t3. Therefore, their QoS parameters can be aggregated and thus the stopped process of service composition can be continued. In this sense, the ontology-based aggregation method assists the designer in maintaining workflows and choosing substitutes for failed services. The concrete decision of what service to select, however, depends on the designer's preferences and is not affected by the aggregation method.

<pre> <wsag:AgreementOffer ...> ... <wsag:ServiceDescriptionTerm ... wsag:ServiceName="S2"> ... <wsla:SLAParameter name="ResponseTime" type="wsla:float" unit="ms"> ... </wsag:ServiceDescriptionTerm> ... </wsag:AgreementOffer> </pre>	<pre> <wsag:AgreementOffer ...> ... <wsag:ServiceDescriptionTerm ... wsag:ServiceName="S2"> ... <wsla:SLAParameter name="ResponseTime" type="wsla:float" unit="ms" sawsdl:modelReference=" &qosns;ExecutionTime"> ... </wsag:ServiceDescriptionTerm> ... </wsag:AgreementOffer> </pre>
--	--

Fig. 4. Example SLA offer by the service provider (left) and with annotation (right)

The second option assumes an ontological commitment of the providers of the services S1 through S3 to the ontology. In this case, each provider still uses his/her own syntax for service parameters, but provides a semantic annotation for each parameter as part of the service description. These annotations are the same as the ones shown in Figure 4. The difference is that annotations are made by the service provider. The annotation process is supported by virtually any XML tool, since the SA-SLA specification provides a standard-conform XML schema for SLA specifications [10].

The advantage of this option is that the process of service composition can be continued immediately by aggregating the QoS parameters under consideration for task t3 and the entire workflow. These aggregates are handed over to a service selection functionality, which selects the best service from a set of services; the selection is either performed by the designer or a component that automates this process. Particular methods for service selection are beyond the scope of our proposal. Existing service selection methods [23] can be combined with our proposal, which complements the selection problem by resolving the heterogeneity of QoS parameters.

5.2 Efficiency

Experimental Setup. We have implemented the proposed aggregation method in a Java prototype application. Workflows are stored in WS-BPEL and SLAs in SA-SLA format. SA-SLA constitutes a combination of WSLA [24] and WS-Agreement for the SLA schema and SAWSDL [25] for semantic annotations. Experiments are performed on Scientific Linux 5.5 on a machine with two Intel Xeon E5630 CPUs (4C, 2.53 GHz) and 16 GB RAM. The prototype is a single-threaded application though.

We study how the proposed aggregation method works for different types of composite services with regard to their complexity and coverage. Complexity is measured

by the following metrics: number of tasks ($|T|$), number of gateways ($|G|$), number of arcs ($|A|$), and number of parameters ($|P|$) in $QoS(CS)$. We consider three workflows $W1$, $W2$, $W3$ with 4, 8, and 12 tasks and bind one service to each task. Fig. 5 shows these workflows as BPMN processes. In BPMN, loop is not represented by a gateway node but a modification of the task node. Therefore, the number of gateways in CS is 3, 6, and 10, with number of arcs being 10, 19, and 31 respectively.

In the experiments, we instantiate each workflow multiple times for different SLAs. We describe each service by five parameters. The reason for this choice is to have full coverage of the ontology's concepts for parameters, i.e., $type1$ to $type5$. Coverage is thus measured by the percentage of the ontology's concepts that the QoS aggregation experiments use. Since we are interested in the effectiveness of the approach, each workflow covers all parameter concepts (5) and composition pattern concepts for *Sequence*, *Loop*, *XOR*, and *AND*. This coverage of CPs is higher than in [11] and [16], and equal to [18].

The experiments study two variations. The first variation is concerned with the workflow's complexity. We will study the influence of $|T|$ on the time required for QoS aggregation. By setting the range of this variable to $\{4, 8, 12\}$, the simulation yields three data sets that provide indications of the system's scalability. The second variation is concerned with the reasoner that is used, since ontology-based systems add computational effort for asserting and retrieving facts. We use the two reference Java ontology frameworks Jena 2.6.4 and OWL API 3.2.2 and three reasoners in the following setup: Jena with OWLMicroReasoner, Jena with Pellet 2.2.2, OWL API with HermiT 1.3.3, OWL API with Pellet 2.2.2, OWL API with FaCT++ 1.5.2.

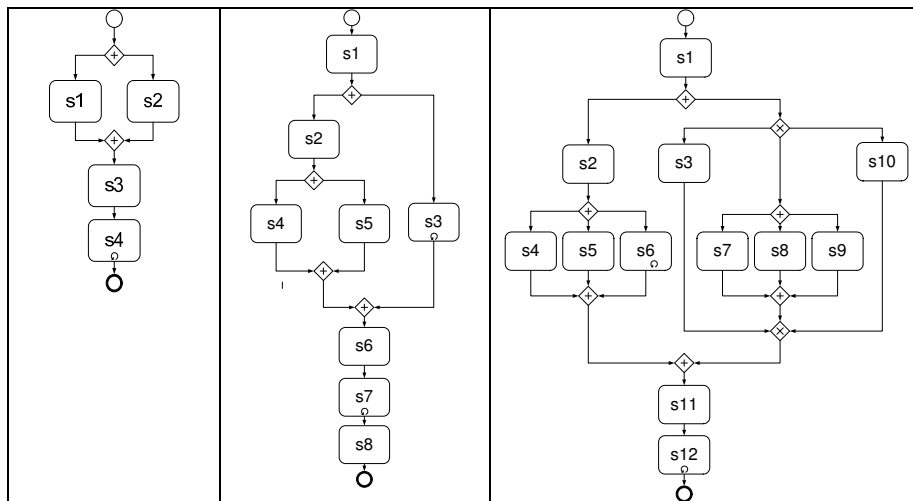


Fig. 5. BPMN workflows $W1$ (left), $W2$ (middle) and $W3$ (right)

We compare the ontology-based method with conventional aggregation, i.e., system with no reasoner, but explicit aggregation functions. This comparison is possible,

because the workflows and SLAs are known prior to execution and have been annotated with aggregation functions exclusively for conventional aggregation.

These two variations span a set of $3 \cdot 6 = 18$ experiments. In each experiment, we apply the aggregation method to five different SLAs per service and compute the QoS 50 times (thus 250 runs per experiment). We arrive at $18 \cdot 5 \cdot 50 = 4,500$ runs. The SLAs are generated prior to executing the experiments and values are assigned to each parameter (uniform distribution, interval $[1,100]$). The concrete values do not affect the method's performance, but are used to check the correctness of the composite QoS (effectiveness of the method).

Results. With regard to varying the number of tasks, Figure 6 shows the mean execution time for all 18 experiments. Each of the six configurations suggests a linear computation time. The figure also indicates that the baseline implementation of the aggregation method without reasoning is valid due to its linear complexity.

First, we assess the distribution of execution time in each experiment by using the following distribution properties: mean, median, minimal value (Min), maximal value (Max), standard deviation (STD), and coefficient of variation (CV). Table 5 shows these measures for the largest workflow *W3*. We observe the following: The data points tend to be very close to the mean (CV between 1.7 and 5.2%). Whereas Min and Max span a relatively wide interval, very few data points are close to Min and Max. These findings hold for all reasoner setups.

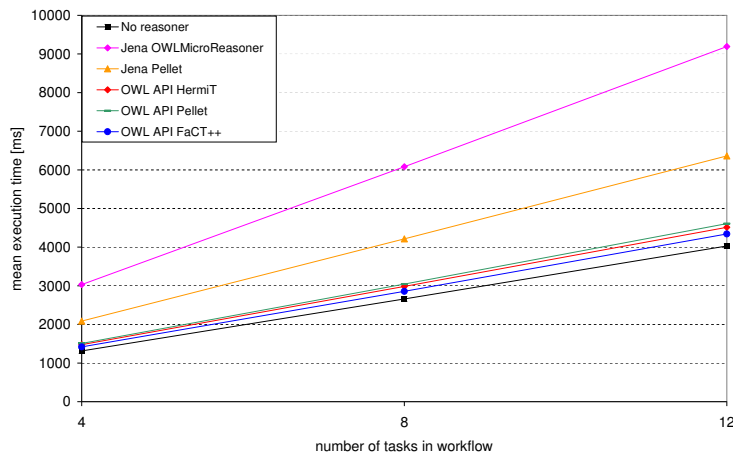


Fig. 6. Mean execution time for six reasoner setups

Second, we assess the efficiency of each setup by comparing their mean execution time to the fastest setup (no reasoner). Table 6 shows the relative increase caused by using the ontology: The increase ranges from $[0.097; 0.101]$ for OWL API FaCT++ to $[1.282; 1.1309]$ for Jena OWLMicroReasoner. These results lead to two important findings: On one hand, all OWL API setups outperform the Jena setups and require between 9.7% and 15.9% additional time compared to no reasoner. This increase is

rather small for an ontology-based method and could be attributed to efficient reasoner systems. On the other hand, the relative additional time does not increase with the number of workflow tasks (as shown by the series *W1*, *W2*, and *W3*). This finding is encouraging for ontology-based QoS aggregation and suggests that the approach is also efficient for workflows of higher complexity.

Table 5. Distribution properties of execution time in ms for workflow *W3*.

Reasoner setup	Mean	Median	Min	Max	STD
No reasoner	4,029	4,003	3,932	6,667	178.91
Jena Micro	9,193	9,157	9,013	13,457	282.81
Jena Pellet	6,363	6,325	6,183	9,062	212.64
OWL API HermiT	4,514	4,486	4,326	6,251	158.97
OWL API Pellet	4,608	4,580	4,508	5,212	87.25
OWL API FaCT++	4,340	4,307	4,231	4,745	85.82

Table 6. Relative increase of mean execution time due to OWL reasoning.

Reasoner setup	W1	W2	W3
Jena Micro	130.9%	128.8%	128.2%
Jena Pellet	60.5%	60.1%	59.5%
OWL API HermiT	13.4%	13.4%	13.2%
OWL API Pellet	15.9%	15.7%	15.5%
OWL API FaCT++	10.1%	9.7%	9.9%

6 Conclusion

We have presented a QoS aggregation ontology and an aggregation method that uses this ontology for composite Web services. We have shown the usefulness of our proposal in a use case scenario of service failure that requires to repair an existing composite service by substituting the failed service. In particular, the approach suggests that designers of composite services have more flexibility in choosing services that have not been aligned to standard QoS parameters. In this sense, the developed artifact helps in balancing the trade-off between standardization and flexibility, which arises from resolving semantic heterogeneity. The assessment of the method's efficiency using the prototype implementation suggests that adding an ontology into the aggregation process increases the computational effort by about 10% for the fastest reasoner setup (OWL API FaCT++). We argue that this increase is compensated by the higher flexibility available for both the service provider and the service consumer.

The results of this work are not limited to a specific domain, since QoS aggregation is a problem in any domain with composite services. The proposed ontology enables the providers to expose their capabilities to potential user, who request a particular service in the course of his/her service composition. Unlike current QoS ontologies, our approach does not rely on a comprehensive semantic description of Web services nor requires commitment to a Web service ontology. Adding annotations to existing SLAs does not affect conformance to the WS-Agreement specification.

The current QoS aggregation ontology contains one example parameter for each parameter type. It is important to note that these parameters represent examples only.

Using the ontology would typically start with linking application- or domain-specific parameters to one parameter type (by making the respective parameter a sub-concept of one of the type-*x* concepts). In our current work, we could have added many more parameters, however, the only added value is to help the user find the right parameter type. Such an extension of the ontology does not require any modification to the concepts *P*, *CP*, and *AF* including all role restrictions. The ontology, however, is limited to its five parameter types. If a new parameter type is needed, then a new sub-concept of *P* must be defined. In addition, the definitions of all *AF* sub-concepts must be extended with additional role restrictions.

Another limitation of our approach is that it requires all service parameters to be annotated. Even if one annotation is missing or incorrect, the entire QoS aggregation process will fail. This limitation is important in cross-organizational workflows incorporating autonomous service providers. A promising solution to circumvent this drawback is the integration of fault-tolerant or probabilistic [17] QoS aggregation methods into our approach.

Acknowledgments. The work of P. Karaenke and J. Leukel has been supported by the German Federal Ministry of Education and Research (BMBF) under the project InterLogGrid (FKZ 01IG09010E) and by the German Federal Ministry of Economics and Technology (BMWi) under the project MIGRATE! (FKZ 01ME11052). The work of V. Sugumaran has been supported by Sogang Business School's World Class University Program (R31-20002) funded by Korea Research Foundation and the Sogang University Research Grant of 2011.

References

1. Dustdar, S.: Web Services Workflows – Composition, Co-Ordination, and Transactions in Service-Oriented Computing. *Concurrent Eng. Res. A*, 12, 237–245 (2004)
2. Ardagna, D., Pernici, B.: Adaptive Service Composition in Flexible Processes. *IEEE T. Software Eng.* 33, 369–384 (2007)
3. W3C: Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language. W3C Recommendation 26 June 2007, <http://www.w3.org/TR/wsd120>
4. Open Grid Forum: Web Services Agreement Specification (WS-Agreement). OGF Proposed Recommendation, <http://www.ogf.org/documents/GFD.107.pdf>
5. Papazoglou, M.P., Traverso, P., Dustdar, S., Leymann, F.: Service-Oriented Computing: State of the Art and Research Challenges. *IEEE Computer* 40, 38–45 (2007)
6. Hevner, A., March, S., Park, J., Ram, S.: Design science in Information Systems Research. *MIS Quarterly* 28, 75–105 (2004)
7. Karaenke, P., Leukel, J.: Towards Ontology-based QoS Aggregation for Composite Web Services. In: Fähnrich, K.-P., Franczyk, B. (eds.): *Informatik 2010. LNI*, Vol. 175, pp. 120–125. GI, Bonn (2010)
8. O'Sullivan, J., Edmond, D., ter Hofstede, A.H.M.: What's in a Service? *Distrib. Parallel Dat.* 12, 117–133 (2002)
9. Dobson, G., Lock, R., Sommerville, I.: QoSOnt: A QoS Ontology for Service-Centric Systems. In: 31st EUROMICRO Conference on Software Engineering and Advanced Applications, pp. 80–87. IEEE Press, New York (2005)

10. Muñoz Frutos, H., Kotsiopoulos, I., Vaquero Gonzalez, L., Rodero Merino, L.: Enhancing Service Selection by Semantic QoS. In: Aroyo, L. et al. (eds.): ESWC 2009. LNCS, Vol. 5554, pp. 565–577. Springer, Heidelberg (2009)
11. Ul Haq, I., Huqqani, A., Schikuta, E.: Hierarchical aggregation of Service Level Agreements. *Data Knowl. Eng.* 70 (5), 435–447 (2011)
12. Cardoso, J., Sheth, A.P., Miller, J.A., Arnold, J., Kochut, K.: Quality of service for workflows and web service processes. *J. Web Semant.* 1, 281–308 (2004)
13. Zeng, L., Benatallah, B., Ngu, A.H.H., Dumas, M., Kalagnanam, J., Chang, H.: QoS-Aware Middleware for Web Services Composition. *IEEE Transactions on Software Eng.* 30, 311–327 (2004)
14. Jaeger, M.C., Rojec-Goldmann, G., Mühl, G.: QoS Aggregation for Web Service Composition using Workflow Patterns. In: 8th IEEE Enterprise Distributed Object Computing Conference (EDOC 2004), pp. 149–159. IEEE Press, New York (2004)
15. van der Aalst, W., ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow patterns. *Distributed and Parallel Databases* 14 (1), 5–51 (2003)
16. Canfora, G., Di Penta, M., Esposito, R., Perfetto, F., Villani, M.L.: A framework for QoS-aware binding and re-binding of composite web services. *J. Syst. Software* 81, 1754–1769 (2008)
17. Hwang, S.-Y., Wang, H., Tang, J., Srivastava, J.A.: Probabilistic approach to modeling and estimating the QoS of web-services-based workflows. *Inform. Sciences* 177, 5484–5503 (2007)
18. Unger, T., Mauchart, S., Leymann, F., Scheibler, T.: Aggregation of Service Level Agreements in the Context of Business Processes. In: 12th IEEE Enterprise Distributed Object Conference (EDOC 2008), pp. 43–52. IEEE Press, New York (2008)
19. Beaty, H.W.: Units, Symbols, Constants, Definitions, and Conversion Factors. In: Fink, D.G., Beaty, H.W. (eds.): *Standard Handbook for Electrical Engineers*. McGraw-Hill Professional, New York (2006)
20. Hodgson, R., Keller, P.J.: QUDT – Quantities, Units, Dimensions and Data Types in OWL and XML. <http://www.qudt.org> (2011)
21. Baader, F., Horrocks, I., Sattler, U.: Description Logic. In: Harmelen, F. van, Lifschitz, V. (eds.): *Handbook of Knowledge Representation*. Elsevier, Amsterdam (2007)
22. W3C: OWL Web Ontology Language. W3C Recommendation 10 February 2004, <http://www.w3.org/TR/owl-ref>
23. Alrifai, M., Risse, T.: Combining Global Optimization with Local Selection for Efficient QoS-aware Service Composition. In: 18th International Conference on World Wide Web (WWW 09), pp. 881–890. ACM, New York (2009)
24. Keller, A., Ludwig, H.: The WSLA Framework: Specifying and Monitoring Service Level Agreements for Web Services. *J. Netw. Syst. Manag.* 11, 57–81 (2003)
25. Kopecky, J., Vitvar, T., Bournez, C., Farrell, J.: SAWSDL: Semantic Annotations for WSDL and XML Schema. *IEEE Internet Comput.* 11, 60–67 (2007)

Eine Bestandsaufnahme von Standardisierungspotentialen und -lücken im Cloud Computing

Robin Fischer¹, Christian Janiesch¹, Joachim Strach²,
Nicolai Bieber², Wolfgang Zink², und Stefan Tai¹

¹ Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany
{robin.fischer, christian.janiesch, stefan.tai}@kit.edu

² Booz & Company, München, Germany
{joachim.strach, nicolai.bieber, wolfgang.zink}@booz.com

Abstract. Die Standardisierung im Cloud Computing ist erst im Entstehen begriffen. Sie gewinnt jedoch zunehmend an Eigendynamik. Bisherige Standardisierungsbemühungen stecken konzeptionell in den Kinderschuhen, da uneinheitliche Definitionen und fehlendes Orientierungswissen ein zielorientiertes Handeln behindern. Die vorliegende Arbeit schlägt deshalb eine konsistente Taxonomie für die strukturierte Betrachtung und begriffliche Eindeutigkeit bei der Beschreibung und Bewertung von Standards vor. Darauf aufbauend wird ein Vorgehensmodell zur Analyse der aktuellen Standardisierungslage vorgestellt. Dieses verwendet eine Standardisierungslandkarte, die das Forschungsfeld anhand der Dimensionen Herausforderungen und Ansatzpunkte aufspannt. Die vorgenommene Analyse erfasst gegenwärtige Standardisierungspotentiale und -lücken im Cloud Computing. Die abschließend vorgenommene Bewertung zeigt Handlungsoptionen künftiger Standardisierungsbemühungen auf.

Keywords: Cloud Computing, Standards, Übersicht, Potentiale, Lücken

1 Einleitung

Cloud Computing ermöglicht den bedarfsgerechten Bezug von Speicherkapazitäten, Rechenleistung und Anwendungen über das Internet. Der Einsatz von Cloud-Diensten ist dabei mit einer nutzungsbezogenen Abrechnung der verwendeten Ressourcen gekoppelt, so dass Investitionskosten reduziert werden können [1].

Anwender von Cloud Computing sehen sich bei der Auswahl und Bewertung von passenden Cloud-Diensten einer Reihe von Herausforderungen ausgesetzt (vgl. bspw. [2]). Das Fehlen einer einheitlichen Begriffswelt, die über die Unterteilung von Cloud Computing über Anwendungs-, Plattform- und Infrastruktur-Dienste hinausgeht, erschwert es Anwendern einzuschätzen, inwieweit ein Cloud-Dienst ihre Bedürfnissen gerecht wird [3]. Das liegt auch daran, dass es nur vereinzelt Standards gibt, die zentrale Aspekte des Cloud Computing, wie bspw. Protokolle und Schnittstellen, Service Level Agreements (SLA) oder rechtliche Vorgaben einheitlich regeln und damit für einen Vergleich herangezogen werden könnten. Dies hemmt die Akzeptanz von

Cloud-Diensten bei potentiellen Anwendern, insbesondere im Mittelstand, und behindert die Entwicklung von interoperablen und offenen Cloud-Diensten.

Wir haben durch die Sichtung von etwa 160 potentiell für das Cloud Computing relevanten Standards sowie durch eingehende Betrachtung einer Auswahl davon eine Standardisierungslandkarte erstellt. Diese katalogisiert die nach unserer Erhebung relevantesten Standards. Die so nach einheitlichen Kriterien erstellte Übersicht für Standards im Cloud Computing, soll Anwender von Cloud-Diensten bei der Auswahl der für ihre Bewertung von Cloud-Diensten relevanten Standards leiten und damit einen standardisierten Vergleich von Cloud-Diensten ermöglichen. Wir haben darüber hinaus eine Potential- und Lückenanalyse durchgeführt, die anhand der entwickelten Standardisierungslandkarte weiteren Standardisierungsbedarf aufzeigt.

Im Folgenden erläutern wir kurz das der Arbeit zugrundeliegende Verständnis von Cloud Computing und Standards und gehen auf verwandte Untersuchungen ein. In Kapitel 3 erläutern wir ausführlich das Vorgehen bei der Erhebung der Standards, den Aufbau der Standardisierungslandkarte sowie das Vorgehen bei der Potential- und Lückenanalyse. Die überblicksartige Einordnung der Standards und eine zusammenfassende Diskussion der Potentiale und Lücken finden in Kapitel 4 statt. Kapitel 5 fasst die Ergebnisse zusammen und versucht Handlungsempfehlungen zu geben.¹

2 Definitionen und verwandte Arbeiten

2.1 Cloud Computing

Das Zusammenspiel von Infrastrukturkomponenten (Server, Speicher, Netze, Middleware) und verfügbaren Diensten erscheint dem Anwender als *Wolke* möglicher Computer- und Kommunikationsanwendungen, wodurch der Begriff des Cloud Computing geprägt wurde. Cloud Computing bietet demnach die Möglichkeit, Speicherkapazitäten, Rechenleistung und Anwendungen nach kundenspezifischen Bedarfen als Dienst über das Internet zu beziehen. Die so ermöglichte bedarfsgerechte, skalierbare und flexible Nutzung von IT-Diensten wird unterstützt durch neuartige Geschäftsmodelle, bei denen die Abrechnung je nach Funktionsumfang, Nutzungsdauer und Anzahl der Nutzer erfolgt [1].

Als Servicemodelle werden Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) und Software-as-a-Service (SaaS) unterschieden (ausführliche Informationen dazu finden sich bspw. bei [1], [4]). Unter Betriebs-, Eigentums- und Organisationsaspekten können Private Clouds (für eine geschlossene Nutzergruppe) und Public Clouds (für eine große Anzahl verschiedener Nutzer) unterschieden werden. In

¹ Die Studie „Das Normierungs- und Standardisierungsumfeld von Cloud Computing“ wurde im Rahmen des Technologieprogramms Trusted Cloud des Bundesministeriums für Wirtschaft und Technologie (BMWi) durchgeführt und ist unter <http://www.trusted-cloud.de/> abrufbar. Dieser Artikel derselben Autoren stellt eine aktualisierte Fassung dar, die das wissenschaftliche Vorgehen in den Vordergrund stellt. Bei der ausführlichen Studie liegt der Fokus mehr auf der Einzelbewertung von Standards und Standardisierungsinitiativen.

der Realität finden sich auch Nutzungskombinationen (Hybrid Clouds) von Private Clouds, Public Clouds und traditionellen on-premise IT-Umgebungen [4].

2.2 Standardbegriff

Die Anwendung von Standards ist freiwillig. In ihnen selbst liegt keine Verbindlichkeit. Eine Anwendungspflicht kann sich jedoch aufgrund von Gesetzen, Rechts- oder Verwaltungsvorschriften, Verträgen oder sonstigen Rechtsgründen ergeben. In der Praxis können Standards zudem durch weichere Steuerungsinstrumente bzw. ihre Bedeutung im Markt Verbindlichkeit entfalten. Wir betrachten im Folgenden nicht nur Standards im engeren Sinn, die durch ein Standardisierungsgremium verabschiedet werden [5], sondern auch Vorarbeiten, Vorgaben und Zertifizierungen.

Der Begriff Cloud-Standards steht dabei vereinfachend und als Sammelbegriff für Orientierungswissen, Spezifikationen, Standards im engeren Sinne, Normen, und rechtliche Vorgaben (Rechtsnormen). In engem Bezug zu Standards stehen (Referenz-)Implementierungen. Diese umfassen etablierte Cloud-Dienste, APIs, Testbeds oder Plugfests. Für Standards sind sie von vorbereitender oder ergänzender Natur, die die Praktikabilität im Einsatz zeigen oder andeuten. Referenzimplementierungen basieren in der Regel auf einer Spezifikation für einen Standard (auch Entwürfen hierfür) oder einem Standard.

Neben der Klassifizierung von Standards steht der Begriff der *Zertifizierung* in einem orthogonalen Zusammenhang: Anbieter und Anwender von Cloud-Diensten, können zur Bestätigung der Einhaltung bestimmter Kriterien, Standards oder rechtlichen Vorgaben *Zertifikate* erwerben. Die Aussagekraft von Zertifikaten wie auch die Nachvollziehbarkeit des Vorgangs der Zertifizierung steigt dabei mit dem Grad der Formalisierung und Verbindlichkeitswirkung.

2.3 Andere Arbeiten zu Cloud-Standards

Jeffery and Burkhard Neidecker-Lutz [6] beschreiben offene Herausforderungen sowie den gegenwärtigen State-of-the-Art im Cloud Computing. Dabei erfolgt auch eine Priorisierung von Handlungsfeldern zur Adressierung der Herausforderungen. Die Smart Cloud Study Group [7] beschreibt Herausforderungen und Potentiale des Cloud Computing im Allgemeinen.

Frameworks der European Network and Information Security Agency (ENISA) [8] sowie der Cloud Security Alliance (CSA) [9] fokussieren Sicherheitsaspekte im Cloud Computing. EuroCloud [10] bietet einen weiteren Überblick und Lösungsansätze. Weiterhin gibt die ENISA [11] einen Überblick über Erwartungen und Hindernisse des Cloud Computing speziell für den europäischen Mittelstand. In der Cloud Control Matrix (CCM) [12] werden für das Management von Cloud-Diensten relevante Herausforderungen (engl. *control areas*) identifiziert und Empfehlungen zur Ausübung der Steuerung (engl. *control specification*) dargestellt. Die CSA [13] nennt weitere Herausforderungen beim Management von Cloud-Diensten und setzt diese in Bezug zu Herausforderungen beim Management von IT-Systemen aus bspw. COBIT 4.1, ISO 27001 oder NIST SP800. Auch die Cloud Computing Use Case Discussion

Group [14] hat Anwendungsfälle für Cloud Computing erarbeitet. Einen weiteren kurzen Überblick von Standards gegliedert nach Forschungsfeldern (engl. *study areas*) verschafft [15].

Das National Institute of Standards and Technology (NIST) [16] analysiert die Standardisierungslandschaft im Cloud Computing und leitet Handlungsempfehlungen zur Standardentwicklung und -adaption ab. Zur Bewertung und Klassifizierung der Standards werden hier konkret auf die öffentliche Verwaltung der USA zugeschnittene Szenarien verwendet [17]. Die Internet Engineering Task Force (IETF) [18] gibt einen Überblick zu aktuellen Standardisierungsinitiativen und Standards. Eine Taxonomie wird dabei nicht verwendet. Das vom NIST initiierte Cloud Standards Wiki² listet existierende Cloud-Standards ohne diese zu klassifizieren. Das zur Erstellung der Liste verwendete methodische Vorgehen ist nicht dokumentiert. Das ITU Telecommunication Standardization Bureau [19] bietet eine Auflistung von Aktivitäten zur Cloud-Standardisierung. Auch hier wird keine Taxonomie zur Kategorisierung der gelisteten Standards verwendet.

3 Methodischer Teil

3.1 Auswahlverfahren der betrachteten Standards

Die Untersuchung der Standardisierung im Cloud Computing basiert auf einer umfassenden Primär- und Sekundärrecherche der in Abschnitt 2.3 angeführten Arbeiten sowie Cloud-Standards entsprechend der obigen Definition. Ergänzend dazu wurden auch aktuelle Forschungsergebnisse gesichtet, insb. aus dem 7. Forschungsrahmenprogramm der EU, um ein umfassenderes Bild aktueller Entwicklungsströme zu bekommen. Die dieser Arbeit zu Grunde liegende Untersuchung war maßgeblich für das Standardverständnis des Technologieprogramms *Trusted-Cloud*³, in dessen Rahmen sie durchgeführt wurde. Zwischenergebnisse der Studie wurden in Interviews mit Experten aus Trusted-Cloud-Projekten validiert und wurden durch die Analyse der durch Fragebögen in der Trusted-Cloud-Community erhobenen Daten substantiiert. Die Ergebnisse der Untersuchung wurden abschließend in Workshops zur Diskussion gestellt. Der Teilnehmerkreis umfasste dabei nationale Standardisierungsbeteiligte aus dem Trusted-Cloud-Programm.

Die Analyse begann mit einer breiten Sichtung möglicher Akteure, Initiativen und Standards im Cloud Computing. Der verfolgte Analyseansatz besteht aus vier Schritten: Fokus, Auswahl, Einzelbewertung und zusammenfassende Bewertung.

Zunächst fanden eine Festlegung des Fokus der Untersuchung und eine Eingrenzung der Rechercheergebnisse statt, die näher betrachtet werden. Wir unterscheiden Standards, die einen expliziten Bezug zum Cloud Computing haben, von Standards, die einen impliziten Bezug zu Cloud Computing haben, d. h. einen Geltungsbezug zu Basistechnologien oder zu Grundprinzipien des Cloud Computing. Vollständig außer-

² <http://cloud-standards.org/>

³ <http://www.trusted-cloud.de/>

halb des Fokus dieser Untersuchung sind Standards, die weder einen expliziten noch impliziten Bezug zum Cloud Computing aufweisen. Weiterhin unterscheiden wir Standards, die nur für spezifische Branchen von Relevanz sind, sowie Standards, die ein branchenübergreifendes Wirkungsfeld haben. Der Fokus der Untersuchung liegt auf branchenübergreifenden Standards, die einen expliziten Bezug zu Cloud Computing besitzen. Insgesamt wurden so etwa 160 potentielle für das Cloud Computing relevante Standards identifiziert und klassifiziert.

Es folgte die Festlegung transparenter Kriterien zur weiteren Auswahl. Die betrachteten Standards sollten dabei ein möglichst großes inhaltliches Spektrum abdecken. Wo möglich, wurden möglichst umfangreiche Standards mit prototypischem Charakter von bekannten Standardisierungsorganisationen bevorzugt. Das Spektrum wurde durch die Herausforderungen und Ansatzpunkte gemäß der im folgenden Abschnitt definierten Taxonomie aufgespannt. In der Untersuchung wurde die Auswahl von Cloud-Standards so auf 21 zu bewertende Standards eingegrenzt (vgl. Tabelle 1).

Danach fand die Einzelbewertung der ausgewählten Standards unter Verwendung einheitlicher Bewertungskriterien statt. Wir untersuchten hierbei Reifegrad, Durchsetzungsfähigkeit und Partizipationsmöglichkeit. Die Einzelbewertung findet sich aus platzgründen nicht in dieser Publikation.

3.2 Erstellung einer Standardisierungslandkarte

Auswahl relevanter Herausforderungen der Standardisierung. Bei der Analyse der oben genannten Dokumente haben wir neun übergeordnete Herausforderungen bestimmt. Insbesondere aus der Sicht der Anbieter von Cloud-Diensten ist die Sicherstellung von Effizienz bei der Dienstbereitstellung von zentraler Bedeutung [6], [11], [14]. Überwiegend aus der Sicht von Cloud-Anwendern sind neben der Effektivität der Dienstnutzung und -steuerung übergreifende Anforderungen an Transparenz, Informationssicherheit und Datenschutz wichtig [6], [8-11], [14]. Letzterer wurde aufgrund seiner Bedeutung für rechtskonformes Cloud Computing in Deutschland und der vielfältigen Anforderungen jenseits der klassischen Informationssicherheit (z. B. Datensparsamkeit, Auskunftsrechte etc.) bewusst gesondert betrachtet. Darüber hinaus sind Interoperabilität sowie Portabilität von Daten und Diensten zwischen verschiedenen Cloud-Anbietern wichtige Themen [6-7], [11], [14].

Eine Sonderrolle kommt zwei aus übergeordneten Interessen getriebenen Herausforderungen zu: Sicherstellung eines funktionierenden Wettbewerbs [6-7], [11] und Sicherstellung von Compliance [6-7], [14]. Beide Herausforderungen stehen nicht auf derselben Ebene, wie die zuvor genannten. So fördert bspw. erhöhte Portabilität und Interoperabilität aber auch verbesserte Transparenz den Gedanken des funktionierenden Wettbewerbs. Datenschutz hingegen ist wiederum ein Teilbereich der Compliance. Eine tabellarische Übersicht der Herausforderungen findet sich in den Spaltenüberschriften der Standardisierungslandkarte in Tabelle 2.

Standards können die *Effizienz der Dienstbereitstellung* von Cloud-Diensten erhöhen. Hier können vier Bereiche unterschieden werden: Die *Nutzung von Entwicklungstools und -komponenten* hilft, aufwändige Eigenentwicklungen im Entwicklungsprozess zu vermeiden. Der *Aufbau skalierbarer Architekturen* stellt eine weitere

Herausforderung dar. Hier müssen insbesondere Eigenschaften wie Redundanz, Fehlertoleranz und Mehr-Mandantenfähigkeit berücksichtigt werden. Eine weitere Herausforderung beim effizienten Betrieb ist *Ressourcenmanagement und Flexibilität*. Dies bezieht sich zunächst auf technische Ressourcen wie Hardware und Software aber auch auf Personalressourcen im Sinne der Kapazitätsplanung und Standardisierung von Qualifikationsanforderungen. Gerade im Cloud-Kontext sind die Erwartungen an die *Verfügbarkeit der Dienste* sehr hoch. Sie hängen von diversen Einflussfaktoren ab (z. B. Rechenzentrumshardware, Netzverfügbarkeit, Anwenderverhalten).

Zur Erhöhung der *Effektivität der Dienstnutzung und -steuerung* werden Lösungen benötigt, die den Dienst an sich bereitstellen (z. B. HTML, Remote-Desktop Protokolle, Standards für Shell-Zugriffe, Internet-Standards). Da diese aber allgemein verfügbar und meist standardisiert sind, werden sie im Weiteren nicht separat betrachtet. Eine höhere Relevanz kommt den folgenden Herausforderungen zu: Vor der eigentlichen Nutzung von Cloud-Diensten sind etwaige Fragen zur *Vertragsgestaltung inkl. Haftungsfragen* zu klären. Während der Nutzung ist einer der Haupterfolgskriterien die Möglichkeit zur eigenständigen *Steuerung der Dienste durch den Anwender*. Im Falle von Problemen während der Nutzung sollte es standardisierte *Governance- und Eskalationsmechanismen* geben.

Die *Transparenz der Leistungserbringung und Abrechnung* vereinfacht die komplexe und oft eher anonyme Auftraggeber-Auftragnehmerbeziehung im Cloud Computing. Konkret bestehen die Anforderungen, dass die *Abrechnung inkl. Lizenzmanagement* transparent für den Anwender ist. Weiterhin sollte Transparenz bzgl. der Leistungsseite bestehen (z. B. durch SLA-Monitoring). Wir subsumieren solche Anforderungen im Bereich *Qualitätssicherung und Überwachung SLA*. Obwohl es dem Kerngedanken des Cloud Computing widerspricht, kann es aufgrund datenschutzrechtlicher und anderer Vorschriften in bestimmten Branchen von hoher Relevanz sein, Transparenz über *Art und Ort der Datenverarbeitung* zu erhalten.

Informationssicherheit wird oft als das Haupthindernis für eine schnelle Verbreitung von Cloud Computing angeführt. Die Verwaltung einer potenziellen Vielfalt von Identitäten sowie die Konfiguration eines effizienten *Identitäts- und Rechtemanagement* für Cloud-Dienste, insbesondere bei der Verwendung einer föderierten Cloud-Architektur, könnte durch einheitliche Standards ermöglicht werden. Im Cloud Computing werden Daten verschiedener Akteure verarbeitet. Daraus resultieren unterschiedliche Anforderungen in Bezug auf das benötigte Niveau an *Vertraulichkeit und Integrität*. Es gilt den gesamten Lebenszyklus von Daten zu betrachten. Dies beginnt bei der technischen Übermittlung und Speicherung und endet erst bei der endgültigen Löschung. Themenbereiche sind u. a. Verschlüsselung und Schlüsselmanagement, anonymisierte Datenverarbeitung etc. *Zugriffsschutz, Logging, Abwehr von Angriffen* ermöglicht eine sichere Trennung von Mandanten sowie Zugriffskontrolle, die sichere Identifizierung und Authentisierung umfasst. Die gezielte Autorisierung sowie zugehörige Protokollierungsmechanismen sind wesentliche Funktionen für funktionierende Informationssicherheit. Darüber hinaus sollten Standards einen Beitrag zum zuverlässigen Schutz vor Angriffen bereitstellen. Neben der rein technischen Erfüllung von Sicherheitsanforderungen sind *Nachweis und Zertifizierung* der IT-Sicherheit von ebenso großer Bedeutung.

Datenschutz als Schutz von personenbezogenen, personenbeziehbaren und sensiblen Daten vor Missbrauch, ist gerade in Deutschland eine der größten Herausforderungen im Kontext des Cloud Computing. Die Sicherstellung der Datenschutz-Compliance bspw. durch geeignete Anbietersauswahl, regelmäßige Kontrolle oder Einforderung einer transparenten Dokumentation stellt viele Unternehmen vor eine große Herausforderung.

Cloud-*Interoperabilität* wird aus drei Gesichtspunkten betrachtet: Unter *Migration in die bzw. aus der Cloud* werden Fähigkeiten benötigt, die es ermöglichen Infrastruktur-, Middleware- oder Anwendungskomponenten sowie vollständige Anwendungen und Daten in die Cloud zu verlagern oder aus der Cloud zu entfernen. *Integrationsfähigkeit in on-premise IT* erlaubt die Interoperabilität von on-premise Systemen und Cloud-Diensten in Form einer Hybrid Cloud. *Cloud-Föderation* umschließt die Fähigkeit Cloud-Dienste unterschiedlicher Ebenen und Anbieter verlässlich und oft ad-hoc miteinander verbinden zu können. Eine Grundlage hierfür können einheitliche oder kompatible Schnittstellen bilden, so dass keine individuelle Integration von Cloud-Diensten notwendig ist.

Zur Vermeidung von Lock-in Effekten ist die *Portabilität zwischen Anbietern* notwendig. So können Dienste und Daten unter Verwendung von einheitlichen Standards einfach und auf regulärer Basis zwischen Anbietern portiert werden. *Dienst-Portabilität* beschreibt Fähigkeiten zur Portierung von Cloud-Diensten. *Daten-Portabilität* beschreibt entsprechend Fähigkeiten zur Portierung von Daten in der Cloud. Standards könnten u. a. einheitliche Datenformate und Exit-Vereinbarungen z. B. mit Datenintegritätszusicherung und Kostenanzeige umfassen.

Aufgrund der Skaleneffekte seitens der Anbieter und potenzieller Lock-in-Effekte, besteht im Cloud Computing die Gefahr einer Beeinträchtigung des Wettbewerbs zwischen den Anbietern und die Herausbildung von marktbeherrschenden Akteuren. Gerade in Deutschland und Europa ist die *Sicherstellung eines funktionierenden Wettbewerbs* von zentraler Bedeutung, der auch die Teilhabe von mittelständischen Unternehmen gewährleistet.

Unter *Compliance mit geltender Rechtslage* wird die Einhaltung von Gesetzen und Richtlinien sowie freiwilliger Vereinbarungen verstanden. Dies ist insb. eine Herausforderung, da der Nutzer von Cloud-Diensten nur eine geringe Transparenz über die Regeleinhaltung des Anbieters hat. Besonders relevant ist dies in den Bereichen IT-Sicherheit, Datenschutz sowie im kommerziellen Bereich.

Auswahl von Ansatzpunkten der Standardisierung. Standards können über unterschiedliche Mittel eine Standardisierung herbeiführen (vgl. auch [7]). Es lassen sich im Cloud-Umfeld die drei grundlegenden Bereiche Technik, Management und Recht unterscheiden. Analog zur Vorgehensweise bei den Herausforderungen haben wir die Ansatzpunkte auf einer zweiten Ebene differenziert. Eine tabellarische Übersicht findet sich in den Zeilenüberschriften der Standardisierungslandkarte in Tabelle 2.

Der Bereich *Technik* umfasst technische Standards. Konkret beinhaltet dies folgende Aspekte: *Datei- und Austauschformate* zur Übermittlung und Speicherung von (teil-)strukturierten Daten. Dies können neben Dokumenten, Bildern oder Mediendateien auch Virtual Machine Images oder ganze Anwendungen sein. *Programmiermo-*

delle bilden die Grundlage für Erstellung und Ausführung von Quellcode. Über die Vorgabe von Programmierkonzepten und -abstraktionen werden die Bausteine von Programmiersprachen definiert. *Protokolle und Schnittstellen* beschreiben einen dynamischen Ablauf zum Austausch von Information zwischen zwei Komponenten, Anwendungen oder Akteuren. *Standardkomponenten und Referenzarchitekturen* erleichtern den Aufbau und die Verwendung von Cloud-Infrastrukturen und Cloud-Diensten. Durch standardisierte Designvorgaben wie bspw. Referenzarchitekturen können Best Practices auf eigene Cloud-Dienste übertragen werden. *Benchmarks und Tests* helfen die Leistungsfähigkeiten unterschiedlicher Cloud-Dienste, bspw. durch die Vorgabe von Lastprofilen und Kennzahlen, zu bemessen und zu beurteilen.

Der Bereich *Management* beinhaltet Standards, die die kommerzielle Abwicklung sowie das Management auf der Seite von Cloud-Anbietern und Cloud-Anwendern unterstützen. *Geschäftsmodelle* bilden die Grundlage für den wirtschaftlichen Betrieb von Cloud-Diensten. Im Bereich des Cloud Computing ist insbesondere der Bereich einer einheitlichen Leistungsbeschreibung für die Standardisierung relevant. *Service Level Agreements (SLA)* erhöhen die Effizienz in der Vertragsgestaltung und erlauben die Festlegung und Sicherstellung gezielter Anforderungen an den Diensteanbieter. SLAs können so signifikant zur Vertrauensbildung beitragen. *Vertragsbedingungen* umfassen Rahmenverträge, die durch SLAs ergänzt werden, Endbenutzer-Lizenzvereinbarungen oder Vertragsbausteine in unterschiedlichen Sprachen. *Managementmodelle und -prozesse* (z. B. im Sinne der im IT-Servicemanagement weit verbreiteten ITIL-Bibliotheken) können helfen einheitliches Vorgehen und Begrifflichkeiten sicherzustellen und Best Practice Prozesse zu fördern. *Controllingmodelle und -prozesse* können durch Vorgaben, bspw. zur Abrechnung und Rechnungslegung, oder dem Risikomanagement von bspw. IT-Systemen einheitliche Dokumente zur Dokumentation der Geschäftstätigkeit fördern. Dies kann ggfs. zur Vereinfachung von Zertifizierungen beitragen. *Leitfäden, Audit etc.* können sowohl potenziellen Anbietern als auch Nutzern von Cloud-Diensten hilfreich sein. Dies wird durch den Transport von Orientierungswissen, bspw. in Form von Best Practice, ermöglicht.

Regelungen und Vorschriften, die den geltenden Rechtsrahmen für die Akteure im Cloud Computing abstecken, können in drei Gruppen des Bereichs *Recht* unterschieden werden. *Rechtliche Vorgaben* stellen verbindliche Vorgaben, die auf entsprechenden Gesetzen, Richtlinien, Verordnungen o. ä. basieren. *Selbstverpflichtungen* fassen freiwillige Vereinbarungen und Kodizes zusammen, die z. B. von (Branchen-) Verbänden herausgegeben werden. *Unternehmensrichtlinien* stellen die schwächste Form von Regelungen und Vorschriften zum geltenden Rechtsrahmen dar. Unternehmensrichtlinien erlauben Unternehmen, sich bei der Geschäftstätigkeit bspw. strengeren Richtlinien zu unterwerfen als durch rechtliche Vorgaben gefordert.

3.3 Analyse von Standardisierungspotentialen und -lücken

In den vorherigen Schritten zur Bewertung des Portfolios von Cloud-Standards haben wir eine durch Herausforderungen und Ansatzpunkte aufgespannte Standardisierungslandkarte erstellt. Darin haben wir die Auswahl von 21 Standards eingeordnet. Die weitergehende Analyse und Bewertung dieser Ergebnisse erlaubt die Identifikation von Standardisierungslücken sowie die Ableitung von Handlungsempfehlungen für die Standardisierung.

Wir haben eine begleitende Potentialanalyse durchgeführt, die allgemeine Beiträge der Standardisierung identifiziert. Dabei wurde die Frage untersucht, inwieweit ein Ansatzpunkt (also beispielsweise ein Datei- und Austauschformat) einen Beitrag zur Lösung einer jeweiligen Herausforderung leisten kann. Die Bestimmung des allgemeinen Lösungsbeitrags wurde dabei auch durch Experten-Interviews sowie den Einsatz von Fragebögen unterstützt. Darin wurde allgemein erhoben, welche Relevanz eine Herausforderung für das Cloud Computing aufweist. Zusätzlich haben wir eine grundsätzliche Einschätzung des Beitrags der Ansatzpunkte zur Lösung der Herausforderung durch Standards erhoben. In beiden Fällen wurden Antwortmöglichkeiten auf einer Likert-Skala mit vier Ausprägungen erhoben:

- *Ja*: Der Ansatzpunkt hat augenscheinliches und umfangreiches Potential zur Lösung der Herausforderung.
- *Eher ja*: Der Ansatzpunkt hat grundsätzliches Potential zur Lösung der Herausforderung.
- *Eher Nein*: Der Ansatzpunkt hat geringes oder nicht direkt ersichtliches Potential zur Lösung der Herausforderung.
- *Nein*: Der Ansatzpunkt hat voraussichtlich kein Potential zur Lösung der Herausforderung.

Die Analyse von Standardisierungslücken verbindet die Potentialanalyse mit der vorgenommenen Bewertung existierender Standards. Dazu wurde eine Standardisierungslücke als vorhanden klassifiziert, wenn tendenziell Potential („eher ja“) zur Lösung einer Herausforderung im Cloud Computing durch einen Standard vorhanden ist, existierende Standards dieses aber noch nicht vollständig ausschöpfen. Weiterer Bedarf zur Standardisierung wurden als „erhöht“ oder „hoch“ klassifiziert werden, wenn Potential der Standardisierung offensichtlich („ja“) ist, existierende Standards dieses aber noch nicht vollständig bzw. unzureichend ausschöpfen. Neben der Bewertung des Lösungsbeitrags vorhandener Standards zur Adressierung der jeweiligen Herausforderung durch Standardisierung eines Ansatzpunktes, wurde auch eine Einschätzung der Dringlichkeit der Standardisierung für jede Herausforderung vorgenommen. Dies ermöglicht die Priorisierung von Handlungsempfehlung. Identifizierte Standardisierungsbedarfe der Kategorie „hoch“ zeichnen sich durch tendenziell kurzfristige Dringlichkeit aus. „Erhöhte“ Bedarfe und „vorhandene“ Lücken sind mittel- bis langfristig dringlich. Die Ergebnisse der Analyse und Einschätzung der Standardisierungslücken wurden ebenfalls in Experteninterviews und Workshops validiert.

4 Ergebnisse und Diskussion

4.1 Relevante Standardisierungsgremien und ihre Standards

Es existiert eine Vielzahl verschiedener Akteure im Normungs- und Standardisierungsumfeld von Cloud Computing. Der Auswahl liegt eine anfängliche Recherche von über 70 verschiedenen Institutionen zu Grunde. Der Fokus liegt auf Normungsorganisationen, Standardentwicklungsorganisationen, Interessensvereinigungen, sonstigen Konsortien oder öffentlichen Einrichtungen. Ihnen allen ist gemein, dass sie Gremien besitzen, die Standards oder Vorarbeiten mit implizitem oder explizitem Bezug zum Cloud Computing forcieren. Einzelne Forschungseinrichtungen oder privat-wirtschaftliche Unternehmen sind nicht im Fokus. Bei Letzteren bestehen keine regulären Mitwirkungsmöglichkeiten für Außenstehende.

In den USA nimmt NIST eine Vorreiterrolle bei der Cloud-Standardisierung ein. Einige internationale Standardisierungsgremien zeigen ebenfalls großes Engagement, während die überwiegende Mehrheit ihren Fokus nur langsam auf Standards für das Cloud Computing ausrichtet. Auf europäischer Ebene wird das ETSI eine koordinierende Rolle einnehmen. EuroCloud ist ein paneuropäischer Unternehmensverband der Anbieter von Cloud Computing mit großem Einfluss. In Deutschland unternehmen das DIN, der BITKOM und das BSI Schritte bei der Anforderungsdefinition.

Es wurden 21 Standards, Vorgaben, Zertifizierungen bzw. Vorarbeiten ausgewählt. Diese wurden im Detail untersucht, bewertet und von ähnlichen Standards abgegrenzt. Bei der Auswahl und Bewertung der 21 Cloud-Standards handelt es sich um eine Momentaufnahme von Mitte 2012 bei der wir insgesamt etwa 160 Standards in Betracht gezogen haben. Die 21 Cloud-Standards besitzen nach Möglichkeit Vorbildcharakter und decken die Bereiche Technik, Management und Recht ab. Kein branchenspezifischer Standard wurde als relevant genug erachtet, um in die engere Auswahl zu gelangen. Eine Übersicht der Standards ist in Tabelle 1 dargestellt.

Tabelle 1. 21 ausgewählte Cloud-Standards für Technik, Management und Recht

Standard	Initiator	Ähnliche
<i>CCRA</i> (Cloud Computing Reference Architecture): Referenzarchitektur für Cloud-Service-Angebote	TOG	Referenzarchitekturen der NIST oder des BSI
<i>CDMI</i> (Cloud Data Management Interface): API zum Zugriff auf Daten in IaaS, DaaS Szenarios	SNIA	XAM, iSCSI, NFS, WebDAV
<i>CIMSVM</i> (CIM System Virtualization Model): Objektmodell und Schnittstellen für Virtuelle Systeme & Komponenten	DMTF	TOSCA
<i>Cloud Audit</i> (Automated Audit, Assertion, Assessment, and Assurance API): API für automatisierte Auditierung	CSA	SCAP
<i>CTP</i> (Cloud Trust Protocol): Einheitliche Techniken und Nomenklatur zur Erhöhung der Transparenz	CSA	SCAP, OCRL
<i>Hive</i> (Apache Hive): Programmiermodell für Datenabfragen	Apache	JAQL, PIG
<i>OCCI</i> (Open Cloud Computing Interface): API zum Management von Clouds (insb. IaaS)	OGF	DeltaCloud, Libcloud, EC2, Eucalyptus, vCloud etc.
<i>OpenStack</i> (OpenStack Cloud Software): Rahmenwerk zum Aufbau von Cloud-Infrastrukturen	(Diverse)	OpenNebula, Nimbus (Schnittstellen: CMDI, OCCI, OVF)
<i>OAuth</i> (Web Authorization Protocol): Protokoll und Schnittstelle zum Identitätsmanagement	IETF	OpenID, WS-Federation, SAML
<i>OVF</i> (Open Virtualization Format): Dateiformat für Virtuelle Maschinen	DMTF, ANSI, ISO	AMI, EMI
<i>SCAP</i> (Security Content Automation Protocol): Protokoll und Schnittstelle zum Abruf von Sicherheitsinformationen	NIST	CloudAudit
<i>TOSCA</i> (Topology and Orchestration Specification for Cloud Applications): Beschreibungssprache für Cloud-Dienste	OASIS	CIMSVM, USDL, e3Value, SNN
<i>USDL</i> (Unified Service Description Language): Beschreibungssprache für virtuelle Dienstleistungen	W3C	OWL-S, WSMO, UDDI, WSDL, WADL, PAS1018
<i>WS-*</i> (Web Service Standards): Spezifikationen, Standards und Normen für Web Services	OASIS, OGF, W3C	WSDL, WS-I, WS-Policy, WS-Security, WS-Agreement etc.
<i>BSI-ESCC</i> (Eckpunktepapier Sicherheitsempfehlungen für Cloud Computing Anbieter): Leitfaden	BSI	Andere Anforderungsdokumente
<i>EuroCloud-SA</i> (EuroCloud Star Audit): Zertifikat für Anbieter von Cloud-Diensten	EuroCloud	EuroPriSe, TiC
<i>GRC Stack</i> (Governance, Risk Management and Compliance Stack): Rahmenwerk zu Risikobewertung von Anbietern	CSA	CloudAudit, CCM, CAIQ, CTP
<i>NIST-UC</i> (Cloud Computing Use Cases): Leitfaden für Cloud-Anwendungsfälle mit Fokus auf US-Behörden	NIST	Use Cases von OGF oder DMTF
<i>SSAE-16</i> (Statement on Standards for Attestation Engagements No. 16): Zertifikat für Anbieter von Cloud-Diensten	AICPA	CobIT, BSI-100, ISAE 3402, ITIL, SAS 70, IDW PS 330/951/FAIT1
<i>OCM</i> (Open Cloud Manifesto): Selbstverpflichtung zu Offenheit für Cloud-Anbieter	(Diverse)	keine
<i>95/46/EG</i> (EU-Richtlinie 95/46/EG „Datenschutzrichtlinie“): Datenschutzvorgaben der EU	EU	BDSG, DSGVO der Länder, Safe Harbor

Die überwiegende Mehrheit der Standards hat internationale Relevanz. Einzelne weisen einen (leichten) europäischen bzw. nationalen Bezug auf (z. B. BSI-ESCC, USDL, NIST-UC, EuroCloud-SA, 95/46/EG). Die Bewertungsergebnisse für die Standards spiegeln den frühen Entwicklungsstand im Cloud Computing wider. Standards, die bereits vor dem Cloud Computing existierten, weisen eine tendenziell grö-

Bere Reife auf (z. B. SCAP, WS-*, OAuth, CIMSVM, SSAE-16) als solche, die aktuell explizit für das Cloud Computing erarbeitet werden. Die Durchsetzungsfähigkeit von Standards mit explizitem Bezug zum Cloud Computing erweist sich hingegen tendenziell höher, als bei solchen mit implizitem Bezug.

4.2 Diskussion der Ergebnisse

Tabelle 2 zeigt Standards sowie Standardisierungspotentiale und -lücken anhand einer Überlagerung in der Landkarte und fasst so die Ergebnisse der Studie zusammen.

Die Mehrzahl der Standards fokussiert die Herausforderungen Informationssicherheit, Effizienz, Interoperabilität oder Portabilität aus technischer Perspektive. Bedarf an technischen neuen oder umfänglicheren Standards besteht bspw. bei Standardkomponenten, Referenzarchitekturen, Benchmarks, Tests oder Protokollen und Schnittstellen. Im Bereich der Managementstandards finden sich nur wenige Standards. Es existieren keine oder nicht ausreichend umfassende Standards für Geschäftsmodelle, Dienstgütevereinbarungen, Managementmodelle sowie -prozesse sowie für das Controlling im Cloud Computing. Das Zusammenspiel des Rechtsrahmens und der Standardisierung im Cloud Computing ist vielschichtig und wird bislang überwiegend auf Herausforderungen im Bereich Datenschutz reduziert. Im Bereich der vertraglichen Regelungen fehlen unter anderem standardisierte, verbindliche unternehmensinterne Vorschriften (Binding Corporate Rules, BCR) für Cloud-Anbieter zum Datenschutz im Zuge einer Selbstregulierung. Auf europäischer und deutscher Ebene ist daher die Klärung des grundsätzlichen strategischen regulatorischen Vorgehens notwendig.

Die Potentialanalyse legt nahe, dass für jede der identifizierten Herausforderungen ein Lösungsansatz durch Standardisierung erfolgsversprechend erscheint. Dabei versprechen technische Ansatzpunkte insbesondere für Herausforderungen, die im Kern eher technischer Natur sind (wie bspw. Effizienz, Effektivität, Informationssicherheit aber auch Interoperabilität und Portabilität) eine Lösung. Im Bereich der Sicherstellung des funktionierenden Wettbewerbs, der Compliance sowie des Datenschutzes liefern technische Ansatzpunkte die geringsten Beiträge. Die Potentialanalyse zeigt weiter, dass die genannten, vermeintlich von technischer Natur stammenden Herausforderungen nur dann angemessen adressiert werden können, wenn auch Standardisierungspotentiale durch Ansatzpunkte auf Management- (Effizienz, Effektivität) und Rechtsebene (Interoperabilität, Portabilität) ausgereizt werden. Auch zeigt sich, dass der Bereich der Informationssicherheit sowie in Teilen auch Portabilität und Compliance nur durch auf Technik-, Management- und Rechtsebene aufeinander abgestimmte Standardisierungsbestrebungen gelöst werden können.

Tabelle 2. Landkarte der Standardisierung im Cloud Computing

	Technik					Management					Recht			
	Datei- und Austauschformate	Programmiermodelle	Protokolle & Schnittstellen	Standardkomponenten & Referenzarchitekturen	Benchmarks und Tests	Geschäftsmodelle	Service Level Agreements	Vertragsbedingungen	Managementmodelle & -prozesse	Controllingmodelle & -prozesse	Leitfäden, Audits etc.	Rechtliche Vorgaben	Selbstverpflichtungen	Unternehmensrichtlinien
Compliance			CloudAudit								EuroCloud-SA, GRC			
Wettbewerb												OCM		
Portabilität	CIMSVM, OVF, TOSCA, USDL	Hive	CIMSVM, OCCL, OpenStack	CIMSVM, OpenStack		USDL	USDL	USDL			EuroCloud-SA	OCM	OCM	
Interoperabilität	CIMSVM, OVF, TOSCA, SCAP, USDL, WS-*	Hive	CDMI, CIMSVM, OCCL, OpenStack, SCAP, WS-*	CDMI, CIMSVM, OpenStack		USDL	USDL	USDL			NIST-UC	OCM	OCM	
Datenschutz				BSI-ESCC							BSI-ESCC, EuroCloud-SA, NIST-UC	95/46/EG		
Informationssicherheit	SCAP, WS-*		CTP, Oauth, SCAP, WS-*	CTP, CCRA, BSI-ESCC				GRC, SSAE-16	SSAE-16		CTP, BSI-ESCC, EuroCloud-SA, GRC, NIST-UC	OCM	OCM	
Transparenz	USDL, WS-*		CloudAudit, CTP, WS-*	CTP, CCRA		USDL	USDL	USDL			CTP, EuroCloud-SA, NIST-UC		OCM	
Effektivität			CDMI, CTP, OCCL, OpenStack	CDMI, CTP, CCRA, OpenStack				GRC			CTP, EuroCloud-SA, GRC, NIST-UC			
Effizienz	CIMSVM, OVF, TOSCA, USDL, WS-*	Hive	CDMI, CIMSVM, CTP, OCCL, OpenStack, WS-*	CDMI, CIMSVM, CTP, CCRA, OpenStack		USDL	USDL	USDL	GRC		CTP, EuroCloud-SA, GRC			

Potential der Standardisierung: Ja Eher ja Eher nein Nein

Es lassen sich folgende Standardisierungsbedarfe hervorheben: Es fehlen Standardkomponenten und Referenzarchitekturen zur Leistungsüberwachung und Sicherstellung von Compliance. Existierende Standards greifen diesen Aspekt bislang nur unzureichend auf. Benchmarks & Tests sind nicht verfügbar. Insbesondere im Bereich der Effektivität der Dienstnutzung könnte durch Vergleichbarkeit mehr Vertrauen geschaffen werden. Protokolle & Schnittstellen zur Transparenzsteigerung durch automatisierten Informationsaustausch sind unzureichend ausgeprägt. Deutlich sind die Lücken bei standardisierten Vertragsbedingungen zur Gewährleistung der Informationssicherheit und Sicherstellung des Datenschutzes speziell für kleine und mittelständische Unternehmen. Darüber hinaus sehen wir weiteres Potential für umfassende und erprobte Datei- und Austauschformate zur Unterstützung von Interoperabilität von Cloud-Diensten sowie für standardisierte Protokolle & Schnittstellen zum Austausch im Sinne der Portabilität. Es finden sich noch keine ausreichenden, rechtlichen Vorgaben zur Sicherstellung des funktionierenden Cloud-Wettbewerbs.

5 Ausblick und Empfehlungen

Wir haben basierend auf einem Screening von etwa 160 Standards sowie über 70 Standardisierungsinitiativen im Cloud Computing 21 Cloud-Standards ausgewählt und in einer Standardisierungslandkarte dargestellt. Diese Übersicht haben wir überlagert mit den Ergebnissen einer Potentialanalyse. Die Ergebnisse haben z. T. deutlich Lücken aufgezeigt, aber auch Bereiche, für die bereits Standards existieren, welche die gestellten Anforderungen nur bedingt erfüllen oder nicht ausreichend erprobt sind.

Aus den Gesamtergebnissen der Untersuchung lassen sich damit einige Handlungsempfehlungen für die Standardisierung im Cloud Computing ableiten: Zunächst sollten offene Standardisierungslücken weiter priorisiert werden und insbesondere öffentliche Anforderungen klar formuliert werden. Bestehende Standards sollten kontinuierlich katalogisiert werden. Darüber hinaus sollte die Offenheit von Standards im Cloud Computing durch das Setzen von Anreizen gefördert werden. Auch sollten existierende Standards auf ihre tatsächliche Offenheit geprüft werden. Wichtig ist dabei Eckpunkte zu definieren, um Doppelarbeit zu vermeiden. Auch erscheint es sinnvoll, Standards für das Vertragswesen vom rechtlichen Rahmen abzugrenzen. Die Standardisierung sollte über nationale, europäische und internationale Verwaltungsebenen hinweg sowie unter Einbeziehung aller Akteure zentral koordiniert werden (z. B. als Standardisierungs-Roadmap). Der bestehende nationale wie europäische Rechtsrahmen sollte auf Angemessenheit und Implikationen für das Cloud Computing umfassend geprüft werden, um geeignete rechtliche Vorgaben abzuleiten. Staatliche Organisationen sollten dabei in fokussierter Weise und begrenztem Maße inhaltlich bei der Standardisierung mitwirken. Der Schwerpunkt liegt auf der Anforderungsdefinition. Die eigentliche Standardisierung ist Aufgabe der Wirtschaft unter Mitwirkung der Wissenschaft. Dabei sollte insbesondere die Rolle der Anwender von Cloud-Diensten mehr Wertschätzung erfahren.

Literatur

1. Baun, C., Kunze, M., Nimis, J., Tai, S.: Cloud Computing: Web-basierte dynamische IT-Services. 2. Auflage, Springer, Berlin (2011)
2. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A View of Cloud Computing. *Communications of the ACM* 53, 50-58 (2010)
3. Marston, S., Lia, Z., Bandyopadhyaya, S., Zhanga, J., Ghalsasi, A.: Cloud Computing: The Business Perspective. *Decision Support Systems* 51, 176-189 (2011)
4. National Institute of Standards and Technology (NIST): NIST SP 800-145, The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology, <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>
5. ISO/IEC: ISO/IEC Guide 2: Standardization and related activities: General vocabulary. ISO/IEC, Geneva (1996)
6. Jeffery, K., Neidecker-Lutz, B.: The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010. Report (2010)
7. Smart Cloud Study Group: Smart Cloud Strategy (2010), http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/councilreport/pdf/100517_1.pdf
8. European Network and Information Security Agency (ENISA): Cloud Computing Benefits, Risks and Recommendations for Information Security (2009)
9. Cloud Security Alliance (CSA): Security Guidance for Critical Areas of Focus in Cloud Computing V2.1, <http://www.cloudsecurityalliance.org/guidance/csaguide.pdf> (2009)
10. EuroCloud: Leitfaden: Recht, Datenschutz & Compliance, <http://www.eurocloud.de/2010/12/02/eurocloud-leitfaden-recht-datenschutz-compliance/> (2010)
11. European Network and Information Security Agency (ENISA): An SME perspective on Cloud Computing: Survey (2009)
12. Cloud Security Alliance (CSA): Cloud Controls Matrix (CCM) Version 1.2, https://cloudsecurityalliance.org/research/ccm/#_overview (2011)
13. Cloud Security Alliance (CSA): Consensus Assessments Initiative Questionnaire v1.1, https://cloudsecurityalliance.org/wp-content/uploads/2011/03/CSA-CAI-Question-Set-v1-1_FINAL_v6.xlsx (2011)
14. Cloud Computing Use Case Discussion Group: Cloud Computing Use Cases. Version 4.0, <http://cloudusecases.org/> (2010)
15. Sakai, H.: Standardization Activities for Cloud Computing. NTT Technical Review, <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201106gls.pdf> (2011)
16. National Institute of Standards and Technology (NIST): NIST-SP 500-291, NIST Cloud Computing Standards Roadmap, http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909024 (2011)
17. National Institute of Standards and Technology (NIST): Cloud Computing Use Cases, <http://www.nist.gov/itl/cloud/use-cases.cfm> (2010)
18. Internet Engineering Task Force (IETF): Cloud SDO Activities Survey and Analysis. draft-khasnabish-cloud-sdo-survey-03.txt, <https://rsync.tools.ietf.org/html/draft-khasnabish-cloud-sdo-survey-03> (2012)
19. ITU Telecommunication Standardization Bureau: Activities in Cloud Computing Standardization, Repository (Version 1.0, May 2010), http://www.itu.int/dms_pub/itu-t/oth/49/01/T49010000020002PDFE.pdf (2010)

Six Principles for Governing Mobile Platforms

Julia Manner, David Nienaber, Michael Schermann, and Helmut Krömer

Technische Universität München, Chair for Information Systems, Garching by Munich,
Germany

{Julia.Manner,david.nienaber,michael.schermann,krcmar}@in.tum.de

Abstract. Mobile platforms enable third-parties to extend and enhance functionalities of products and services by mediating these to consumers. A tremendously successful example is Apple's App Store where third-party developers contribute applications that add to the usefulness of the mobile device. However, research and practice still struggle with understanding why and when such mobile platforms prosper or wither. In this paper, we develop a framework to study how companies use governance mechanisms to attract customers and third-party developers. We demonstrate the usefulness of our framework by analyzing how governance mechanisms were used to change the interaction of platform, customers, and third-party developers. We derive six principles for governing mobile platforms that highlight the need to further investigate how governance mechanisms align the value propositions to platform operator, customers and third-party developers.

Keywords: mobile platforms, governance, controls, services

1 Introduction

Enabling third-party developers to add functionalities to a core product by providing a mediating mobile platform is an increasingly popular model to raise earnings and enhance customer satisfaction [1-3]. Mobile platforms are also known as App Stores providing a development and marketplace environment.

However, many companies struggle to provide a sustainable and thriving mediating platform. This challenge is illustrated by the fact that 97 percent of the mobile services market share is held by only seven platforms, despite the fact that 40 to 50 platforms exist [4]. As platforms are affected by network effects, the success of a platform is determined by having enough participants on the development side to attract the customer side and vice versa. These stakeholder interests need to be aligned by governance [5-6].

The struggle with platform success in practice highlights the complexity of mobile platform governance and reflects the immaturity of practical understanding and research on platform governance [7-9]. In particular, we observe that in practice platform providers use imitation strategies rather than developing a proactive governance

concept [10]. Yet, imitation ignores the fact that platform governance needs to fit to the ecosystem of the platform which is never identical across platforms [4], [11].

In this paper, we triangulate results from heterogeneous background - a previous literature review, a basic governance concept for platforms and real world example - to posit initial principles for using governance to align the value propositions for platform operator, customers, and third-party developers. We thereby contribute to research on the important topic of mobile platforms by developing a theoretical framework that helps to study how companies use governance mechanisms to become attractive to customers and third-party developers. The usefulness of our framework will be demonstrated by analyzing how governance mechanisms were used to change the interaction of platform customers and third-party providers. In conclusion, we derive six principles for governing mobile platforms that highlight the need to further investigate how governance mechanisms align stakeholder interests.

The remainder of this paper is structured as follows: First, we discuss the literature on platform governance identified in the literature review of Manner et al. [8]. Based upon this, we adapt the basic concept of platform governance proposed by Manner et al. [8] to provide a practically applicable platform governance framework to study and classify the use of governance mechanisms in the mobile platforms of Apple, Google and Microsoft. We compile evidence from the field with literature to propose initial principles for using governance mechanisms to align the value propositions to platform provider, customers and third-party developers. Finally, we discuss our findings and conclude with an outlook on future research opportunities.

2 Theoretical Background

2.1 A Definition of Mobile Platform Governance

In this paper, we focus on mobile platforms such as Apple's App Store with two-market sides: the developer environment and the buying environment. Mobile platforms like Apple's App Store are socio-economic layered, consisting of an IT-based artifact which enables external knowledge holders to contribute functionality to a core product [7], [12]. In this way, platform providers are mediating the process of adding functionality to the IT-based marketplace and therefore to the consumers to gain further economic value; for example, by extracting revenue from transaction and to achieve their strategic aim.

Aside from platform providers, there are two major stakeholders to be considered in a platform ecosystem: developers and consumers. They are part of the definition of platform ecosystems as a functional unit consisting of the platform provider, developers, consumers as well as strategic partners of the platform provider [13-14]. The integration of third-party developers into the value creating process is a challenging concept. A platform provider needs to establish regulating guidelines, documentations and rules to enable and guide developers which assist them in creating compatible applications to extend the core product [9]. On the one hand, developers need support to be able to create functionality as well as creative freedom to be innovative [15]. On

the other hand, the platform provider must maintain control and prevent developers to bypass the platform, building a direct developer consumer relationship. Thereby, providers lose their possibility to integrate their strategic aims like a specific customer experience and gaining money from the platform [13], [16].

The regulating framework to shape a viable and sustainable platform is known as platform governance. It provides structures to manage the stakeholder relationships [17], determines the allocation and distribution of power for actions [18] and controls the actions of the platform stakeholders [19]. Hence, platform governance is defined as “the structure, power, processes, and control mechanisms that are applied by the platform owner to achieve his aims” [8].

2.2 Theories Related to Mobile Platform Governance

The key element to a viable platform is the balance of its ecosystem by sound platform governance [7], [11], [20]. It is a multi-dimensional concept that links the technical architecture of a platform with the ecosystem, and the platform provider’s power. Hence, studying mobile platforms demands the investigation of technical, organizational and financial aspects [21].

Since the concept of platform governance touches many different disciplines, the existing knowledge of governance is very fragmented [7], [10]. Many researchers in the fields of economics, strategic management and organizational science have examined aspects of platform governance corresponding to their discipline [8], [11]. Information systems research now combines some of this fragmented work originating from its reference disciplines [8], [22].

Economics was one of the first research disciplines to study the underlying concept of platforms: pricing in two- and multi-sided markets [5-6]. They discuss pricing mechanisms for platforms and price elasticity in the competitive ecosystem of dynamic systems [6], [17], [23]. Still, research in this field provides little managerial implications for platform providers and governance [7]. Strategic management literature regards the role of platforms as complementary markets [3], [24], starting a broad discussion about “how open is open enough” [25], observing the field in a practical way and explaining the past [14], but not providing process implications for practitioners. Organizational science literature addresses platform governance from a control perspective [26]. Controls are considered a powerful instrument to synchronize activities [27] and to “assure ongoing alignment between their investment in the community and related product goals” [28].

Across all streams of literature, there is consensus that successful platform governance fits to the ecosystem of a platform and thus cannot be easily imitated [7], [9]. Hence, actively developing distinct platform governance mechanisms that fit to the platform ecosystem might be a solution [7]. Yet, a comprehensive understanding of platform governance, its elements and how they are linked together is necessary.

3 A Comprehensive View on Mobile Platform Governance

In this chapter, we discuss the elements of platform governance derived from literature and their links to each other, building a comprehensive view on mobile platform governance based on the findings of Manner et al. [8]. Figure 1 illustrates the three levels of the platform governance framework. The top level is the market structure, the second level is the governance policies and the bottom level is the governance configuration. Taking this perspective, we begin to elaborate on the elements of governance and its implications.

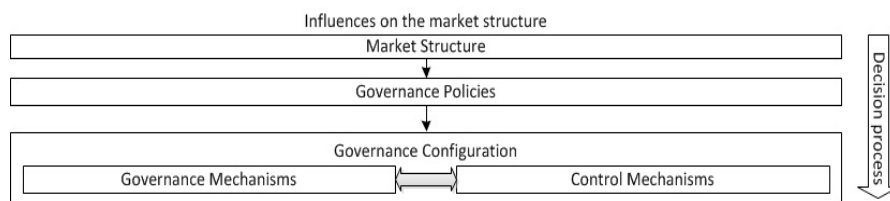


Fig. 1. Top level governance framework [8]

3.1 Market Structure

The value propositions of a platform depend on the congruence of the platform with a dynamic market structure [4], [11]. The discipline of marketing defines the market structure as the macro-ecosystem of the platform [29]. Thus, a fit between platform governance and the ecosystem is vital for platform success [7]. As an example, software providers “are constantly evolving, increasing their depth and/or reach and in the process redefining their boundaries”, which stresses the importance of ecosystem alignment [1], [30]. Following a similar line of arguments, de Reuver et al. [31] suggest important triggers that describe a dynamic market structure, such as the emergence of new technologies, regulatory changes, or competitive changes from changing consumer behavior or behavior of competitors. These dynamics in the market structure require to constantly realign the value propositions of platform operators, customers, and third-party developers [2], [6]. Consolidating literature on classical market structures [29] and electronic markets [32], we condense the following triggers of a market structure for mobile platforms. These triggers constitute important cues to the platform operator when formulating the objectives and the strategy along with the business model [32]: **Legal triggers** describe events originating from legislation, government agencies, or jurisprudence that affect the platform, e.g., restrictions to access user data [8]. **Technical triggers** originate from technical progress, e.g. the increasing availability of fourth generation broadband Internet access. Also pre-commercial developments have to be taken into account as they may have significant impact on the design of a platform [4], [16]. **Socio-economic triggers** refer to economic events and actions that also influence social reality, for example the increasing awareness for privacy issues in society [4], [7], [32]. **Competitive triggers** originate from changes in the competitor’s behavior. For instance, such triggers comprise of

strategic moves of competitors that shift competition advantages between market actors [7], [19]. **Stakeholder triggers** refer to actions of customers, third-party developers or other strategic partners of the platform operator. To give an example, developers who are disgruntled by poor technical support, might switch to competing platforms. Changes in customer behavior are also important triggers [11], [21], [31].

3.2 Governance Policies

The next task is the transition from ecosystem to governance, which is a sub form of the market structure [33]. This is achieved by the formulation of policies. We define policies as an organizing logic which is derived from the market structure. Heading up, we assessed the basement of literature identified by Manner et al. [8]. Thereby, we observed a variety of issues which lead to seven policy sections. These aim at resolving particular questions.

Standardization policy: To what extent shall the platform be standardized? This implies a trade-off, considering that a too high degree of standardization means poor usability for developers but assures flexibility to adapt to ecosystem changes [18]. The interaction between different functionalities can be supported by standardization which can deliver an added value by complementary use. However, innovation can be hampered. Ensuring usability is a further trade-off for platform providers. Ease of use is supported by a uniform interaction design but it demands high and expensive control efforts. These rules aim at aligning features, services and platform experience. Providing a new API or SDK sets a standard as well as implementing a new section within a marketplace.

Incentive policy: How are the platform stakeholders incentivized, especially developers? Users as well as developers need to adapt to the platform. It is not well known what really motivates developers to develop for a certain platform. According to Holzner et al. [15], financial aspects, prestige and creative freedom are important incentives for developers. Providing creative freedom contradicts to the user's demand of ease of use and compatibility of functionality. The design of the incentive policy must therefore be balanced with the standardization policy. Many other examples exist. Yet, they pose interdependencies to other policies [7], [15]. Incentives can be either positive like new incentive programs or negative when a previously implemented incentive is withdrawn by the provider.

Legal policy: Who is allowed to use, change, transfer and benefit from the platform and its facilities? Property rights for apps could remain with the developer, but could also be transferred to the provider. This issue may be addressed – for example – by shared income and licenses [19]. With increasing access to the core product by applications of third-parties, new legal issues arise.

Quality of Service policy: Which quality of service is offered to the customer? Quality is an important aspect for platforms regulated by governance [33]. This topic addresses issues of functionality, availability, performance and perceived service quality from the viewpoint of service consumers [21], [34].

Distribution policy: How are services, support and tools distributed by the platform? Making content accessible in the right way to the customers is vital for custom-

er satisfaction and service usability. Developers have to be provided with the latest SDK and processes for application integration have to be implemented [19]. In general, the whole value network from developer to service consumer is covered. There are also different ways to enable customers to access the offering. For instance, Android customers can choose between several distribution platforms, but Apple customers must use Apple's App Store or install illegal software applications like Cydia [12], [34].

Finance policy: How is revenue generated and how are the financial aspects organized on the platform? More important is the share of profit between the platform stakeholders [21]. Customers need to pay reasonable prices for their added functionalities. Developers need to make stimulating profits or they will not continue developing for the platform. A platform provider needs to skim enough off the platform revenue streams to be able to continue offering and nurturing the platform. Detailed aspects are the right to set prices on applications, discretization of prices and entry fees for developers [18-19], [21].

Security policy: What security standards have to be implemented? This refers to data, communication and distribution security, access to personal user data and similar aspects [16], [18-19]. The platform provider decides what APIs are offered and thereby which access the third-party applications are allowed to the device. The more APIs are offered, the more functionality can be built. On the downside, possibilities for security gaps that need to be controlled increase in number. The security policy also ensures the authentication process and determines the authorization of third-party applications and users.

The goal is to align policies perfectly to the market structure [7], [35]. The five core papers on platform governance, identified by Manner et al. [8] revert to more than twenty different aspects, which have to be resolved in a platform context. These all can be categorized into the above mentioned policies. Based on literature, we believe that a main challenge on this governance level is to find the correct fit between these policies [7]. For example, the more technical support (quality of service policy) is provided to the developers, the more skilled they get. Thus, the platform provider can grant skilled developers a higher degree of design clearance, which stimulates them positively (incentive policy). But a too high degree could infringe the platform, when allowing too much access to programming interfaces and other means of the standardization policy. The implementation of low admittance criteria for developers to foster immigration (incentive policy) is another aspect. But to keep the desired quality-of-service, one has to adapt the technical support to the needs of unskilled developers. All interdependencies are of a varying nature and must be analyzed in the specific context. The policies defined here provide view of issues every platform has to consider.

3.3 Governance Configuration

The constituted governance policies must be implemented to manage consumer experience and to align with stakeholder objectives [20]. Platform operators need to choose from a plethora of formal and informal mechanisms to implement their gov-

ernance policies [16]. Every mechanism can be described from a governance perspective as well as from a control perspective [7], [18].

De Reuver and Bouwman [18] classify governance mechanisms into three not-mutually exclusive categories: **Authority-based governance mechanisms** employ the platform provider's power to enforce behavior as desired using policies. **Contract-based governance mechanisms** refer to legally binding agreements, whereas **trust-based governance mechanisms** are based on the assumption that the target group has strong incentives to reach the desired goal and levers these incentives [18].

Control mechanisms inform the platform operator about the extent to which the governance mechanisms are effective and efficient. Originally, Ouchi [36] defined three different types of control mechanism: outcome, behavior and social control. They all aim at achieving coordination between parties with different intentions and goals in organizations. Kirsch [37] expanded social control through self-control. Thereby, she distinguishes between group and individual level. Cardinal et al. [38] adds input control where meeting desired inputs such as employee skills to an objective are appreciated. However, we differentiate the following control mechanisms: **Input control mechanisms** define required characteristics of provided input, e.g. developers and apps. Meeting these characteristics is rewarded. An **output control mechanism** sets a desired output. Again, reaching such goals is rewarded. **Behavior control** provides procedures and routines to reach a goal, where compliance with the procedures and routines is rewarded. **Social control mechanisms** strengthen common values and shared beliefs. People acting according to these are rewarded by social acceptance to the system. Such a control is implemented by rewarding behavior that has a "substantially overlap" with the interests of the subject under control [36], e.g. developers that code in a certain style are rewarded by thorough feedback on their code.

Output and behavior control are formal mechanisms, whereas input and social control are informal mechanisms. According to previous findings, control is implemented by a mixture of control mechanisms [36-37]. Applying the right balance of each control mode is essential [39].

3.4 A Framework for Mobile Platform Governance

We have three levels to consider when designing a platform governance concept. The first level is the market structure giving an overview of the platform's current situation which is influenced by legal, technical, socio-economic, competition and stakeholder behavior aspects. Changes to these imply an adaption of the present governance concept. Therefore, it resembles the starting point for the development of sound governance.

Table 1. Governance Framework: Important elements and corresponding levels.

Level Aspects								
Market Structure	Factors	Legal	Technical	Socio-Economic	Competition	Stakeholder Behavior		
Governance Policy	Policies	Standardization	Incentive	Legal	Quality of Service	Distribution	Finance	Security
Governance Configuration	Governance Mechanisms	Authority-based	Contract-based	Trust-based				
	Control Mechanisms	Input	Outcome	Behavior	Social			

On the second level, governance policies are derived to implement the business strategy in conformance with the market structure and to achieve corresponding goals. At the bottom layer, governance mechanisms implement the policies from a means point of view, focusing on its characteristics, whereas controls focus on targets, mainly specific stakeholders and their behavior. Table 1 aggregates the critical issues of platform governance into a morphological framework.

4 Insights from Changes in Governance Concepts

In order to understand the effectiveness of governance concepts, we analyzed three governance changes from different platform providers: Apple, Android, and Microsoft. The goal of the multiple change analysis was to find out, if the framework is applicable for governance change analysis and findings on how successful platforms are governed. Apple’s App Store and Google’s Google Play mobile platforms have often been described in research as completely different (closed vs. open). Therefore, they are suitable platforms to gain insights. We also chose the Windows Phone Store in contrast to the two mobile platform leaders Apple and Google [34]. At present, Microsoft battles with establishing a developer and consumer base rather than base extension. The chosen cases are suitable to demonstrate the framework applicability and to provide further insights as they provide a wide variety.

4.1 Governance Changes in the Apple Platform

Earlier this year, **Apple** made a significant privacy change. This was triggered by a broad discussion in politics and society overall (**socio-economic and stakeholder behavior trigger**) about the unauthorized extraction of personal data like the address book by application providers. Apple may have feared that the discussion could start to unsettle users and make them more reluctant to buy applications. Then, the attractiveness of the platform for users would have additionally suffered.

In consequence, Apple used its power as platform provider and applied an adapted security policy prohibiting unauthorized access via applications without prior asking for access to secure the trust of its mobile device customers.

To enforce the security policy Apple changed from a **contract-based** governance mechanism, which only prohibited the unauthorized access in the guidelines, to **au-**

authority-based governance. Previously, although it was forbidden to access the data, it was not controlled, whether the application recorded data without permission. This means it was **social control (informal)**. The new control consists of a mixture of input and behavior control. The implemented **input control mechanisms** are a change in the SDK and the release of a new operating system (**authority-based**), whereas contracts remained as **behavioral control** steering the developers behavior (**contract-based**) [40]. In this real world change, we find two principles applied. It presents a formal input control supported by an informal behavioral control presenting a **control mix**. Since Apple aimed to limit uncertainty by generating new trust of users in Apple who otherwise would have left the platform, we also find high **mechanism strength** that is **proactively** implemented, as there was no lawsuit or law at the moment of change.

4.2 Governance Changes in the Android Platform

The second example analyzes **Google's** change of their platform governance concept regarding the Android billing system. Google had installed a **contract-based** mechanism controlled by **social control** to make developers use Google's in-app billing system for in-app purchases. However, Google recognized that more and more developers circumvent Google's in-app billing system.

Due to this fact, consumers were confronted with different billing systems, which negatively influenced their experience and Google's revenue. This omitting behavior of the developers as **stakeholders** triggered a governance change. Android changed the approval process regarding in-app billing. Google established its own billing system, implementing an **authority-based** governance mechanism via an adapted **standardization policy**. This policy change is controlled by an adapted approval process. The approval process, as a form of **output control**, now rejects any applications infringing this rule [41]. We identify that governance changes trigger control changes in the same direction. More strict governance leads to a formal strict control. Therefore, we derive a **congruency** between governance and control.

4.3 Governance Changes in the Windows Platform

The last case deals about the **Windows Phone** Store. Being a new entrant with already two major **competitors** like Android and Apple, Windows needed to catch up with the amount of application offerings. Only few developers worked for Windows Phone due to the lack of advantages they saw in providing their applications for Microsofts' small consumer base [15]. Microsoft needed to establish a relationship to potential application developers to gain a certain amount of developers providing innovative and various applications to also attract consumers.

Hence, **trust-based governance** in the form of a new **incentive policy** for developers was introduced. By providing free phones for developers, promising visibility of their apps as well as advance payment, money and prestige was given as incentive. From a control perspective, Microsoft also implemented an informal **social control**. By providing these incentives they reached a substantially overlap between the devel-

opers' interests and its Windows Phone Store [36], [42]. In this real world example, we observe two principles. Firstly, the Windows Phone platform implemented **proactively** the incentive policy to achieve more participants. Secondly, the governance and control mechanisms were implemented **congruently** as a light governance mechanism is supported by a light informal control.

5 Six Principles for Governing Mobile Platforms

Principles for governing mobile platforms have not yet been suggested in literature. We therefore propose to derive principles for governing mobile platforms based on our framework, evaluated literature and real world examples of governance changes. The suggested principles are reflected between these sources and identified in at least two of them to achieve more robust results [43]. They are summarized in Table 2.

We argued that all three governance levels (market structure, governance policies and the governance configuration) are important. Each governance level sheds light on different aspects that need to be considered when deriving platform governance. Thus, we propose “consider every governance level” as the first governance principle.

The next topic to be considered is the dynamic nature of the ecosystem of the platform. To be viable, platform owners need to quickly react to changes in the ecosystem [7]. This requires an ongoing evaluation and anticipation of changes and their impacts to the market structure. Platforms should constantly adapt their governance and learn from these changes. Our second governance principal is therefore: “Use proactivity to ensure viability”.

The direct link between the market structure and the governance policies implies the need to align both levels. We argue that mechanisms are the means to enforce policies. Thus, changing policies always trigger a subsequent change in the governance configuration. Therefore, we propose the third and the fourth governance principle “Adapt policies to changes in the market structure” and “Constantly align governance and control mechanisms”.

Platform providers differentiate primary and secondary objectives, e.g. being profitable or establishing a large user base [2], [15]. Primary objectives need to be enforced by authority-based governance and formal control to ensure their pursuit. Secondary objectives are mainly supported by trust-based and informal control. Thus, our fifth governance principle is: “Enforce primary objectives through authority-based governance and formal control”. Yet, control modes are implemented by a mixture of controls, where formal controls are the main part, aided by informal instantiations [37]. Applying the right fraction of each is essential. Our sixth governance principle is therefore: “Aid formal control with the right amount of informal control”.

Table 2. Six principles for governing mobile platforms from a provider’s perspective

#	Principle	Explanation	Rationale	Source*
1	Complete Use	Platform providers should think through all three governance levels to structure their governance decision process and evaluate the impact of the change.	Governance can be divided into three abstract and interdependent levels. All of them have to be considered when implementing governance as each has significant implications on the platform.	L, FW, RW
2	Proactivity	Anticipate changes in the ecosystem to be ahead of your competitors and thus ensure competitive advantage and success.	Governance is about the alignment with the ecosystem. Anticipating changes enable to adapt faster than your competitors.	L, RW
3	Congruency I	Governance policies need to be aligned with the market structure to guarantee the correct implementation of governance and control mechanisms	Governance policies contain implementation guidelines. Implementing the right strategy is vital for platform success. Thus, the correct alignment of market structure and policies is a precondition to provide a thriving platform.	FW, RW
4	Congruency II	Governance and control mechanisms have to be aligned with the policies to ensure correct implementation of the strategic aims	Governance and control mechanisms are the means through which the platform provider takes action. After formulating the correct strategy in the policies, its manifestation in tangible mechanisms is essential to ensure platform enforcement.	FW, RW
5	Mechanism strength	When pursuing a primary business objective, providers should only use authority-based governance and formal control (and no other means)	Platform providers need to meet primary goals to be successful. Mechanisms which ensure the fulfillment of the goals are necessary. Relying on trust and goodwill is not a viable option.	RW, L
6	Control mix	Formal controls to ensure primary objectives should be aided by informal controls.	Primary goals should not be enforced by informal controls as they are essential and fulfillment should not be exposed to uncertainty. Consequently, formal controls are required. To ensure that primary goals are reached, further aid is necessary to limit uncertainty by adding extra informal controls.	RW, L

*Source: L - Literature on platform governance; FW - Governance framework; RW - Real-world Examples

6 Discussion

Research on platform governance is fragmented [7], [9]. As many companies fail in providing a thriving platform by aligning stakeholder interests in real world, we argue that platform governance research could advance understanding of why some mobile platforms prosper and others wither. Although research on platforms and governance is increasing, there are little empirical insights on this complex concept to this point. When a complex phenomenon is studied, a triangulation of sources is especially appropriate since it contributes to the robustness of results and a more complete understanding [43]. Therefore, we combined insights from literature qualitatively, reflecting these on a previously developed framework and deriving insight from real world examples.

We adapted a basic mobile platform governance concept provided by Manner et al. [8]. Unlike its predecessor, the framework combines the elements together and provides classification characteristics within each level making it applicable for real word governance change analysis. The final framework structures platform governance into three distinct levels, the market structure, governance policies and the governance configuration as well as provides classification characteristics within the

level, enabling platform providers to structure their decision. Moreover, we analyzed three real world governance changes demonstrating the applicability and usefulness of the framework to analyze and understand governance.

To support the framework with guidance, we proposed six principles for governing mobile platforms which are based on literature findings, elements of our framework and three real world governance changes. Each principle was derived from at least two of these three sources. Hence, we claim to have proposed robust findings.

The first principle **complete use** was found in literature, derived from the framework and within each real world example as we were able to fully apply the framework. In fact, we could also find the **proactivity** principle proposed by literature as it is reflected in all presented governance changes. Every store actively initiated a change without being forced by for example law or bankruptcy.

The third principle of **congruency I** between market influence factor and the governance policy is validated by the fact that a market influence factor triggering the governance policy change was identified within each change. The fourth principle **congruency II**, which proposes the alignment between governance and control, is reflected by the Windows Phone Store change where the implementation of a light governance mechanism is supported by a light informal control. The counterexample is provided by Google Play presenting the implementation of strict governance that leads to a formal strict control which in the end results in a high validity of this principle. Ensuring the enforcement with mechanisms is especially important when business interests are depending. Apple's interest not to lose trust of market participants was undeniably strong. Analyzing the change and detecting strong mechanisms supports and validates the literature derived principle of mechanism strength on the condition of business objectives. At last, literature is proposing that control mechanisms do not occur mutually exclusive. Evidence on our proposed principle of **control mix** was also found in reality where an input as well as behavioral control was applied by Apple's App Store.

Altogether, we were able to supply reliable verification for each principle. Furthermore, the principles were reflected and rationally analyzed to provide quality control. Our real world examples clearly support the claim that governance needs to be designed carefully. A fitting governance concept can only be designed for one certain platform, not for several and even such a concept does only fit until it needs to be adapted to the ecosystem again [7], [9], [11].

Our findings propose the first support for platform providers to develop governance rather than imitate as the framework proposes a structured way to analyze governance changes and enables to identify more easily what consequences a change could cause. Hence, it enables a well-structured decision. As the principles were derived from the three big players in the mobile service market, they will help practitioners to provide a thriving platform when combined with the presented framework. However, to this point the framework and the principles do not support the development of governance for a new platform or advise what governance policies should be adapted on the platform when for example a new competitor enters the market. Moreover, at the moment, the principles are of general nature that need to be analyzed more intensely in detail by future research and there may be more factors in real

world considered by practitioners that are not yet implemented in the framework and the principles derived from literature.

We acknowledge several limitations of our work. First of all, our framework is for the most part based on a literature review provided by Manner et al. [8]. Furthermore, although our examples provide a wide range as we chose three different platforms to prove the applicability of the framework, more changes should be looked at. Finally, our examples only observe changes from a secondary source. Interviews with the platform providers questioning their motives could provide more insights

7 Conclusion and Future Research

We analyzed the literature, proposed applicable mobile platform governance framework and applied the framework to analyze multiple governance change examples. Six principles for governing mobile platforms were derived from the framework, real world and literature insights.

Providing such a structured view on the multi-dimensional concept of governance based on research findings of several disciplines, builds a more mature view on platform governance and is a valuable contribution to research and practice.

However, we believe future empirical research could enlarge the proposed mechanisms. For example, we believe a legal factor, triggering changes, has to be integrated by contract- or authority-based governance mechanisms as well as by formal controls rather than by trust-based mechanisms and informal controls. As shown in the examples, governance policies can be ignored by stakeholders if not enforced by the right tangible mechanisms. Implementing weak governance, although a legal framework does exist, is to others an opportunity for a lawsuit claim.

A cross-case study analyzing a large amount of governance changes would be an adequate method for finally evaluating the framework and enlarging the governance principles. Interviews with platform providers would contribute to this research and enable to build a platform theory [43].

References

1. Jansen, S., Brinkkemper, S., Finkelstein, A.: Business Network Management as a Survival Strategy: A Tale of Two Software Ecosystems. In: Jansen, S., Brinkkemper, S., Finkelstein, A., Bosch, J. (eds.): IWSECO 2009. CEUR-WS.org, Vol. 505, pp. 34-48, CEUR Workshop Proceedings (2009)
2. Iyer, B., Lee, C.-H., Venkatramen, N.: Monitoring Platform Emergence: Guidelines from Software Networks. Communications of the Association for Information Systems 19, 1-13 (2007)
3. Gawer, A., Henderson, R.: Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel. Journal of Economics & Management Strategy 16, 1-34 (2007)
4. Basole, R. C., Karla, J.: Entwicklung von Mobile-Plattform-Ecosystem-Strukturen und -Strategien. Wirtschaftsinformatik 53, 301-311 (2011)

5. Economides, N., Katsamakas, E.: Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry. *Management Science* 52, 1057-1071 (2006)
6. Rochet, J.-C., Tirole, J.: Platform Competition in Two-Sided Markets. *Journal of the European Economic Association* 1, 990-1029 (2003)
7. Tiwana, A., Konsynski, B., Bush, A.A.: Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research* 21, 1-23 (2010)
8. Manner, J., Nienaber, D., Schermann, M., Krcmar, H.: Governance for mobile service platforms: A literature review and research agenda. *ICMB 2012. AIS, Delft, NL* (2012)
9. Ghazawneh, A., Henfridsson, O.: Balancing platform control and external contribution in third-party development: the boundary resources model. *Information Systems Journal* 23 (2), 173-192 (2013)
10. Burkard, C., Widjaja, T., Buxmann, P.: Software Ecosystems. *Wirtschaftsinformatik* 54, 43-47 (2012)
11. Haaker, T., Faber, E., Bouwman, H.: Balancing customer and network value in business models for mobile services. *International Journal of Mobile Communications* 4, 645-661 (2006)
12. Ballon, P., Walravens, N., Spedalieri, A., Venezia, C.: The reconfiguration of mobile service provision: towards platform business models. In: *19th ITS European Regional Conference, Rome, Italy* (2008)
13. Ghazawneh, A., Henfridsson, O.: Micro-Strategizing in Platform Ecosystems: A Multiple Case Study. In: *ICIS 2011 Proceedings, Shanghai* (2011)
14. Kouris, I., Kleer, R.: Business models in two-sided markets: an assessment of strategies for app platforms. In: *ICMB 2012. AIS, Delft, NL* (2012)
15. Holzer, A., Ondrus, J.: Trends in Mobile Application Development. In: Hesselman, C., Giannelli, C. (eds.): *Mobile Wireless Middleware, Operating Systems, and Applications-Workshops*. Vol. 12, pp. 55-64. Springer, Berlin (2009)
16. Rudmark, D., Ghazawneh, A.: Third-Party Development for Multi-Contextual Services: On the Mechanisms of Control. In: *European Conference on Information Systems* (2011)
17. Brousseau, E., Penard, T.: The Economics of Digital Business Models: A Framework for Analyzing the Economics of Platforms. *Review of Network Economics* 6, 81-114 (2007)
18. De Reuver, M., Bouwman, H.: Governance mechanisms for mobile service innovation in value networks. *Journal of Business Research* 65, 347-354 (2011)
19. Ghazawneh, A., Henfridsson, O.: Governing third-party development through platform boundary resources. In: *ICIS 2010 Proceedings, Vol. 48, St. Louis* (2010)
20. Jain, A.: Apps Marketplaces and the telecom value chain. *IEEE Wireless Communications* 18 (4), 4-5 (2011)
21. Bouwman, H., Haaker, T., Faber, E.: Developing Mobile Services: Balancing Customer and Network Value. In: *The Second IEEE International Workshop on Mobile Commerce and Services*, pp. 21-31, Munich (2005)
22. Grover, V., Ayyagari, R., Gokhale, R., Lim, J., Coffey, J.: A Citation Analysis of the Evolution and State of Information Systems within a Constellation of Reference. *Journal of the Association for Information Systems* 7, 270-325 (2006)
23. Caillaud, B., Jullien, B.: Chicken & Egg: Competition among Intermediation Service Providers. *RAND Journal of Economics* 34 (2), 309-328 (2003)
24. Eisenmann, T., Parker, G., van Alstyne, M.W.: Strategies for two-sided markets. *Harvard business review* 84 (2006)

25. West, J.: How open is open enough? Melding proprietary and open source platform strategies. *Research Policy* 32, 1259-1285 (2003)
26. Eaton, B., Elaluf-Calderwood, S., Sørensen, C., Yoo, Y.: *Dynamic Structures of Control and Generativity in Digital Ecosystem Service Innovation: The Cases of the Apple and Google Mobile App Stores*. LSE, London Report 44, 1-25 (2011)
27. Busquets, J.: Orchestrating Smart Business Network dynamics for innovation. *European Journal of Information Systems* 19, 481-493 (2010)
28. West, J., O'Mahony, S.: The role of participation architecture in growing sponsored open source communities. *Industry and Innovation* 15, 145-168 (2008)
29. Kotler, P., Keller, K.L., Brady, M., Goodman, M., Hansen, T.: *Marketing Management: European Edition*. Prentice Hall, Harlow, UK (2009)
30. Hagiu, A.: Multi-sided platforms: From microfoundations to design and expansion strategies. pp. 1-26. Harvard Business School, Harvard Business School Strategy Unit Working Paper (2009)
31. De Reuver, M., Visser, A., Prieto, G., Bouwman, H.: Governance of flexible mobile service platforms. *Intelligence in Next Generation Networks* 1-6 (2010)
32. Neumann, D.: *Market Engineering: A Structured Design Process for Electronic Markets*. UVB, Karlsruhe (2007)
33. Methlie, L.B., Pedersen, P.E.: Business model choices for value creation of mobile services. *Info* 9 (5), 70-85 (2007)
34. Müller, R.M., Kijl, B., Martens, J.K.J.: A Comparison of Inter-Organizational Business Models of Mobile App Stores: There is more than Open vs. Closed. *Journal of Theoretical and Applied Electronic Commerce Research* 6, 63-76 (2011)
35. Tee, R., Gawer, A.: Industry architecture as a determinant of successful platform strategies: a case study of the i-mode mobile Internet service. *European Management Review* 6, 217-232 (2009)
36. Ouchi, W.G.: A Conceptual Framework for the Design of Organizational Control Mechanisms. *Management science* 25, 833-848 (1979)
37. Kirsch, L.S.: Portfolios of control modes and IS project management. *Information Systems Research* 8, 215-239 (1997)
38. Cardinal, L.B., Sitkin, S.B., Long, C.P.: A configurational theory of control. In: Sitkin, S., Cardinal, L.B., Bijlsma-Frankema, K. (eds.): *Control in Organizations: New Directions in Theory and Research*, pp. 51-107. Cambridge University Press, Cambridge (2009)
39. Schermann, M., Wiesche, M., Krcmar, H.: The Role of Information Systems in Supporting Exploitative and Exploratory Management Control Activities. *Journal of Management Accounting Research* (to appear) 24, 1-42 (2012)
40. Davies, C.: Apple privacy changes could impact 1000s of apps. In: Nguyen, V. (ed.): *Slashgear*, Vol. 2012. Slashgear, <http://www.slashgear.com/apple-privacy-changes-could-impact-1000s-of-apps-16213785/> (2012)
41. Lunden, I.: Google Now Playing At Apple's Game For In-App Payments? No, Just Business As Usual, Says Google. In: Butcher, M. (ed.): *Techcrunch*, Vol. 2012. <http://techcrunch.com>, <http://techcrunch.com/2012/03/09/google-now-playing-at-apples-game-for-in-app-payments-no-just-business-as-usual-says-google/> (2012)
42. Bowers, T.: App developers getting extra incentives from Microsoft. In: Hiner, J. (ed.): *TechRepublic*, Vol. 2012. TechRepublic, <http://www.techrepublic.com/blog/career/app-developers-getting-extra-incentives-from-microsoft/4183> (2012)
43. Yin, R.K.: *Case study research: Design and methods*. Sage Publications, Inc, Thousand Oaks, CA (2009)

Integration von Planungssystemen in der Logistik – Ansatz und Anwendung

Christoph Augenstein, Stefan Mutke, und André Ludwig

Universität Leipzig, Institut für Wirtschaftsinformatik, Leipzig, Germany
{augenstein,mutke,ludwig}@wifa.uni-leipzig.de

Abstract. Während noch vor wenigen Jahren die meisten Produktions- oder Handelsunternehmen eigene Logistikabteilungen unterhielten, so sind Logistikfunktionen heute weitestgehend an Logistikdienstleister in Kontraktform ausgelagert. Eine der Hauptaufgaben von Logistikdienstleistern liegt in der mehrstufigen Planung logistischer Dienstleistungen für den Kunden. Sie schließt zahlreiche, meist isoliert betrachtete Planungsschritte ein. Für die Logistikplanung stellt dieser Beitrag einen Ansatz zur Integration von Planungssystemen in der Logistik auf Basis der Modellintegration vor. Am Beispiel der Prozessplanung und der Prozesssimulation wird gezeigt wie mit diesem Ansatz Modelle zwischen Planungssystemen konsistent wiederverwendet werden können.

Keywords: Logistikplanung, Modellintegration, Modelltransformation, Planungssysteme, Simulation

1 Einleitung

Die Logistik beschäftigt sich mit der zeitlichen und räumlichen Transformation von logistischen Objekten (Material, Personen oder Informationen) und deren bedarfsgerechten Verfügbarkeit [1]. Ihre Hauptaufgabe besteht im effizienten Bereitstellen der richtigen Ware, in der richtigen Menge, in der richtigen Qualität, zur richtigen Zeit, zu den richtigen Kosten, am richtigen Ort (6R-Regel) [2].

Während noch vor wenigen Jahren die meisten Produktions- oder Handelsunternehmen eigene Logistikabteilungen unterhielten, sind Logistikfunktionen heute weitestgehend an Logistikdienstleister (LDL) in Kontraktform ausgelagert. Ausgelagerte und durch LDL in Form von logistischen Dienstleistungen angebotene Logistikfunktionen umfassen dabei die Bereiche Transport, Umschlag, Lagerung aber auch zahlreiche Zusatzbereiche wie Verpackung, Veredelung, Verzollung oder Reifung. Aufgrund unternehmensspezifischer Anforderungen (z. B. nachgefragter Funktionsumfang, Prozesse oder Informationsaustausch) und branchenspezifischer Besonderheiten (z. B. Sicherheitsanforderungen in der Chemiebranche/Gefahrgut oder exakte Rückverfolgbarkeitsanforderung in der Lebensmittelbranche) ist dabei nahezu jede ausgelagerte Dienstleistung in ihrem Funktions-/Leistungsumfang und ihren Geschäftsbedingungen individuell. Die individuellen, aus unterschiedlichen Logistikfunktionen bestehenden logistischen Dienstleistungen werden dabei als Mehrwertlogistik oder

Fourth-Party-Logistics (4PL) bezeichnet [3-4]. So genannte Mehrwert-Logistikdienstleister (Mehrwert-LDL) integrieren und organisieren Teilleistungen unterschiedlicher LDL und bieten diese als kundenspezifische Dienstleistung am Markt an. Sie wählen zusammen mit dem Kunden die jeweils am besten geeigneten LDL zur Leistungserbringung aus, optimieren und steuern den gesamten Güter- und Informationsfluss über alle beteiligten LDL. Mehrwert-LDL fungieren dabei als logistische Generalunternehmer. Sie stellen die zentrale Schnittstelle zwischen dem Kunden und den einbezogenen LDL dar, tragen die Verantwortung für die Leistung und die Qualität der in die Kette integrierten Anbieter.

Eine der Hauptaufgaben von Mehrwert-LDL liegt in der mehrstufigen Planung logistischer Dienstleistungen für den Kunden. Sie schließt zahlreiche, abhängig vom Kunden unterschiedliche Planungsschritte wie die Auswahl von Dienstleistern, die Prozessplanung oder das Aufstellen langfristiger Prognosen zur Planungsabsicherung ein (s. Abbildung 1). Jeder dieser Planungsschritte baut auf eigenen Planungsmethoden und den in ihnen verwendeten Planungssystemen auf. Dadurch können zwar die für den konkreten Planungsschritt relevanten Aspekte der realen Welt abstrahiert werden, es entstehen so jedoch nur isolierte Planungsergebnisse im Gesamtplanungsprozess, die einer isolierten Interpretation unterliegen. In [5] dagegen wurde die Integration der einzelnen Planungsschritte als wesentlicher Faktor hervorgehoben, um den Modellierungsaufwand, die Kosten und die Fehleranfälligkeit gering zu halten.

In vorliegendem Beitrag wird ein Ansatz vorgestellt, mit dem sich heterogene Modelle von Logistikplanungssystemen integrieren lassen. Auf Basis der Modellintegration kann eine ganzheitliche Sicht auf die Planung logistischer Mehrwertdienste hergestellt werden. Zuerst wird ein Überblick über verwandte Ansätze gegeben (Abschnitt 2). Am Beispiel heterogener Planungsmodelle und Leistungsbeschreibungen wird herausgestellt, wie sich Modelle in unterschiedlichen Planungsschritten grundsätzlich aufeinander abbilden und einheitlich wiederverwenden lassen (Abschnitt 3). Die Modellintegration wird am Beispiel der Prozessplanung und der anschließenden Prozesssimulation vorgestellt (Abschnitt 4). Abschließend werden die wesentlichen Elemente der Planung zusammengefasst (Abschnitt 5).

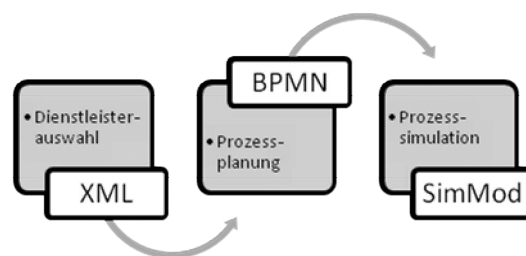


Abb. 1. Beispiel einer mehrstufigen Logistikplanung und den darin verwendeten Planungssystemen und Modellen

2 Verwandte Ansätze

Die Umsetzung einer integrierten Planung auf Seiten des Mehrwert-LDL führt zur Integration mehrerer Modelle und zu einer Überführung dieser Modelle in Weitere. In diesem Beitrag wird ausgehend von Prozessmodellen und einer formalisierten Beschreibung von Dienstleistern gezeigt, wie diese Modellinformationen integriert und schließlich in ein Simulationsmodell überführt werden. Für die spezifische Aufgabe einer Modellintegration und Transformation im Rahmen der Logistikdomäne sind bislang jedoch keine umfassenden Ansätze bekannt. Dagegen lassen sich verwandte Ansätze für den Bereich der Modellintegration in Arbeiten zur Metamodellierung bzw. Metamodellierungsplattformen sowie Modelltransformationen finden. Für den Ansatz zur Absicherung der Planung mittels einer Simulation sind ebenfalls Modelltransformationen, im Rahmen der Überführung von Prozessmodellen zu Simulationsmodellen, sowie Simulationsansätze für logistische Systeme vorhanden. Nachfolgend werden in drei Sektionen Beispiele für solche Ansätze genannt und kurz beschrieben.

Metamodellierung. Metamodellierungsplattformen [6-7] sind Ansätze zur Verwaltung von Metamodellen bzw. deren Modellen u. a. zum Zwecke der Versionierung. Braun und Winter modellieren in ihrem Ansatz schichtenweise eine Unternehmensarchitektur, auf der Metamodelle pro Schicht verwaltet und in einen integrativen Rahmen gebracht werden sollen. Generell wird jedoch nicht die Verbindung verschiedener Metamodelle angestrebt.

Modelltransformationen. Modelltransformationen werden in den Bereichen (modellgetriebene) Softwareentwicklung bzw. Softwarearchitekturen verwendet. Einen Überblick über Transformationsarten liefert [8]. Sie sind für diesen Ansatz aus zwei Perspektiven relevant. Zum einen dienen Transformationen dazu Modellkorrespondenzen zwischen unterschiedlichen Modellen zu identifizieren und zu realisieren. Beispiele für diese Art der Nutzung sind sowohl [9] als auch [10], die jeweils für einen multiperspektivischen Ansatz Korrespondenzen zwischen den Perspektiven ermitteln. Erstere definieren übergreifende Aspekte für Softwarearchitekturen sowohl intensional auf Metamodellebene als auch extensional auf Modellebene. Selonen und Kettunen erarbeiteten einen Ansatz für die Softwareentwicklung, um Modelle, auf demselben Metamodell basierend, mittels Regeln miteinander zu verbinden. Schließlich existieren mit [11] und [12] Ansätze, die Softwaremodelle mit Hilfe von UML zueinander in Beziehung setzen.

Zum anderen sind Transformationsansätze auch für die Entwicklung von Transformationsmodellen bei der Überführung von Prozess- in Simulationsmodelle relevant. Sowohl Petsch et al. als auch Kloos et al. stellen in [13-14] ein Transformationsmodell als Zwischenschritt bei der Umwandlung von Prozess- zu Simulationsmodellen vor. Bei beiden Ansätzen wird davon ausgegangen, dass Prozessmodelle ohne Berücksichtigung von Simulationsanforderungen vorhanden sind, um bspw. die Transparenz von Prozessen zu erhöhen, um Mitarbeiter zu schulen und zur Dokumen-

tation für Zertifizierungen. Die Ansätze setzen allerdings voraus, dass die Prozesse als EPK (ereignisgesteuerte Prozessketten) modelliert sind.

Simulationsansätze. Es gibt eine Reihe von weitgefächerten Ansätzen die sich mit Simulation bei der Planung von Logistiksystemen befassen. Ingalls diskutiert in [15] den Nutzen von Simulation als Analysemethode zur Untersuchung von Supply Chains. Zudem umfasst die Arbeit eine Auflistung von Vor- und Nachteilen. Ein konkreter Ansatz wird jedoch nicht vorgestellt. Cimino et al. stellt ein allgemeines Simulations-Framework zur Modellierung von Supply Chains vor [16]. Zudem wird in diesem Beitrag ein Überblick über verschiedene ereignisdiskrete Simulationswerkzeuge im Hinblick auf die Anwendungsdomäne und die Nutzung von Programmiersprachen als Alternative zu einer ereignisdiskreten Simulation (EDS) gegeben.

3 Integration von Planungssystemen durch Modellintegration

Mit den Modellen in der Logistikplanung (s. Abbildung 1) werden spezifische Informationen eines Dienstes festgehalten. Jedes Planungswerkzeug besitzt zur Modellierung ein Metamodell als formale Basis, um bestimmte Aspekte zu modellieren. Die Ablaufplanung einer Dienstleistung kann beispielsweise in der Prozessmodellierung mit Hilfe der Sprache „Business Process Modeling and Notation“ (BPMN) modelliert werden. Eine logistische Dienstleistung wird dabei sukzessive durch verschiedene Modelle umfänglich beschrieben. Während der Planung werden dadurch Modelle erstellt, die je nach Art disjunkte oder überlappende Informationen beinhalten. Verstärkt wird dieser Zustand dadurch, dass mehrere Personen beteiligt sind, sodass derselbe Sachverhalt einer unterschiedlichen Modellierung (Synonyme, Homonyme, alternative Modellierungsarten) unterworfen sein kann. Gleichzeitig muss gewährleistet werden, dass weitere Werkzeuge im Planungsprozess hinzugefügt werden können und dass die Durchführung der Planung, Steuerung und Kontrolle möglichst effizient geschieht. Das wird ermöglicht, indem bereits modellierte Informationen in anderen Modellen wiederverwendet werden, womit gleichzeitig auch die Problematik der unterschiedlichen Modellierung verringert wird.

Das durch Augenstein et al. in [17] vorgestellte Service Repository als prototypische Implementierung eines Service Modeling Frameworks (SMF) legt den Grundstein für eine Modellintegration von Planungssystemen. Im Service Repository werden die in Werkzeugen verwendete Modelle zentral verwaltet und auf Metaebene so miteinander verknüpft, dass Informationen aus bestehenden Modellen extrahiert und wiederverwendet werden. Im SMF wird analog zur adaptiven Referenzmodellierung nach [18] jedes Modell als eine Projektion auf ein Gesamtmodell angesehen. Allerdings wird für eine Dienstleistungsbeschreibung nicht die a priori Erstellung eines Referenzmodells bevorzugt, da sich gezeigt hat, dass die Entwicklung eines solchen umfänglichen Dienstleistungsmodells bereits mehrfach (s. z. B. [19-21]) vorangetrieben wurde, ohne jedoch einen Konsens bzgl. des Inhalts zu finden. In [17] wurde dazu auch detailliert die Problematik eines Gesamtmodells für Dienstleistungen diskutiert. Einzelne Modelle einer Dienstleistung werden jeweils als ein Aspekt der Dienstleis-

tung und ein Gesamtmodell dieser Dienstleistung somit als Obermenge der verfügbaren Modelle aufgefasst.

Um Modelle verschiedener Modellierungssprachen miteinander verknüpfen zu können, nutzt das SMF eine Modellhierarchie, wie in Abbildung 2 dargestellt. Da nach [7] die wesentlichen Elemente einer Modellierungssprache ihre Syntax, ihre Semantik und ihre Notation sind, müssen zur Verbindung von Modellen verschiedener Sprachen zumindest Syntax als auch Semantik vereinheitlicht werden. Die Notation als reine Darstellungsform kann für diese Zwecke zunächst vernachlässigt werden. Eine einheitliche Syntax wird im SMF durch ein vorgegebenes Metametamodell erreicht, mit dem alle verwendeten Modellierungssprachen beschrieben sein müssen. Somit wird die Lücke auf syntaktischer Ebene zwischen diesen Sprachen geschlossen und eine gemeinsame Verarbeitung ermöglicht. Der Ansatz und das Repository bauen hierzu auf dem EMOF¹ (Essential Meta Object Facility) kompatiblen Metamodell Ecore² der Eclipse Foundation auf.

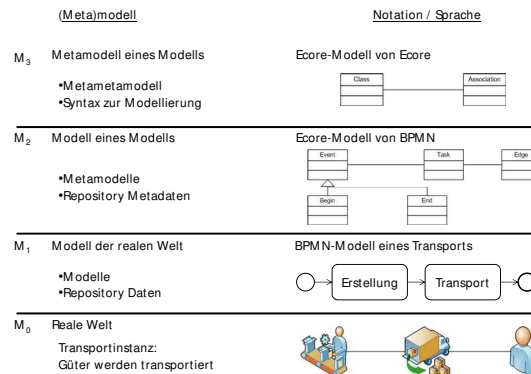


Abb. 2. Modellhierarchie im Repository

Zusätzlich muss auch die semantische Lücke zwischen Modellen mit vergleichsweise hohem Aufwand geschlossen werden. Um ein Modell in ein semantisch äquivalentes Modell zu überführen, werden Transformationsregeln auf der Metaebene definiert. Im Vergleich zu bestehenden Ansätzen der Modelltransformation (s. Abschnitt 2) wird mit dem SMF ein anderer Ansatz zur Modellintegration verfolgt. Bestehende Ansätze der Modell-zu-Modell-Transformation erstellen auf Metaebene Beziehungen zwischen Elementen und führen teilautomatisiert eine Transformation auf Modellebene durch. Problematisch sind die folgenden Aspekte: Zum einen werden Modelle auf Metaebene direkt miteinander verknüpft, sodass bei Hinzunahme eines weiteren Modelltyps zunächst weitere Transformationen definiert werden müssen. Zum anderen können Transformationen nur durch Vergleich der abstrakten Syntax automatisiert durchgeführt werden, so dass ein zusätzlicher manueller Schritt notwendig wird, um eine vollständige Transformation zu gewährleisten (s. hierzu die Etablierung

¹ Siehe hierzu: <http://www.omg.org/mof/>

² Siehe hierzu: <http://www.eclipse.org/modeling/emf/?project=emf>

extensionaler Verbindungen bei Romero et al. [9] oder Intermodellkorrespondenzen bei Selunen et al. [10]). Modelltransformationen sind dann sinnvoll, wenn sowohl mit dem Ausgangs- als auch mit dem Zielmodell ein ähnlicher Sachverhalt modelliert werden soll (z. B. die Transformation eines BPMN-Modells in ein BPEL-Modell zur Automatisierung). Das SMF umfasst jedoch Modelle, die unterschiedliche Aspekte einer Dienstleistung betreffen (z. B. Prozess, technische Beschreibungen, Kennzahlen oder Preismodelle). Daher wird die semantische Lücke zwischen Modellen nicht durch die Angabe von Transformationsregeln geschlossen, sondern durch ein separates Modell, das relevante Modellelemente deskriptiv miteinander verbindet. Zusätzlich fließen Elemente der Fachdomäne, in diesem Fall der Logistik, über ein weiteres Modell mit in die Beschreibung der Verbindung ein und erlauben damit eine noch präzisere Beschreibung.

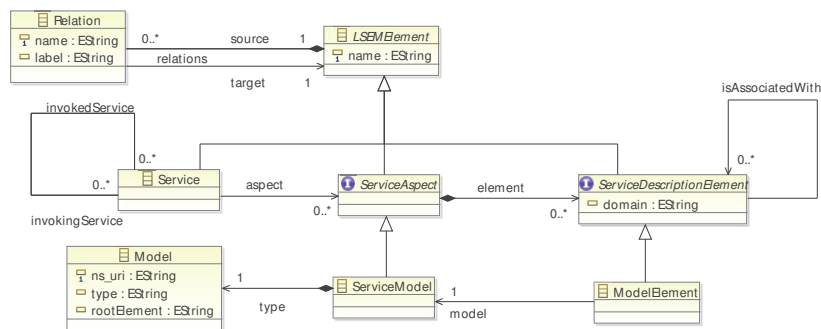


Abb. 3. Basismetamodell (Ausschnitt)

Das Basismetamodell, welches in [17] als Common Service Metamodel (CSM) bezeichnet und in Abb. 3 auszugsweise dargestellt wird, übernimmt die Aufgabe, semantisch ähnliche Modellelemente miteinander zu verbinden. Es beschreibt zum einen die Grundstruktur des SMF, indem Services, Modelle und deren Elemente zueinander in Beziehung gesetzt werden. Zum anderen setzt es die Modelle bzw. deren Elemente über eine zusätzliche Abhängigkeit (**ServiceAspect** und **ServiceDescriptionElement**) zueinander in Beziehung. Jedes Modell wird als ein Aspekt einer Dienstleistung und jedes Modellelement, das zu anderen in Beziehung gesetzt werden soll, als ein **ServiceDescriptionElement** beschrieben.

Am Beispiel eines Dienstleistermetamodells (s. Abbildung 4), welches Kennzahlen zu LDL enthält, soll diese Verbindung näher erläutert werden. Das Modell charakterisiert einen LDL durch die Angabe von Kontaktinformationen, durch Kennzahlen, die zusätzlich in die Kategorien Kosten, Qualität und Zeit eingeteilt werden und durch die Dienste, die er anbietet. Neben den Diensten besitzt das Dienstleistermetamodell zwei weitere Verbindungen (Vererbungsbeziehungen) zum Basismetamodell. Jeder LDL erbringt eine spezifische logistische Funktionalität (z. B. Transport, Umschlag oder Lagerung). Das Modell selbst wird als ein Aspekt (**ServiceAspect**) eines Dienstes gekennzeichnet und die Elemente, die die Kennzahlen des Dienstleisters im Modell repräsentieren, werden als **ServiceDescriptionElement** modelliert.

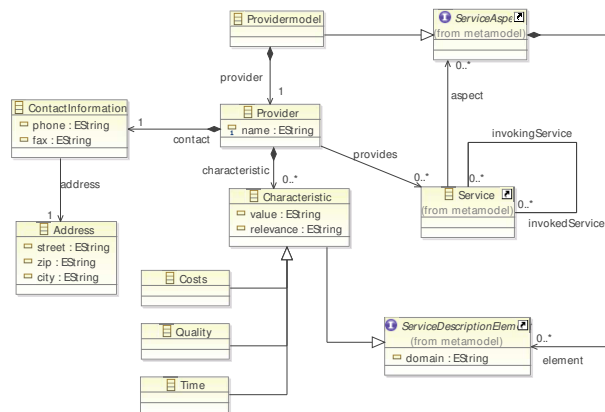


Abb. 4. Dienstleistermetamodell

Damit stehen diese Informationen auch anderen Modellen zur Verfügung, im Gegensatz zu den Kontaktinformationen, die ohne weitere Verbindung zum Basismetamodell nur innerhalb eines Dienstleistermodells Verwendung finden. Mit Hilfe des Basismetamodells wird damit auch die semantische Lücke zwischen Modellen im SMF geschlossen. Auf Modellebene semantisch äquivalente Informationen oder Informationen, die in anderen Modellen wiederverwendet werden sollen, werden auf die beschriebene Art gekennzeichnet und stehen im weiteren Verlauf der Planung zur Verfügung. Dazu werden in einem separaten Modell des SMF Beziehungen auf Modell- und Modellelementebene modelliert. Die Beziehungen, die zwischen den Service-DescriptionElementen modelliert werden, steuern dabei die Extraktion der Modellinformation für andere Werkzeuge. Die eingangs geforderten Eigenschaften der Informationswiederverwendung und Erweiterbarkeit werden durch das einheitliche Metamodell und das Basismetamodell erfüllt. Einzige Einschränkung besteht darin, dass nur solche Modelle integriert werden können, deren Metamodelle sich durch das Metametamodell beschreiben lassen.

4 Anwendung am Beispiel der Prozessplanung und -simulation

Die Planung logistischer Mehrwertdienste wird mittels mehrerer Modelle durchgeführt. Eine grobe Planung der einzelnen Teilleistungen und deren zeitlichen Abhängigkeiten werden mit einem Prozess dargestellt. Nach [22] wird ein solcher Prozess als zusammenhängende, abgeschlossene Folge von Tätigkeiten beschrieben, die zur Erfüllung einer betrieblichen Aufgabe notwendig sind. [23] ergänzt, dass diese Tätigkeiten zur Erzeugung eines Produktes oder einer Dienstleistung dienen. Wenngleich hier zunächst die betriebliche Dimension eines Prozesses genannt wird, so werden Prozesse im überbetrieblichen Kontext mit demselben Zweck verwendet. Je nach

Ausrichtung werden Bausteine, die für die Modellierung eines Prozesses relevant sind, funktional oder strukturell beschrieben. [22] fokussiert die funktionale Sicht und nennt daher Funktionen, Ereignisse, Organisationseinheiten, Informationsobjekte und Operatoren als wesentliche Elemente. [23] fokussiert die strukturelle Sicht eines Prozesses, bzw. dessen Ablauf und benennt Ereignisse, Zerlegungen, Sequenzen, Auswahlen und Parallelitäten. Die Prozessmodellierung als Aktivität zur Definition und Beschreibung eines Ablaufs verbindet Ausführende (Organisationseinheiten) mit Ausführungsobjekten (Informationsobjekte) über Aufgaben (Tätigkeiten) und verbindet diese über unterschiedliche Kontrollflüsse miteinander. Anwendungsgebiete bzw. Zweck der Prozessmodellierung sind beispielhaft die Dokumentation, die Vorbereitung zur Automatisierung oder auch die Optimierung. [24] benennt die Simulation zur Untersuchung des Systemverhaltens im Zeitablauf als weiteres Anwendungsgebiet.

Die Anwendung von Simulationstechniken in der Logistik helfen dem LDL logistische Mehrwertdienstleistungen effizienter zu gestalten. Da sie für eine längere Kontraktdauer konzipiert sind, sind tiefgreifende Änderungen während der Laufzeit sehr kostenintensiv und sehr zeitaufwendig [25]. Somit ist es bereits während der Planungsphase notwendig, Aussagen über das zukünftige Verhalten der Lieferkette zu treffen. Zur methodischen Absicherung der Planung, Steuerung und Überwachung der Material-, und Informationsflüsse haben Simulationstechniken in der Logistik inzwischen einen hohen Stellenwert eingenommen [26]. EDS von Logistiknetzwerken eignen sich, um die Entscheidungsunterstützung innerhalb der Planungsphase zu verbessern, indem mehrere Systemvarianten analysiert werden, die sich in Struktur und Verhalten unterscheiden [27]. Somit kann das System vor der eigentlichen Umsetzung untersucht werden. Eine Simulation benötigt neben dem eigentlichen Modell zusätzlich eine Umgebung, innerhalb der das Modell lauffähig ist. Ein Simulationsmodell wird daher grundsätzlich nur für eine bestimmte Umgebung entwickelt. Sind diese beiden Voraussetzungen gegeben, kann mit einem Simulationsexperiment das Modellverhalten zielgerichtet untersucht werden. Allgemein besteht ein Simulationsmodell aus folgenden Konzepten:

Tabelle 1. Konzepte eines Simulationsmodells

Entities	Repräsentiert ein Realwelt-Objekt. Dynamische Entities bewegen sich durch das System (z. B. Produkt, Kunde), statische Entities bedienen andere Entities (z. B. Förderband, Lager)
Events	Ein Event verändert den Zustand eines Simulationssystems. Events sind immer Start und Ende einer Activity oder Delay (z. B. Fracht ist verladen)
Attributes	Attributes charakterisieren Entities (z. B. Bearbeitungszeit, Kapazität, Farbe). Eine Gruppe von Entities kann die gleichen Attributes besitzen und variiert nur in der Ausprägung
Activities	Repräsentiert eine bestimmte Zeitspanne. Die Zeitdauer ist im Vorfeld bekannt (z. B. Bearbeitungszeit einer Maschine)
Delays	Repräsentiert eine unbestimmte Zeitspanne. Diese hängt von den Kombinationen des Systemzustandes ab (z. B. Fracht wartet auf Einlagerung)

Die wesentliche Aufgabenstellung einer Simulation in der Logistik ist die Untersuchung der Transportmengen und Kapazitäten der Teildienstleistungen, um sicherzustellen, dass der Kundenbedarf erfüllt werden kann. So ist es möglich, die Durchläufe der Güter durch das Logistiksystems hinsichtlich der Kapazitäten zu analysieren, um Engpässe frühzeitig identifizieren zu können. Für die Bildung dieser Simulationsmo-

delle werden in der Logistikdomäne hauptsächlich anwendungsorientierte Modellierungskonzepte mit anwendungsnahen Beschreibungsmitteln eingesetzt. Typische Vertreter sind „Bausteinkonzepte“. Diese stellen topologische, organisatorische und/oder informatorische Elemente – zweckmäßig aggregiert und vordefiniert sowie aus Anwendungssicht parametrisierbar – für ein bestimmtes Anwendungsfeld zur Verfügung [26]. In [5] wurden darüber hinaus speziell die Ziele und Anforderungen eines Mehrwert-LDL an einen Simulationsansatz näher diskutiert und die Notwendigkeit eines Ansatzes zur Transformation von Prozess- zu Simulationsmodellen hervorgehoben. Als wichtigste Anforderung wurde die Integration des Simulationsansatzes in den gesamten Planungsprozess herausgestellt. Der Mehrwert des Ansatzes liegt somit in der automatischen Transformation von bestehenden Prozessmodellen zu Simulationsmodellen.

Zwar können Prozessmodelle als Basis für die Erzeugung eines Simulationsmodells verwendet werden, für eine Simulation sind jedoch zusätzliche Informationen als zur reinen Visualisierung der Prozesse notwendig [14]. In [28] wird ein Ansatz zur formalen und semantischen Beschreibung von Dienstleistungen aus der Logistikdomäne unter Verwendung von Konzepten der Serviceorientierung und von Semantic-Web Technologien vorgestellt. Der Ansatz kategorisiert und beschreibt zudem die modularen Logistikdienstleistungen Transport, Umschlag, Lagerung, Mehrwertdienstleistungen, usw. in Form einer Ontologie. Diese wurde im Rahmen der Forschungsarbeit verwendet, um die BPMN Prozessmodelle mit zusätzlichen Informationen zu annotieren. Aktivitätselemente von BPMN repräsentieren Logistikdienstleistungen, die durch die Ontologie mit mehr Informationen angereichert werden. Um diese Idee umzusetzen wurde das BPMN Ecore-Metamodell erweitert (s. Abbildung 5 links). Die identifizierten Elemente und Attribute, die für die Simulation logistischer Prozesse benötigt werden, wurden durch ein eigenes Ecore-Modell umgesetzt und bilden zusammen mit den BPMN Ecore-Modell das logistische BPMN Metamodell. Die rechte Seite in Abbildung 5 zeigt die Umsetzung in einem Editor. Jeder Aktivität (BPMN Task) kann mittels bereitgestellter Annotation der Typ der Dienstleistung sowie die Zeitdauer, Abweichung usw. zugeordnet werden. Zudem kann für „Sequence Flows“ (Kontrollstruktur) nach einem Gateway die Wahrscheinlichkeit des Ablaufs des Pfades zugewiesen werden. Diese Daten werden aus dem Providermodell im Repository bereitgestellt, können aber zudem manuell angegeben werden. Zudem wurde das Simulationswerkzeug angepasst. Wie oben beschrieben wird innerhalb des Ansatzes ein „Bausteinkonzept“ verwendet. So wurden in dem Werkzeug Entitäten mit den dazugehörigen Attributen als Bausteine erstellt und können somit jederzeit wiederverwendet werden. Somit entspricht eine Aktivität im Prozessmodell einem Baustein im Simulationsmodell. Die Ontologie wird als Basis für beide Modelle verwendet.

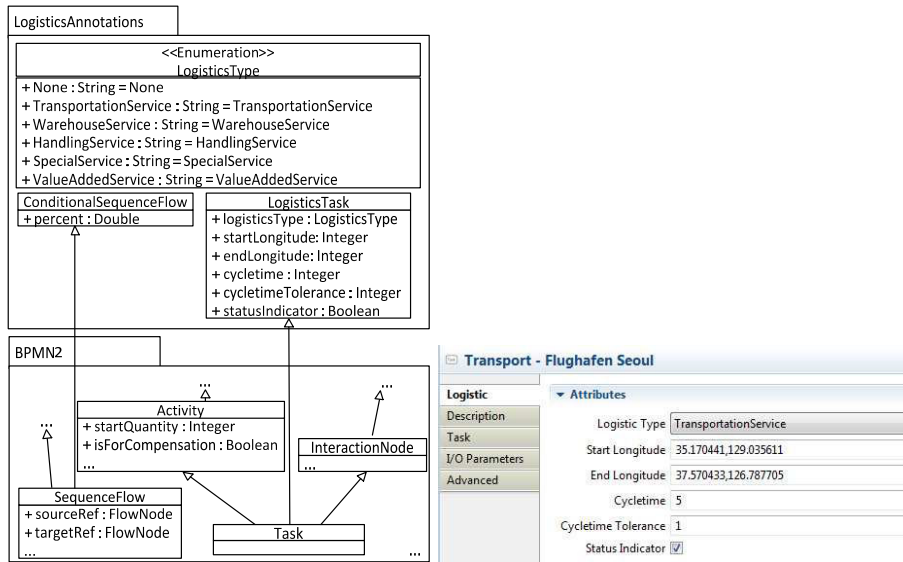


Abb. 5. UML Klassendiagramm der logistischen Elemente, Umsetzung im BPMN Editor

Für die Transformation von Prozessmodellen zu Simulationsmodellen wird eine Modell-zu-Text (M2T) Transformation angewendet. Diese Aufgabe kann durch das Eclipse Projekt Xtext³ übernommen werden. Abbildung 6 verdeutlicht den Ablauf und zeigt die einzelnen Bestandteile der Transformation.

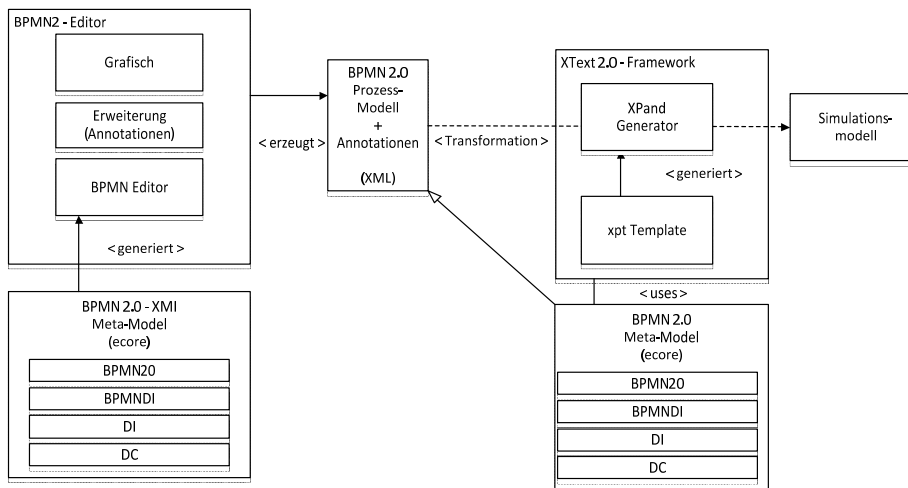


Abb. 6. Transformationsvorgehen vom Prozess- zum Simulationsmodell

³ <http://www.eclipse.org/Xtext/>

Mit dem BPMN Prozesseditor, dem das logistische Ecore Metamodell zugrunde liegt, wird ein BPMN Prozess modelliert und mit Annotationen angereichert. Das Prozessmodell steht neben der grafischen Repräsentation auch als XML Repräsentation zur Verfügung, welche für die Transformation verwendet wird. Das Xtext Framework nutzt ebenfalls das logistische Ecore Metamodell zur Auswertung des Prozessmodells in XML Repräsentation. Der Xpand Generator wertet die einzelnen Elemente des Quellmodells aus und erzeugt auf Basis eines xpt-Template das Simulationsmodell.

Beispielszenario

Zum besseren Verständnis der Aktivitäten innerhalb der Grobplanung wird ein mögliches Anwendungsszenario beschrieben. Dabei wird das Zusammenspiel des Service Repository mit den Planungswerkzeugen veranschaulicht. Das nachfolgend beschriebene Szenario fokussiert den Entwicklungsaspekt einer Logistikdienstleistung und beschreibt Schritte und Modelle, die innerhalb der Planungsphase auftreten:

Ein Produktionsunternehmen benötigt für die Endmontage an den Fertigungsbändern diverse Kleinteile (z. B. Schrauben, Befestigungsmaterial, Spritzgusskleinteile, Sicherungen). Aus Kostengründen bezieht es diese von mehreren Anbietern aus Asien. Dazu spezifiziert es Anforderungen (bspw. Art der Güter, durchschnittliche Bestellmengen, spezielle Verpackungen, Lieferzeiten/Durchlaufzeiten, Start-/Zielort, usw.), die beispielhaft in Tabelle 2 (links) definiert sind. Als Generalunternehmer übernimmt ein Mehrwert-LDL die Beschaffungslogistik und führt aufgrund der Kundenanforderung eine Grobplanung der Lieferkette durch (s. Abbildung 7).

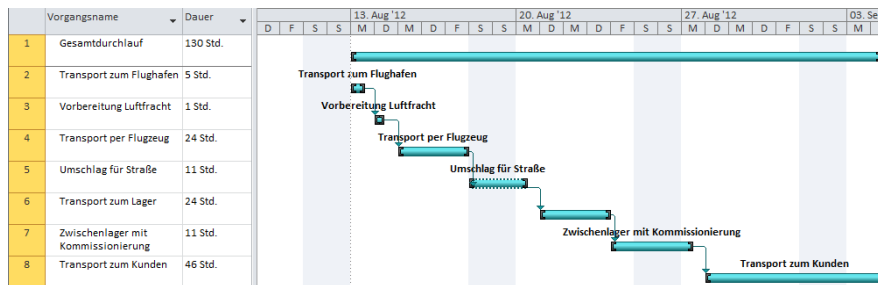


Abb. 7. Zeitliche Grobplanung des Mehrwert-LDL

Im Szenario werden die Kleinteile in zwei Werken hergestellt und anschließend von den Produzenten in den Werken auf Paletten verpackt. Der Mehrwert-LDL ist ab diesem Punkt für die gesamte Logistik bis zum Kunden verantwortlich. Im ersten Schritt muss der Transport vorbereitet werden. Dazu gehören die Zollabwicklung und das Einholen aller benötigten Papiere. Danach werden die Paletten im Werk verladen und zum Flughafen per LKW transportiert. Am Flughafen angekommen, werden die Paletten in spezielle Container luftfrachttauglich verladen und für den Lufttransport vorbereitet. Die Container werden per Flugzeug nach Deutschland transportiert. Am Zielflughafen werden die Kleinteile wieder auf Paletten verpackt und per LKW in ein Zentrallager befördert und nach einer Qualitätsprüfung eingelagert. Aus dem Zentral-

lager werden gemäß der durchschnittlichen Bestellmenge die Zwischenlager an den Produktionsstandorten des Produktionsunternehmens und schließlich die Produktionsbänder „Just in Time“ per LKW beliefert.

Tabelle 2. Kundenanforderung (links), Anforderung an Teildienstleistung (rechts)

Transportgut	Autozubehör	Name der Dienstleistung	Transport Flughafen Seoul 2	Umschlag Luftfracht Flughafen Seoul	Transport Flughafen Berlin
Bestellmenge	1 Palette/36 Std.	Art der Dienstleistung	Transport Straße	Umschlag Luftfracht	Transport Luft
Durchlaufzeit	6 Tage	Durchlaufzeit	5 Std.	1 Std.	24 Std.
Servicegrad	98%	Servicegrad	99,50%	99,80%	99,50%
Produktionsort	Südkorea	Startort	Busan	Seoul	Seoul
Zielort	Kassel (DE), Toledo (ES), Sibiu (RO), Östersund (SW)	Zielort	Seoul	Seoul	Seoul

Die zeitliche Grobplanung ist zugleich die Grundlage für die Identifizierung der Teilleistungen und damit für die Anbietersauswahl. In einem weiteren Schritt werden die Anforderungen des Kunden auf Anforderungen an die Teilleistungen herunter gebrochen (s. Tabelle 2 rechts) und potenzielle Anbieter auf Basis der verfügbaren Dienstleistermodelle ausgewählt. Das Verfahren zum Vergleich der Anforderungen mit den Kapazitäten der Anbieter wird aus Platzgründen nicht näher beschrieben. Im Ergebnis wurde am Ende des Verfahrens zu jedem Teilschritt der Gesamtleistung, dargestellt als Prozessmodell, ein Dienstleister zugeordnet. Ermöglicht wird dies über die in Abschnitt 3 beschriebene Verknüpfung des Dienstleistermetamodells und des Prozessmetamodells mit dem Basismetamodell. Für die anschließende Transformation wird das annotierte und erweiterte Prozessmodell aus dem Repository geladen und mit Hilfe des Generators in das Simulationszielmodell umgewandelt. Das Ergebnis ist ein in Enterprise Dynamics⁴ (ED) ausführbares Simulationsmodell der Logistikkette (s. Abbildung 8). Führt der Mehrwert-LDL das Simulationsmodell aus, stellt er fest, dass innerhalb des Prozesses ein Engpass besteht. Die Anforderungen an diese Teilleistung für den Transport vom Flughafen zum Zwischenlager müssen daher nachträglich angepasst werden, da nach Einschwingen des Systems die Auslastung des Dienstleisters demnach 99,8% beträgt und somit das Lager am Flughafen vollläuft.

Mit diesem integrativen Planungsansatz kann der Mehrwert-LDL effizient Abweichungen von den Anforderungen erkennen und über Änderungen im Prozessmodell alternative Simulationsmodelle erzeugen. Durch geänderte Anforderungen kann beispielsweise ein neuer Dienstleister ausgewählt werden. Eine weitere Möglichkeit besteht darin, eine weitere Dienstleistung und somit einen weiteren Prozessschritt für die Logistikkette zu modellieren. Durch die modellgetriebene Entwicklung des Simulationsmodells mittels eines Generators kann durch die beschriebene Transformation bzw. Erweiterung ein neues Simulationsmodell erzeugt und ausgeführt werden.

⁴ incontrol Simulation Solutions – <http://www.incontrolsim.com/>

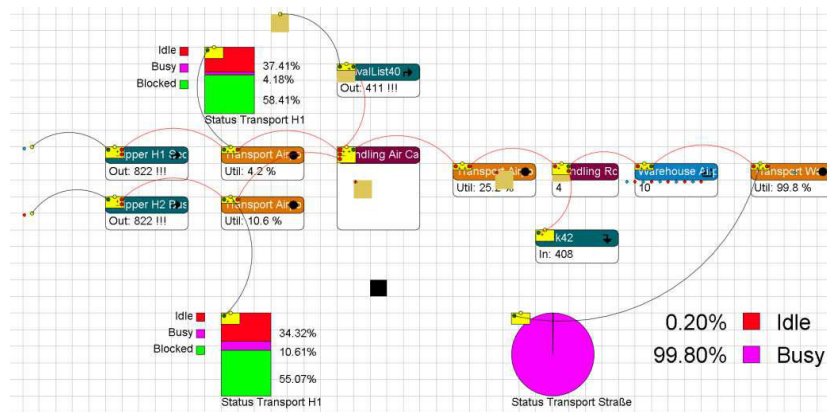


Abb. 8. Simulationsmodell (Ausschnitt) des Beispielszenarios in ED

5 Zusammenfassung

Eine der Hauptaufgaben von Mehrwertlogistikdienstleistern liegt in der mehrstufigen Planung logistischer Dienstleistungen für den Kunden. Sie schließt zahlreiche, meist isoliert betrachtete Planungsschritte und Planungssysteme ein. Zur integrierten Logistikplanung stellte dieser Beitrag einen Ansatz zur Integration von Planungssystemen in der Logistik auf Basis der Modellintegration vor. Aus diesem Grund wurde das Service Modeling Framework und das auf dieser Grundlage entwickelte Service Repository vorgestellt. Am Beispiel der Prozessplanung und der Prozesssimulation wurde dargestellt, wie mit diesem Ansatz Modelle zwischen Planungssystemen konsistent wiederverwendet werden können und wie die technische Umsetzung des Lösungsansatzes in diesem Anwendungsfall gestaltet wurde. Dadurch kann mit einem simulationsbasierten Ansatz die Planung so abgesichert werden, dass Anforderungen des Kunden möglichst präzise zugleich aber kosteneffizient eingehalten werden. Die Entwicklung des Simulationsmodells wird durch die im Repository vorgehaltenen Modelle ermöglicht. Die Schwierigkeit der Transformation des Prozessmodells wurde gelöst, indem das existierende BPMN-Metamodell mit Annotationen erweitert wurde, um dann in einem modellgetriebenen Verfahren das Simulationsmodell zu generieren. Wird mit Hilfe der Simulation eine Kombination von Teilleistungen und deren Zuordnung zu Dienstleistern gefunden, so ist damit die Planungsphase abgeschlossen.

Danksagung

Die Autoren bedanken sich für die Unterstützung durch das Bundesministerium für Bildung und Forschung (BMBF) welches im Rahmen des Förderprojekts Logistik Service Engineering und Management (BMBF 03IPT504X) die Erarbeitung dieses Beitrags gefördert hat.

Literatur

1. Arnold, D., Isermann, H., Kuhn, A., Tempelmeier, H.: Handbuch Logistik. Springer, Berlin Heidelberg (2008)
2. ten Hompel, M., Schmidt, T., Nagel, L.: Materialflusssysteme: Förder- und Lagertechnik. Springer, Berlin Heidelberg (2007)
3. Schmitt, A.: 4PL-Providing-TM als strategische Option für Kontraktlogistikdienstleister: eine konzeptionell-empirische Betrachtung. Dt. Univ.-Verl, Wiesbaden (2006)
4. Nissen, V., Bothe, M.: Fourth Party Logistics - Ein Überblick. Logistik Management 4 (1), 16-25 (2002)
5. Mutke, S., Klinkmüller, C., Ludwig, A., Franczyk, B.: Towards an Integrated Simulation Approach for Planning Logistics Service Systems. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.): BPM 2011 Workshops. LNBIP, Vol. 99, pp. 306-317. Springer, Berlin Heidelberg (2012)
6. Braun, C., Winter, R.: A Comprehensive Enterprise Architecture Metamodel and Its Implementation Using a Metamodeling Platform. In: Desel, J., Frank, U. (eds.): Enterprise Modelling and Information Systems Architectures Workshop 2005. LNI, Vol. P-75, pp. 64-79. GI, Bonn (2005)
7. Karagiannis, D., Kühn, H.: Metamodelling Platforms. In: Bauknecht, K., Tjoa, A., Quirchmayr, G., (eds.): E-Commerce and Web Technologies. LNCS, Vol. 2455, pp. 451-464. Springer, Berlin Heidelberg (2002)
8. Czarnecki, K., Helsen, S.: Classification of Model Transformation Approaches. In: Bettin, J., van Emde, G., Agrawal, A., Willink, E., Bezivin, J., (eds.), 2nd OOPSLA Workshop on Generative Techniques in the context of Model Driven Architecture, Anaheim, USA, pp. 1-17 (2003)
9. Romero, J.R., Jaén, J.I., Vallecillo, A.: Realizing Correspondences in Multi-viewpoint Specifications. In: IEEE International Enterprise Distributed Object Computing Conference (EDOC 2009) pp. 163-172. IEEE (2009)
10. Selonen, P., Kettunen, M.: Metamodel-based inference of inter-model correspondence. In: Krikhaar, R., Verhoef, C., Di Lucca, G.A. (eds.): 11th European Conference on Software Maintenance and Reengineering (CSMR '07), pp. 71-80. IEEE (2007)
11. Akehurst, D.H.: Model Translation: A UML-based specification technique and active implementation approach. University of Kent, Kent (2000)
12. Anwar, A., Dkaki, T., Ebersold, S., Coulette, B., Nassar, M.: A Formal Approach to Model Composition Applied to VUML. In: 16th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS 2011), pp. 188-197. IEEE (2011)
13. Petsch, M., Schorcht, H., Nissen, V., Himmelreich, K.: Ein Transformationsmodell zur Überführung von Prozessmodellen in eine Simulationsumgebung. In: Loos, P., Nüttgens, M., Turowski, K., Werth, D. (eds.): Modellierung betrieblicher Informationssysteme - Modellierung zwischen SOA und Compliance Management. Köllen (2008)
14. Kloos, O., Schorcht, H., Petsch, M., Nissen, V.: Dienstleistungsmodellierung als Grundlage für eine Simulation. In: Dienstleistungsmodellierung 2010. Physica (2010)
15. Ingalls, R.G.: The value of simulation in modeling supply chains. In: WSC '98 Proceedings, pp. 1371-1376. IEEE Computer Society Press (1998)
16. Cimino, A., Longo, F., Mirabelli, G.: A general simulation framework for supply chain modeling: state of the art and case study. Arxiv (2010)
17. Augenstein, C., Ludwig, A., Franczyk, B.: Integration of service models – preliminary results for consistent logistics service management. In: SRII Global Conference, pp. 100-109. IEEE (2012)

18. Delfmann, P.: Adaptive Referenzmodellierung: Methodische Konzepte zur Konstruktion und Anwendung wiederverwendungsorientierter Informationsmodelle. Logos, Berlin (2006)
19. Cardoso, J., Barros, A., May, N., Kylau, U.: Towards a Unified Service Description Language for the Internet of Services: Requirements and First Developments. In: IEEE International Conference on Services Computing, pp. 602-609. IEEE (2010)
20. Walkerdine, J., Hutchinson, J., Sawyer, P., Dobson, G., Onditi, V.: A Faceted Approach to Service Specification. In: 2nd International Conference on Internet and Web Applications and Services (ICIW 07). IEEE Computer Society (2007)
21. Weigand, H., Johannesson, P., Andersson, B., Berholtz, M.: Value-Based Service Modeling and Design: Toward a Unified View of Services. In: Advanced Information Systems Engineering. LNCS, Vol. 5565, pp. 410-424. Springer (2009)
22. Staud, J.L.: Geschäftsprozessanalyse: Ereignisgesteuerte Prozessketten und objektorientierte Geschäftsprozessmodellierung für Betriebswirtschaftliche Standardsoftware. Springer (2006)
23. Richter- von Hagen, C., Stucky, W.: Business-Process- und Workflow-Management: Prozessverbesserung durch Prozess-Management. Teubner, Stuttgart (2004)
24. Rosemann, M., Schwegmann, A., Delfmann, P.: Vorbereitung der Prozessmodellierung. In: Becker, J., (ed.): Prozessmanagement: ein Leitfaden zur prozessorientierten Organisationsgestaltung. Springer, Berlin (2003)
25. Alves, G., Roßmann, J., Wischniewski, R.: A discrete-event-simulation approach for logistic systems with real time resource routing and VR integration. In: International Conference on Computational Systems Engineering (ICCSE 2009), World Academy of Science, Engineering and Technology WASET, pp. 476-481 (2009)
26. Kuhn, A., Wenzel, S.: Simulation logistischer Systeme. In: Arnold, D., Furmans, K., Isermann, H., Kuhn, A., Tempelmeier, H., (eds.): Handbuch der Logistik. Springer, Berlin Heidelberg (2008)
27. Richtlinie, V.: 3633: Simulation von Logistik-, Materialfluß- und Produktionssystemen, Grundlagen, Beuth (2000)
28. Hoxha, J., Scheuermann, A., Bloehdorn, S.: An Approach to Formal and Semantic Representation of Logistics Services. In: ECAI 2010, pp. 73-78 (2010)

Evaluating the App-Store Model for Enterprise Application Software and Related Services

Stefan Wenzel^{1,*}, Francesco Novelli^{2,*}, and Christoph Burkard^{3,*}

¹ SAP AG, Commercial Platform, Walldorf, Germany
stefan.wenzel@sap.com

² SAP AG, SAP Research, Darmstadt, Germany
francesco.novelli@sap.com

³ TU Darmstadt, Software Business & Information Management, Darmstadt, Germany
burkard@is.tu-darmstadt.de

Abstract. Spurred by the overwhelming success of mobile “app stores”, enterprise software vendors are increasingly embracing the use of online sales channels. Whereas organizational buying behavior has already been investigated with regard to the acquisition of enterprise software via traditional offline channels, researchers have not yet focused on its provision using an electronic sales channel. In this explorative study, we laid the first bricks to fill this research gap and provided valuable insight to practitioners confronting this marketing innovation. Relying on a qualitative research strategy, we assessed the extent to which an electronic sales channel may support the different phases of a generic software buying process, and the impact of product characteristics on enterprise customers’ channel adoption. We identified a set of adoption drivers and barriers which ought to be taken into account by channel providers, and elicited technological and organizational solutions from the interviews.

Keywords: app store, enterprise application software, electronic marketplace, buying situation, online channel

1 Introduction

In consumer markets, e-commerce has undoubtedly grown a widely-accepted distribution channel. As a case in point, the online retailer Amazon has become the world’s leading retailer [1], building on a well-accepted e-commerce model for books and subsequently extending its strategy to turn into a general online store. Something comparable has been happening in the mobile software market with the Apple’s “App Store”, thanks to its intuitive buying experience and seamless consumption capabilities. In fact, the “app-store model” has changed the way consumer software is built, packaged, sold, and delivered.

* These authors contributed equally to this work.

Contrariwise, in the market for enterprise application software – which encompasses standalone business applications such as Customer Relationship Management (CRM) or integrated ones such as Enterprise Resource Planning (ERP) – a traditional sales model is still predominantly pursued: that of a long-lasting, personnel-intensive process. As a matter of fact, sales cycles of several months up to years are common ([2] and [3]). The buying process is often highly centralized and driven by IT staff or purchasing departments. It involves the evaluation of multiple solutions and generates high costs for both the purchasing company and the software vendor [4]. This is not only due to historical and organizational reasons but also determined by the nature of enterprise software itself.

However, the app-store model could be emulated by enterprise software vendors by establishing online marketplaces offering complementary software solutions and services provided by them and their ecosystems. The platform should serve enterprise software consumers throughout the buying process from information search to purchase and software delivery with minimum human interaction. The set of organizational and technological means constituting this centralized e-commerce platform is what we refer to as the “online sales channel”.

Initially, this strategy has been pursued by Software-as-a-Service (SaaS) vendors such as Salesforce.com (AppExchange) and NetSuite (SuiteApp.com) [5]. Other companies have followed a similar approach with electronic B2B marketplaces or app-stores for their diverse solutions (including both on-demand and on-premise applications, and related services): SAP (SAP Store), Deutsche Telekom (Business Marketplace), Microsoft (Pinpoint), Fujitsu (Business Solution Store), Amazon (AWS Marketplace), SugarCRM (SugarExchange), and Google (Google Apps Marketplace).

In this exploratory study we assessed to which extent the app-store model can be successfully transferred to the enterprise software market, and which barriers and drivers might influence the customers’ adoption of such a channel. Furthermore, we identified relevant e-commerce instruments and methodologies to overcome existing barriers. Finally, we derived implications for decision makers and managers of enterprise software companies contemplating the use of online sales.

2 Related Research and Methodology

Research on organizational buying of enterprise software is a multidisciplinary field of study, in particular in the context of an electronic distribution channel. One stream of research relevant to our work is that on organizational buying behavior in general. Foscht and Swoboda [6] compared the buying behaviors of consumers and organizations. Robinson et al. [7] proposed three dimensions to comprehensively identify buying situations: newness of the problem, information requirements, and consideration of new alternatives. Building on these findings, they introduced three buying classes corresponding to specific patterns of purchase behavior: new task, modified rebuy, and straight rebuy. Webster and Wind [8] proposed a buying process model and differentiated between four variables influencing the buying behavior: individual, social, organizational, and environmental. Based on those seminal publications, many au-

thors have then investigated factors of influence on the organizational buying process. Sheth [9] distinguished individual, environmental, and group-organizational aspects of the buying process. The influence of organizational actors was the focus of the analysis by Ronchetto et al. [10]. Johnston and Lewin [11] performed an extensive literature review on the topic and found that the most investigated constructs among the determinants of organizational buying behavior are the characteristics of the purchase, organization, group, participants, process, seller, and information. Recent publications analyzed the change of buying center structures between different situations and phases [12] .

A second relevant stream of research is the one focusing on organizational software acquisitions. Based on Webster and Wind, Verville and Halington [13] classified relevant factors in ERP acquisitions into environmental (e.g., technological, cultural), organizational (e.g., project management approach), group/interpersonal, and individual (e.g., acquisition team's composition, individual leadership). In a second study, Verville and Halington [14] modeled the ERP acquisition process through the phases of planning, information search, selection, evaluation, choice, and negotiation. The planning phase deals with team formation, and selection and evaluation criteria. Information search addresses information screening and information sources, while the selection process comprises the evaluation of requests for quotes, and the shortlisting of vendors. Functional, technical, and vendor-specific aspects are appraised in the evaluation phase. The choice phase encompasses the final product choice and precedes the negotiation phase. Benlian et al. [15] analysed the influence of application specificity, application adoption uncertainty, strategic value of application, application inimitability, the attitude toward SaaS in general, and subjective norms on the degree of software as a service adoption. Whereas Loebbecke [16] identified information-related drivers (i.e., customer references, expert network recommendations, and demonstration team presentations) and feature-related ones (i.e. price performance, functionality, sales team service) impacting the business to business (B2B) software purchase decision, Palanisamy et al. [17] identified five factors influencing the enterprise software acquisition process: enterprise-systems strategy and performance, business process re-engineering and adaptability, management commitment and users' buy-in, single vendor solution, and consultants, team-location, and vendor's financing.

To the best of our knowledge, the provision of enterprise software or related services using an online channel has not been investigated yet. Whereas Benlian et al. [15] and others who studied SaaS adoption assumed an electronic purchase of that SaaS solution, the acceptance of an online channel for enterprise software or related services has not been investigated in detail yet. One exception is the work of Wenzel et al. [18] who proposed a framework to distinguish enterprise software, which may be suitable for a transactional sales model (i.e., electronic sales channels) or a consultative sales model (i.e., traditional, offline sales channels). They grouped the dimensions in their framework into "what is being bought" (i.e., price of software, risk connected with the purchase, specificity of software, complexity of software), "who is buying" (i.e., number of users, buying authority, level of IT expertise), and "how it is

bought” (i.e., buying class, ability to evaluate, implementation type, software deployment).

3 Research Methodology

Our explorative research relied on a qualitative research design and semi-structured interviews. On the one hand, the semi-standardized interviews allowed to provide a shared understanding of the channel, the buying-process, and the products/services. On the other, each interviewee was free to enrich the discussion with elements of enthusiasm or concern peculiar to her profile and organizational environment.

Interview Design. A tentative set of questions was assembled drawing from the literature (see section 2). Test interviews (10 in total; up to 30 minutes each) were conducted at the ICT trade fair CeBIT targeting software customers (4 test-interviews) and software providers (6 test-interviews). A revised version of the interview guide was discussed with a sales member from a global enterprise software vendor (from now on, simply “ESV”). Thereafter, the guide was only subject to minor adjustments in wording.

The interview structure was articulated in four sections. The first one comprised an introduction to the topic of electronic channels, the explanation of our simplified product portfolio and buying process, and questions on the interviewee’s familiarity with online purchasing of software, both privately and professionally. The second part focused on the buying process and examined the possibility of relying only on an online channel to support it. The third one explored the factors influencing the channel decision. Eventually, the interviewees were given the possibility to raise additional points and further comments before concluding the interview.

Stylized Product Portfolio. Since the market for enterprise software encompasses a wide range of products and services, and since we expected at least some product characteristics to play a role in the adoption of an online channel, we devised a stylized product portfolio to be used as a common reference.

Core solutions are either company-wide information systems (e.g. ERP) or systems spanning one functional area (such as CRM). *On-top solutions* are components which complement core solutions by providing additional functionalities, business-process support, or front-ends (such as mobile applications). *Non-executable items* are post-purchase goods affecting the solution without modifying its code base (e.g., user licenses, usage contingences, and service level agreements). *Services* are distinguished into those related to the technological facet of the solution (*IT Services*, e.g., implementation, data migration, training) and those related to the supported business processes (*Professional Services*, e.g., accounting).

Buying Process. The buying decision process was derived from Verville and Halington [14] and slightly simplified to comprise 5 phases. In the *problem recognition* phase the organization gains awareness of an opportunity or threat which can be

dealt with by acquiring an enterprise solution. The *information search* covers the acquisition of information material, followed by the *evaluation*, where selected solutions and vendors are ranked. The *negotiation and purchase* phase encompasses finalizing the terms of the transaction, stipulating contracts, and executing the purchase. The *aftersales* phase covers additional purchases (i.e., non-executable items).

Each phase was treated similarly. The interviewee was asked to describe the generic enactment of the phase as-is in her organization. Hereupon, the possibility of performing the phase in the online channel was evaluated covering also the (dis-) advantages of the channel. Whenever the interviewee did not mention the existence (or absence) of product specificities, she was reminded of the portfolio structure and prompted to think of any difference which may arise in sourcing the items online.

Sample Design. We employed a non-probability sample-design, that is, the researchers' own networks of professional relationships were exploited to find suitable candidates, relying on a combination of convenience and snowball sampling. Twelve interviews were conducted overall, including 4 with ESV sales members, 7 with German companies (covering both, medium and large enterprises), and one with a public administration agency. The interviewees cover senior roles in IT as well as business departments. Our sample size lies within the range of what is commonly considered acceptable for a qualitative study [19].

Data Collection and Analysis. Each interview took from 50 to 70 minutes and was conducted between March and June 2012 either face-to-face (3) or telephonically (9). To support the discussion, visual representations of the buying process and product portfolio, were employed, either as hard copy or sent to the interviewee in advance of the appointment. The participating persons were preliminary made aware of the study's goals (including the aim of publishing a scientific contribution) and guaranteed of the anonymous treatment of all data. Conversations took place in German and were recorded and transcribed. Coding was performed by all authors (partly in pairs to ensure consistency) to analyze the transcripts.

4 Empirical Results

4.1 Buying Process

Table 3 and 4 gather respectively all the barriers and drivers which we could identify along the buying process and relate to a specific dimension of the purchase decision.

Problem Recognition. In the problem recognition phase, the online channel should act as an external stimulus to arouse the needs of the customer-companies by proactively recommending additional products and signaling adoption trends. While this role was positively seen by the majority of the interviewees across all product categories (Table 1), issues were raised as to whether the online platform would be capable of targeting the right recipients in a trustworthy manner (Table 3).

A crucial barrier in this phase is complexity, partly inherent in the marketed software solutions and partly due to the diversity of user-profiles targeted by the channel, making it a daunting task to deliver the appropriate message or recommendation to a specific user. Other vendor's electronic marketing initiatives may also be inconsistent or confusing for online channel's users. Generally, unfocused and/or uncoordinated communication in an organizational context could annoy members of the buying center, especially when they internally aim at conflicting organizational goals. Prior experience with the channel also represents a barrier, in that unprompted Internet messages are mainly associated with advertisement.

These issues notwithstanding, an online platform represents a unique opportunity for the vendor to extend its reach within the relevant buying centers (Table 4). It namely allows bypassing gatekeepers, that is, the buying team members (usually IT personnel) who filter the information flows and grant the vendor access to key employees. The platform could be accessed in all autonomy by any initiator, influencer or decider, who may in turn use the published information to persuade others. In case recommendations were integrated into the vendor's own software applications, the users base itself could be leveraged. Two interviewees further believed that some disseminated information be of value to customers by itself, for it allows benchmarking against competitors and industry trends.

Instruments mentioned during the interviews mainly dealt with overcoming the complexity threat, the difficulty in targeting appropriate recipients, and the distrust attached to Internet media. A number of technical solutions were suggested which could help reducing the complexity by easing the navigation to the wanted piece of information. Among those the use of dedicated stores for different buying centers, search capabilities for customer references, success-stories, and best-practices. Recommenders which intelligently utilize the user's contextual information and take into account similar users' decisions could more effectively couple recipients with the appropriate recommendation. The sales personnel's knowledge of the customer's organizational environment could also be employed to fine-tune or personalize the messages. Coherently with the marketing literature, messages may be legitimized by exploiting the customer's attention to third-parties' opinions, and, therefore, the channel's trustworthiness would be increased by using social networking components to involve experts and other users.

Information Search. In the information search phase the online channel acts as one-stop shop for all the information requirements potential customers may have. Most points raised with regard to problem recognition apply just as well to this phase. As a matter of fact, interviewees often tackled both phases simultaneously or developed their remarks further from one to the other. Being free of any contractual binding for the channel user and given the already widespread acceptance of the Internet as a means of information retrieval (in particular for on-demand solutions), there is an overall positive attitude towards the online platform as a source of information (Table 1).

Complexity and channel conflicts are the main barriers for the channel's adoption at this stage (Table 3). Both portfolio complexity and solution complexity may negatively affect the user's satisfaction with the channel. The latter complexity issue is

more acute when the solution is large in scope or demanding in terms of integration, or when multiple decision criteria (and thus information requirements) are considered at once. This may be worsened by the customer's lack of in-house knowledge of the solution and its underlying technologies.

Partly as means of addressing this complexity, and partly out of established habits in the enterprise software market, an online channel is seen as competing with third-parties and the vendor's own marketing and sales to win the customer's attention. While the autonomy and anonymity guaranteed by the online platform is valued by some, a direct contact point, such as a sales representative or a consultant, may still be considered a more convenient, comprehensive and reliable source, especially as the solution's criticality increases.

All instruments listed for the problem recognition phase apply to the information search as well, although the piece of information to be effectively retrieved, communicated, or legitimized revolves around the solution to the problem. For that reason, potential consumers require hands-on experience with the solution in the forms of demos and trials. Moreover, new media could be employed to lessen or circumscribe the need for consultancy (e.g., video-conferencing and online seminars).

Evaluation of Alternatives. In the evaluation phase, the channel should allow an in-depth analysis and comparison of selected solutions. Whereas the general tendency was positive, responses varied considerably depending on the type of product under investigation. Most interviewees dismissed online evaluation of core solutions and services, but all contemplated it for on-top solutions (Table 1).

These different attitudes towards different product types are determined by multiple factors. Overall, the interviewees mentioned 16 of them, influencing the channel choice positively (drivers, Table 4) or negatively (barriers, Table 3) towards the online evaluation. The major barriers appear to be solution complexity, scope and evaluability, mentioned by the lion's share of interviewees. Functionally complex products with breadth and depth of features, such as ERP, are rather considered suitable for traditional enterprise software channels where direct contacts with the vendor's sales force can take place. Furthermore, the evaluation of solutions may be hindered by the absence of customer data to produce meaningful test-scenarios. Whereas the latter may not exclusively be a barrier for online channels, the provision of interactive tools for easier online evaluation is boosted by online channels. Moreover, functionally limited solutions may be a driver of online channels. Thus, the limited solution scope or the limited integration are mentioned as driver in conjunction with the prior experience with the solution, the evaluability of solution in general and the integration/implementation effort. Concerning IT and professional services, the interviewees responded rather unfavorably to the online channel as means of evaluation given the intangible nature and the relevant human component of services.

The interviewees stated instruments and methods to overcome barriers mentioned above. Many proposed instruments address the evaluability (e.g., demos, trials, expert contact). Furthermore, the complexity / the scope of the solution may be reduced, for example, by know-how databases, guided tours or a wizard (i.e., answering questions on customers' needs and getting product recommendations). One Interviewee remarked that the online channel should address different information requirements:

“The executive board has another information demand what the solution does and how this is explained. They won’t go into detail. ... Of course, a member of the financial accounting ... has to know whether dunning processes or credit vouchers can be done. That means, there is a layered information concept for different information requirements”.

Negotiation and Purchase. In the negotiation and purchase phase, the online channel should support the settlement of contracts, billing and payment, and the delivery of the solution. When analyzing whether the interview participants could imagine conducting this phase via an electronic channel, the results are rather heterogeneous (Table 1). Whereas the general tendency is almost balanced between pro and contra positions, most interviewees would not conduct this phase online for core solutions. However, some of them did distinguish between integrated solutions (like ERP) and more self-contained solutions (e.g., travel expense management) – the latter category perceived as better suited for an online channel. Contrariwise, on-top solutions were perceived as highly suitable for online purchases. Online-suitability was also attributed to standardized IT services (like training or maintenance services). IT and professional services with high personal components, however, are preferred to be purchased traditionally.

The identified influence dimensions for negotiation and purchase are broadly spread. Similar to the information and evaluation phases, multiple factors relating to the nature of the purchased solution were mentioned, e.g. solution complexity or integration and implementation effort. However, in contrast to these previous phases, a greater focus was given to contracts and pricing models as adoption barriers (Table 3). Complexity in contracts and pricing models, as well as the need for individual contract conditions to comply with customer’s requirements, is seen as a major barrier for purchasing enterprise software and services online. Further barriers mentioned by the interviewees included complexity in the solution deployment, and incompatibility with customers’ internal purchasing policies and processes. Lack of integration with internal purchasing systems would lead to undesirable replication duties to maintain synchronization between the records.

Furthermore, the recognition of customers’ individual contract conditions and existing frame contracts (covering online purchasing conditions) were mentioned as enablers for the online channel. The fact that online stores foster transparency in pricing and discount mechanisms is perceived as intrinsic advantageous to the online channel.

To counter the mentioned barriers, the interviewees proposed multiple instruments to address contract issues (e.g., frame agreements and standardization of contracts) and also pricing ones (e.g., online acknowledgment of individual discounts). Another group of proposed improvements dealt with customers’ internal purchasing policies, processes and systems (e.g., support of budget allocation and approval, license management, integration with the customer’s purchase system) and to handle the complexity of deployment processes (e.g., automatic deployment of the solution). Finally, tools to enable personal consultation with experts (such as chat, phone support, video conferencing) and certifications to build the solution provider’s online reputation.

Aftersales. In the aftersales phase, the electronic channel should allow purchasing additional non-executable goods related to already acquired solutions, and managing them together with other solution aspects such as service level agreements. Most interviewees agreed on the viability of conducting this phase through an electronic channel (Table 1) and coherently focused on adoption drivers when discussing it (Table 4). In particular, the possibility to perform license management online may be a driver for the online channel. Interviewees proposed online contract management (e.g., online availability of existing contracts) as a helpful instrument.

Table 1. Interviewees' attitude towards the online channel's support of the buying process

Buying process phase	Product class	Online channel adoption decision			
		Positive	Neutral	Negative	N/A
Problem recognition	Overall assessment	9	1	1	1
Information search	Core Solutions	6	2	3	1
	On Top Solutions	11	0	0	1
	IT Services	10	1	0	1
	Prof. Services	6	0	3	3
	Overall assessment	8	2	0	2
Evaluation of alternatives	Core Solutions	1	3	6	2
	On Top Solutions	9	0	0	3
	IT Services	2	3	3	4
	Prof. Services	0	1	5	6
	Overall assessment	4	2	1	5
Negotiation and purchase	Core Solutions	0	2	6	4
	On Top Solutions	5	0	1	6
	IT Services	2	2	0	8
	Prof. Services	1	1	2	8
	Overall assessment	4	2	3	3
Aftersales	Overall assessment	6	1	0	5

Table 2. Dimensions influencing customers' adoption of the online channel

Dimension	Relevant	Irrelevant	Unprompted mention
Solution complexity	12	0	11
Solution criticality	10	1	7
Solution evaluability	9	1	9
Solution scope	10	0	8
Solution pricing	10	2	7
Solution specificity	9	1	7
Integration & implementation effort	11	1	7
Solution deployment	8	3	5
Number of end-users	6	5	7
Provider's reputation & trustworthiness	5	0	7
Existing relationships	8	0	9
Contracts in place with the provider	7	0	8
Prior experience with solution	7	0	8
Prior experience with channel	5	1	5
IT competences	10	1	5
IT governance & procurement	11	1	8
Innovativeness	3	0	4

Table 3. Barriers to the adoption of the electronic channel (*continued*)

Buying process phase				Dimension	
Problem recognition	Information Search	Evaluation of alternatives	Negotiation and Purchase	Aftersales	Relationship
					Customer
	A must, but building them requires personal contacts and thus a traditional sales channel				Provider's reputation & trustworthiness
Personal relationships a prerequisite to trust recommendations; only an account manager knows which is the right target for a certain message	The lack of existing business relationship with the provider increases perceived risk; in presence of a dedicated account representatives, the customer may find it more convenient to inquire or act through that channel				Existing relationships
	Contracts contain complex elements to evaluate (e.g., pricing models, data hosting for on-demand solutions)	Nonstandardized contracts with individual SLA and price arrangements are demanded by customers	Customers want to avoid lock-in of long-term contracts and subscriptions		Contracts
	Customers with limited or absent in-house expertise with regard to the solution will rather rely on third parties (experts, consultants and partners)				Prior experience with the solution
Automatic online recommendations perceived just as ads; confusion with corporate marketing web pages; erroneously targeted messages annoy users; lack of awareness of the online platform	Confusion with the provider's other electronic marketing initiatives	Customers not yet used to advanced online instruments such as wizards or system configurators			Prior experience with the channel
	Customers with limited or absent in-house IT expertise rather rely on third parties' consultancy	Low IT affinity negatively impacts evaluability	Lack of IT competences increases perceived risk		IT competences
Uncoordinated messages to multiple persons with conflicting stakes cause irritation	Customers advised by consultants or partners who fulfill their information requirements	Opposed by decision makers or forbidden by existing procurement processes	A plethora of people, hierarchies, and customer-specific policies and processes to be considered	The online platform do not reflect the customer's own license management	IT governance & procurement
A conservative attitude (generally associated with European customers) will delay adoption					Innovativeness

Table 4. Drivers for the adoption of the electronic channel

Buying process phase				Dimension		
Problem recognition	Information Search	Evaluation of alternatives	Negotiation and Purchase	Aftersales	Solution	Customer
	Simple, standardized, functionally-focused and self-contained commodity applications will be a driver for the electronic channel				Complexity, criticality, scope, specificity, Implementation & integration effort	
	Availability of free-of-charge trials	Transparency of pricing models and discounts			Pricing	
	On-demand solutions more readily evaluated online	Automatic deployment			Deployment	
	A strong brand will drive adoption	Low sales volumes			Number of end-users	
	Possibility for the vendor to bypass gatekeepers and reach out to relevant customer's employees; possibility for the customer to avoid personal contacts and act in anonymity				Provider's reputation & trustworthiness	
	Freedom from any contractual binding at this stage	Short-term contracts	Frame agreements and recognition of customer-specific contractual conditions		Relationships with the provider	
Integrated with the vendor's own applications to leverage user base	Customers with solution expertise will be keen to inform themselves online	Positive prior experience with smaller or similar solutions encourage deploying larger solutions from the channel			Contracts in place	
	Internet already used to collect information	The more activities and instruments are moved online, the more customers will get used to perform them there			Prior experience with the solution	
	Customers with IT expertise keen to inform themselves online				Prior experience with the channel	
Buying center members use online information to persuade other decision makers	Online channel as an instrument to enforce governance				IT in-house competences	
A pioneering attitude (generally associated with American customers) will drive adoption					IT governance & procurement	
					Innovativeness	

4.2 Dimensions Influencing the Adoption of Online Channels

One goal of the survey was to identify the relevant dimensions influencing – as either barriers or drivers – online channel adoption. Table 2 lists all the explicitly mentioned dimensions. This qualitatively substantiates the framework proposed by Wenzel et al. [18] (see the reference for the dimensions' detailed descriptions). The column “relevant” reports whether the interviewees judge the dimension relevant for the electronic channel adoption, the column “irrelevant” shows the number of interviewees who *explicitly* rated the factor as irrelevant. Both columns take the whole interview into account. The last column indicates whether the interviewees mentioned the specific factor without being prompted, that is, without an explicit question by the interviewer. Thus, this column refers to the second part of the interview only, since the third part of the interview purposely targeted judgments on the *given* dimensions.

5 Management Implications

On the basis of our exploratory findings, enterprise software customers will be more or less inclined to use an electronic channel largely depending on the solution's specific attributes and on the buying process phase they are in. Enterprise software vendors aiming at the establishment of a B2B app-store ought to take that into account and consequently shape their strategy.

First of all, the online channel's role should fit the vendor's portfolio and the served market segments [18]. If buying situations predominantly revolve around transactional sales, then the entire buying process from problem recognition to aftersales may be conducted online, and the electronic channel could very well become the primary sales channel. When certain customer segments (e.g., small enterprises) would adopt the online channel and others (e.g., large enterprises) refrain from it, the electronic platform could be a complementary sales channel in alternative to the personal sales channels. A third possibility encompasses a tight integration of offline and online channels via pre-defined “entry and exit points” or “channel hopping” [20] to let both channels cross-fertilize and reap the advantages of both ([21] and [22]). As an illustrative scenario, the buying process would be supported by the electronic platform until the customer requests a formal quotation. Subsequent phases would then be conducted together with a personal sales agent.

Secondly, our analysis has shown that some existing products and services are tailored for an offline buying mode and could not be offered “as-is” through an electronic marketplace. Therefore, it may be necessary to re-design them or adjust some of their characteristics to better suit the online channel. Along with those mentioned in section 3, the following changes might simplify the buying situation and reduce barriers on multiple levels for a selected solution: disaggregation of the solution in terms of scope or usage (initially free, then paid); one-dimensional, time or usage-based pricing models; and instant-deployment capabilities.

Finally, e-commerce capabilities of today's consumer-oriented app-stores must be enhanced to meet the enterprise customers' unique requirements. Electronic marketplaces and online stores for enterprise application software and services namely need

additional capabilities and new technologies to cope with the complexity of the enterprise market and reduce the adoption barriers we identified. The instruments mentioned throughout section 3 are possible improvements we elicited with our qualitative study. Among the capabilities not currently available in consumer app-stores: acknowledgment of corporate buying roles and governance structures (e.g., approval workflows and group decision support) and support/adaptation to those; automatic compatibility and integration checks with the customer's existing landscape; acknowledgment of individual and volume discounts; online involvement of the vendor's sales and service personnel via collaboration technologies (e.g., co-browsing, chat, video conferencing).

6 Outlook

In this exploratory study, we investigated the suitability of an online channel for enterprise software and services and offered first insights into the electronic channel's aptness to support the buying process with regard to the different types of enterprise software products and services. We elicited barriers and drivers of the electronic channel adoption. Whereas our study allows to draw preliminary conclusions concerning the acquisition of enterprise software via electronic channels, some limitations do apply. First of all, the sample size did not allow to thoroughly take into account different company characteristics. Subsequent research endeavors may increase the sample size to investigate this aspect as well. Moreover, the influencing dimensions we investigated may be overlapping. While some interviewees expressed thoughts on a general level (e.g., in terms of complexity), others went into details (e.g., evaluability, which may affect the perceived complexity as well). Following studies may build on our results to investigate dimensions' correlations and interdependences as well, something which was not contemplated here. Besides, further research could employ a quantitative research approach. Repeated studies could be especially useful to track the acceptance of an enterprise software online channel over time.

Acknowledgements. This work was partially financed by the European Commission under grant agreement 285248 (project FI-WARE).

References

1. comScore: Amazon Sites Visited by 1 in 5 Global Internet Users in June, http://www.comscore.com/Insights/Press_Releases/2011/8/Amazon_Sites_Visited_by_1_in_5_Global_Internet_Users_in_June [accessed 19 December 2012]
2. Liao, X., Li, Y., Lu, B.: A model for selecting an ERP system based on linguistic information processing. *Information Systems* 32, 1005–1017 (2007)
3. Anwenderstudie, E., Mediengruppe, D.K.: Einsatz von ERP-Lösungen in der Industrie. (2009)

4. Cusumano, M.A.: The business of software: What every manager, programmer, and entrepreneur must know to thrive and survive in good times and bad. Free press (2004)
5. Burkard, C., Draisbach, T.: Software ecosystems: Vendor-sided characteristics of online marketplaces. In: Informatik 2011. LNI, Vol. P-192. GI, Bonn (2011)
6. Foscht, T., Swoboda, B.: Käuferverhalten: Grundlagen-Perspektiven-Anwendungen. Gabler (2007)
7. Robinson, P.J., Faris, C.W., Wind, Y.: Industrial buying and creative marketing. Allyn & Bacon (1967)
8. Webster, F.E., Wind, Y.: Organizational buying behavior. Prentice-Hall, NJ (1972)
9. Sheth, J.N.: A model of industrial buyer behavior. *The Journal of Marketing* 37 (4), 50–56 (1973)
10. Ronchetto Jr, J.R., Hutt, M.D., Reingen, P.H.: Embedded influence patterns in organizational buying systems. *The Journal of Marketing* 53 (4), 51–62 (1989)
11. Johnston, W.J., Lewin, J.E.: Organizational buying behavior: Toward an integrative framework. *Journal of Business research* 35, 1–15 (1996)
12. Järvi, P., Munnukka, J.: The dynamics and characteristics of buying centre networks: A qualitative study on Finnish firms. *Marketing Intelligence & Planning* 27, 439–457 (2009)
13. Verville, J.C., Halington, A.: A qualitative study of the influencing factors on the decision process for acquiring ERP software. *Qualitative Market Research: An International Journal* 5, 188–198 (2002)
14. Verville, J., Halington, A.: A six-stage model of the buying process for ERP software. *Industrial Marketing Management* 32, 585–594 (2003)
15. Benlian, A., Hess, T., Buxmann, P.: Drivers of SaaS-adoption—an empirical study of different application types. *Business & Information Systems Engineering* 1 (5), 357–369 (2009)
16. Loebbecke, C., Weiss, T.: Drivers of B2B Software Purchase Decisions. In: *Proceedings of the conference on Bridging the socio-technical Gap in Decision Support Systems: Challenges for the Next Decade*, pp.535–546. IOS Press (2010)
17. Palanisamy, R., Verville, J., Bernadas, C., Taskin, N.: An empirical study on the influences on the acquisition of enterprise software decisions: A practitioner's perspective. *Journal of Enterprise Information Management* 23, 610–639 (2010)
18. Wenzel, S., Faisst, W., Burkard, C., Buxmann, P.: New Sales and Buying Models in the Internet: App Store Model for Enterprise Application Software. In: *MKWI 2012 Proceedings*, pp. 639–651 (2012)
19. Luborsky, M.R., Rubinstein, R.L.: Sampling in Qualitative Research: Rationale, Issues, and Methods. *Research on aging* 17, 89–113 (1995)
20. Ahlert, D., Blut, M., Michaelis, M.: Erfolgsfaktoren des Multi-Channel-Marketings. In: *Handbuch Multi-Channel-Marketing*. (2007)
21. Steinfield, C.: Understanding click and mortar e-commerce approaches: A conceptual framework and research agenda. *Journal of Interactive Advertising* 2, 1–10 (2002)
22. Armbruster-Reif, K.: Multi-Channel Commerce: Hybridstrategien und Controlling. In: *6. Internationale Tagung Wirtschaftsinformatik 2003-Band 1*. 553 (2003)

Research Methods and Philosophy

Ulrich Frank¹ und Frédéric Thiesse²

¹ University Duisburg-Essen, Essen, Germany

² University Würzburg, Würzburg, Germany

Due to its strong interrelations with disciplines as diverse as management, computer science, psychology, and others, Information Systems research today is characterized by manifold paradigmatic views and methodological approaches. The IS field has a long history of discussing its research methods and different schools of thought compete for attention and legitimacy. Prominent examples include the positivist and the constructivist paradigm, each based on their own epistemological assumptions and accompanied by sets of criteria for conducting methodologically sound research. In recent years, it was particularly the design research paradigm which has attracted the interest of IS researchers worldwide. Whether it is acknowledged or not, the philosophical premises underlying any form of scientific investigation come with fundamental consequences for the practice of IS research and its results. Not least, the interplay between philosophy of science, the choice and employment of different methods, and the very nature of our research results is reflected by the perpetual debate on the trade-off between rigor and relevance in IS. Against this backdrop, the objective of the present track is to consider philosophy and methodology in the IS discipline and discuss their importance for the future of the field as a whole.

Program Committee

Roman Beck, University Frankfurt
Peter Fettke, University Saarbrücken
Stefan Klein, University Münster
Franz Lehner, University Passau
Björn Niehaves, University Münster
René Riedl, University Linz
Kai Riemer, University Sydney
Bernd Carsten Stahl, De Montfort University
Stefan Strecker, FernUniversität Hagen
Volker Wulf, University Siegen

Theorieentwicklung in der Akzeptanzforschung: Entwicklung eines Modells auf Basis einer qualitativen Studie

Kristin Vogelsang, Melanie Steinhüser, und Uwe Hoppe

Universität Osnabrück, Institut für Organisation und Wirtschaftsinformatik, Osnabrück,
Germany

{kristin.vogelsang,melanie.steinhueser,uwe.hoppe}@uni-osnabrueck.de

Abstract. Die Untersuchung der Akzeptanz von Technologien im Allgemeinen und Software im Besonderen ist ein fruchtbares Feld in der anglo-amerikanischen Forschung der Disziplinen (Management) Information Systems und der deutschen Wirtschaftsinformatik. Trotz zahlreicher Untersuchungen, die ihren Ursprung in dem Technologie-Akzeptanzmodell und verwandten Theorien haben, mehren sich Beiträge, welche die Defizite bisherigerer Studien und Forschungsansätze herausstellen. Eine wesentliche Ursache ist Fokussierung auf quantitative Forschungsmethoden, wie wir anhand von Metastudien und einer eigenen Literaturrecherche aufzeigen. Während quantitative Verfahren sich in der Regel gut für eine Überprüfung gegebener Theorien eignen, ist ihr Beitrag für die Bildung neuer Theorien begrenzt. Im vorliegenden Beitrag wird aufgezeigt, wie ein qualitatives Verfahren zur besseren Theoriebildung genutzt werden kann. Am Beispiel der Untersuchung der Akzeptanz von Projektmanagement-Software (PMS) kann aufgezeigt werden, dass dieses Vorgehen zu neuen Konstrukten führt, während einige Konstrukte bestehender Akzeptanz-Theorien nicht bestätigt werden konnten.

Keywords: Akzeptanz, qualitative Forschung, quantitative Forschung, Forschungsansatz, TAM

1 Einleitung

Bereits seit den 80er Jahren ist die Erforschung der Akzeptanz von Anwendungssystemen ein Kerngebiet der Wirtschaftsinformatik. Die bis heute immer wieder erweiterten und veränderten Modelle spiegeln den Wandel der Bedeutung und Fähigkeiten von Informationstechnologien (IT) wider [1]. Sie haben ihre Wurzeln in der anglo-amerikanischen IS- und MIS-Forschung. Das Technology Acceptance Modell (TAM) von Davis [2] gilt als Initialmodell für zahlreiche Folge Modelle des Forschungsgebiets. Es enthält den Kern der Akzeptanzforschung, die Erfassung und Deutung der Perceived Ease of Use (PEU), Perceived Usefulness (PU) und Intention to Use (IU) als Einflussgrößen auf die Akzeptanz. Noch immer entstehen immer wieder Varianten des TAM. Die Untersuchung der Akzeptanz verschiedener Softwareprodukte erfolgt

überwiegend anhand quantitativer Verfahren. Diese eignen sich jedoch nur bedingt für eine praxisnahe Umsetzung der gewonnenen Kenntnisse. Zudem ist ein Hauptproblem bei der Anwendung der auf dem TAM basierenden Modelle, dass die Konstrukte oft durch Kombination verschiedener vorhandener Theorien gebildet werden und somit der Prozess der Theoriebildung unvollständig bleibt.

In diesem Artikel soll aufgezeigt werden, welche Vorteile ein qualitatives Vorgehen, im Sinne einer Expertenbefragung bei der Analyse der Softwareakzeptanz aufweist, und wie der Theoriebildungsprozess verbessert werden kann. Zu diesem Zweck haben wir das Expertenwissen verschiedener Gruppen wie Anwender, Hersteller und erfahrener Berater zusammengetragen und die formulierten Erfahrungen und Best Practices zusammengeführt. Zudem wird ein Handlungsrahmen vorgestellt, der die Lücken der quantitativen Forschung im Bereich der Softwareakzeptanz ausgleicht. Zusätzlich werden Ergebnisse des Forschungsprozesses dargestellt.

2 Softwareakzeptanzforschung

2.1 Überblick

Das TAM gilt als das bekannteste Modell zur Erforschung der Softwareakzeptanz [3]. Es wurde 1986 von Fred D. Davis [2] entwickelt und gehört seitdem zu den am meisten zitierten und variierten Akzeptanztheorien [4]. Zentrales Element im TAM ist die Einstellung eines Anwenders gegenüber der Nutzung einer Software. Sie beeinflusst die tatsächliche Nutzung, mit der die Akzeptanz der Software gemessen wird. Das TAM geht auf die Theorie of Reasoned Action (TRA) und die Theorie of Planned Behavior (TPB) zurück. Daher liegt die Behavioral Intention – die Nutzungsabsicht – im Erkundungsfokus. Das TAM in seiner Grundform kommt mit vier Konstrukten aus. Die Einstellung wird durch Perceived Usefulness (PU) und der wahrgenommenen Einfachheit des Systems, der Perceived Ease of Use (PEU), dargestellt. Diese Konstrukte beeinflussen die Nutzungsabsicht (Intention to Use) welche im Modell der Nutzung (Use) vorausgeht.

Im weiteren Zeitverlauf wurde das TAM in zahlreichen Modifikationen erweitert oder auf verschiedene, meist generische Softwareprodukte, angewandt. Es lassen sich folgende Modifikationsströme identifizieren [3]:

1. Erweiterung um Faktoren anderer Modelle wie Subjektive Norm, wahrgenommene Verhaltenskontrolle und Selbstwirksamkeit (Self Efficacy). Sowie Erweiterungen um alternative Faktoren, entnommen aus der Theorie zur Diffusion von Innovationen, z. B. Erprobbarkeit, Kompatibilität, Sichtbarkeit oder Ergebnisbeweisbarkeit.
2. Untersuchung externer Variablen, wie demographische Faktoren und Persönlichkeitsmerkmale, welche die Beziehungen zwischen der Einfachheit der Nutzung und der wahrgenommenen Nützlichkeit und ihren Explikationen moderieren.
3. Studien, welche die Relevanz des TAM für bestimmte situative Anwendungsbereiche (bspw. für eine bestimmte Art von Software) untersuchen.

Zu den Studien der ersten und zweiten Kategorie gehören vor allem Untersuchungen aus dem anglo-amerikanischen Raum. Venkatesh und verschiedene Mitautoren haben sich hier in der Vergangenheit federführend gezeigt. Bemerkenswert ist beispielsweise das TAM3 [5], das 11 Explikationen der PEU und PU enthält sowie zahlreiche moderierende Beziehungen der Konstrukte zu externen Faktoren wie Experience und Voluntariness. Aktuell befasst sich die Akzeptanzforschung primär mit der Relevanz des TAM im Bezug auf neuere Technologien wie sozialen Medien [6-8] sowie Internettechnologien [9] und in diesem Zusammenhang mit Vertrauen [10].

Eine weiter gefasste Perspektive vertritt die Aneignungsforschung, die den gesamten Lebenszyklus der Nutzung [11] unter Berücksichtigung der Nutzungspraktiken [12] untersucht. Ziel dieses Ansatzes ist es, neben den Prädiktoren der Nutzung, wie sie in TAM Studien untersucht werden, auch die Interaktionen zwischen Mensch und Technologie zu analysieren [13] und dabei Veränderungen im Laufe des Aneignungsprozesses sowie unvorhergesehene Praktiken [14] zu berücksichtigen. Dabei wird ein völlig anderes Paradigma zugrunde gelegt, welches anstrebt, die Dualität zwischen Mensch und Technologie aufzulösen. Dieser Ansatz ist anders als typische TAM Studien von einer größeren Methodenvielfalt geprägt (z. B. [15-17]). Er wird an dieser Stelle allerdings nicht ausführlicher diskutiert, da der Fokus unserer Arbeit speziell auf der Akzeptanz von Softwareprodukten liegt und nicht auf einer weitergehenden Betrachtung des Aneignungsprozesses.

2.2 Defizite bisheriger Softwareakzeptanzforschung

Das TAM und seine Modifikationen dominieren die Akzeptanzforschung und auch die Methodik, mit der die Akzeptanz erhoben wird. Quantitative Ansätze, basierend auf dem TAM, sind vorherrschend in diesem Forschungsbereich [18]. Für Forscher birgt die starke Orientierung am TAM einige Vorteile. Es ist ein erprobtes und verlässliches Modell, das aufgrund seiner langen Tradition als anerkanntes Modell sowohl ein definiertes Vorgehen als auch gebilligte Interpretationen der Ergebnisse umfasst. Zudem ist unter Verwendung des Modells nahezu immer ein interpretierbares Ergebnis zu erwarten. Etwa 50% der Varianz von Intention to Use / Behavioral Intention konnten bislang maximal mit Hilfe des TAM aufgedeckt werden [3], [19-20]. Zudem ist das TAM, aufgrund seines definierten Erhebungsinstrumentes, einem Fragebogen, leicht zu handhaben und vielseitig einsetzbar. Die Einfachheit des TAM ist ein wesentlicher Treiber der starken Verbreitung des Modells [4] und trägt maßgeblich zu seiner Anerkennung bei. Aufgrund seiner allgemeinen Formulierung wird es in Untersuchungen für jegliche Art von Software eingesetzt. Diese Generalisierbarkeit des Erhebungsinstrumentes sorgt zudem dafür, dass Ergebnisse verschiedener TAM-Studien sich miteinander vergleichen lassen. Diese Merkmale haben das TAM zu dem meist zitierten Akzeptanzmodell gemacht.

An genau diesen Merkmalen setzt jedoch die Kritik an der Aussagefähigkeit des TAM an. Das TAM, in seiner Grundform beschränkt auf PEU und PU, steht im Verdacht, sich nur bedingt dazu zu eignen, komplexe Entscheidungsprozesse erklären zu können und das tatsächliche Verhalten in der Praxis zu prognostizieren [4]. Zudem wird die Fokussierung der Nutzungsabsicht als einziger Prädiktor der Nutzung oft-

mals als kritisch betrachtet [20-21]. Die erklärte Varianz der Zielvariablen Use liegt in den meisten Modellen unter 40% (TAM3 weist beispielsweise ein R^2 von 31-36% auf [20]) und weicht damit von den Ergebnissen der erklärten Varianz der Intention to Use ab. Dies ist ein Indikator dafür, dass die Schlussfolgerung, dass der Absicht etwas zu tun auch eine entsprechende Handlung folgt, nur bedingt richtig ist. Anzumerken ist auch, dass es sich sowohl bei Intention to Use als auch bei Use selbst um umstrittene Hilfskonstrukte der Akzeptanz handelt (vgl. [22] für Use und allgemein [20]). Auf der anderen Seite werden die Modelle durch immer mehr Explikationen von PEU und PU zu komplexen Theorien (vgl. bspw. [9] und [23], die v. a. aufgrund ihrer Annahmen bezüglich moderierender Effekte nicht mehr zu handhaben sind. Die tatsächliche Kausalität warum beispielsweise demographische Faktoren die Beziehungen der Konstrukte moderieren, ist bisher noch nicht hinreichend geklärt. Oftmals fehlen theoriegeleitete Begründungen der moderierenden Modellannahmen und Konstrukte [4].

Im Diskurs der Akzeptanzforscher unter dem Titel „Quo Vadis TAM“ [24-26] wird zudem die Relevanz einer Forschung, welche die Arbeitsumgebung berücksichtigt, herausgestellt. Dabei erkennen die Autoren die wichtige Rolle der Art der Software und auch des Einflusses des Unternehmens und fordern zu weiteren Untersuchungen dieser Kategorien auf [27]. Auch wichtige Schlüsselvariablen, wie beispielsweise die besondere Unternehmenskultur, die in einem Unternehmen vorherrscht, werden in den Untersuchungen nicht berücksichtigt oder nur verkürzt (beispielsweise in Form von Facilitating Conditions, die primär die umgebende Infrastruktur widerspiegeln (vgl. [5]) dargestellt.

Dies liegt vor allem an der Art und dem Fokus der Untersuchungen. Viele Autoren (vgl. beispielhaft [8]) untersuchen vor allem die Akzeptanz gegenüber generischen Softwareprodukten wie Textverarbeitung oder Internet, die über eine weite Verbreitung inhaltsgleicher Installationen verfügen. Jedoch treffen die Eigenschaften dieser Produkte wie einfache Anwendung, homogene Anwendungsziele (wie das Verfassen eines Textes), Nutzung unabhängig von der Qualifikation und Rolle der Nutzer oder auch die allenfalls lose Einbindung der Anwendung an die Unternehmensprozesse nur bedingt auf Informationssysteme zu, die spezifische betriebliche Aufgaben unterstützen. Bestehende Akzeptanzmodelle sind daher nur begrenzt dazu geeignet, die Akzeptanz ganzheitlich zu erklären. Erschwerend kommt hinzu, dass oft nur die Anwenderperspektive betrachtet wird.

Ein weiteres Defizit der Forschung zur Softwareakzeptanz muss darin gesehen werden, dass viele Autoren sich darauf beschränken, Akzeptanz zu erklären. Ein gestaltendes Wissenschaftsziel, im Sinne einer planvollen Erhöhung der Akzeptanz über eine Ableitung von Handlungsempfehlungen, steht nicht in ihrem Fokus. Darüber hinaus greift der Betrachtungszeitraum der Akzeptanzforschung recht kurz. Auch in Zeitreihenstudien werden oft nur der Initialkontakt der Anwender mit der betrachteten Software sowie eine begrenzte Zeitspanne an Folgemonaten betrachtet (vgl. bspw. [28]). Dabei sollten alle Phasen der Einführung (vgl. [29]), beginnend bei der Vorbereitung und Auswahl der Software bis hin zu ihrem flächendeckenden dauerhaften Einsatz bei der Betrachtung mit einbezogen werden. Es ist zu erwarten, dass der Einfluss identifizierter Akzeptanzfaktoren im Zeitverlauf der Softwareverwendung variiert. Dies wird in bisherigen quantitativen Studien nicht hinreichend berücksichtigt.

Im Folgenden zeigen wir auf, wie qualitative Forschung zur Behebung dieser Defizite beitragen kann.

2.3 Bedeutung qualitativer Methoden in der Technologieakzeptanzforschung

Seit Ende der 90er Jahre werden Angehörige der wissenschaftlichen Gemeinschaft der Bereiche Wirtschaftsinformatik / (Management) Information Systems ermutigt, qualitative Forschung anzuwenden [30]. Während quantitative Methoden zwar weiterhin dominant sind, ist eine Tendenz zur steigenden Akzeptanz qualitativer Forschung erkennbar. Dies spiegelt sich vor allem in einer zunehmenden Zahl von Veröffentlichungen wider [27]. Dabei ist der Anteil der Arbeiten, die Ergebnisse qualitativer Forschung präsentieren, im europäischen Raum traditionell größer als im amerikanischen [31]. Mögliche Gründe für die Dominanz der Veröffentlichungen quantitativer Forschung in renommierten Wirtschaftsinformatikpublikationen werden viel diskutiert und sollen an dieser Stelle nicht weiter erörtert werden. Zu betonen ist jedoch, dass die relativ geringe Präsenz bedauerlich ist, da sich qualitative Methoden prinzipiell gut eignen, um tiefe Einblicke in organisationale Kontexte zu erhalten [27], wie es bei Fragestellungen der Wirtschaftsinformatik oft gefordert ist.

Wie im vorangegangenen Abschnitt bereits erwähnt, ist die Dominanz quantitativer Methoden in der Akzeptanzforschung besonders deutlich. So stellen Lee et al. [3] in einer Meta-Analyse fest, dass über 90% der von ihnen betrachteten Akzeptanzstudien mit quantitativen Verfahren durchgeführt wurde. Andere Meta-Analysen betrachten die Forschungsmethode erst gar nicht und berücksichtigen lediglich statistische Ergebnisse der Artikel, die in die neue Studie einfließen (vgl. [1], [19], [32-36]).

Die Arbeit von Lee et al. stammt aus dem Jahr 2003. Um zu evaluieren, ob Wissenschaftler seit dieser Zeit methodisch andere Vorgehensweisen gewählt haben oder gar ein neuer Trend erkennbar ist, haben wir in führenden Wirtschaftsinformatikzeitschriften¹ nach Arbeiten zu diesem Forschungsgebiet gesucht, die nach 2003 publiziert wurden. Wir konnten 32 Artikel identifizieren, die sich mit Technologieakzeptanz befassen. Diese basieren ausschließlich auf theoretischen Überlegungen zur Struktur der Modelle sowie auf Daten, die aus quantitativen Erhebungen gewonnen wurden. Um zu prüfen, ob und ggf. inwieweit es Unterschiede hinsichtlich der angewandten Forschungsmethode im deutschsprachigen Raum gibt, haben wir zusätzlich die Beiträge der Wirtschaftsinformatik Tagungen 2009 und 2011 analysiert. Dabei konnten wir zehn Forschungsberichte identifizieren, die sich schwerpunktmäßig mit dem Thema Technologieakzeptanz befassen. Darunter sind vier, deren Ergebnisse hauptsächlich auf qualitativen Analysen basieren und sechs, deren Aussagen ausschließlich durch quantitative Daten belegt werden. Dies untermauert die Aussage, dass qualitative Forschung im europäischen Bereich keine derart untergeordnete Rolle spielt, wie

¹ Information Systems Research, MIS Quarterly, Journal of Management Information Systems, Information Systems Journal, Journal of the Association for Information Systems, The Journal of Strategic Information Systems, Journal of the ACM International Journal of Electronic Commerce, Journal of Information Technology, Wirtschaftsinformatik

es im amerikanischen Raum der Fall ist. Andererseits ist festzustellen, dass derartige Beiträge ihren Weg in qualitativ hoch gerankte Journals noch nicht gefunden haben.

Hinsichtlich der Ergebnisse lässt sich feststellen, dass quantitative Analysen in der Regel darauf abzielen, die Zusammenhänge zwischen Konstrukten und Variablen mit statistischen Methoden zu überprüfen. Dabei werden entweder gezielt der Einfluss bestimmter Variablen wie beispielsweise Kultur [37] oder Vertrauen [38] auf die Akzeptanz untersucht, oder bereits bekannte Zusammenhänge näher beleuchtet. Der weitaus größte Teil der Akzeptanzanalysen geht dabei wie folgt vor: Verwandte Literatur wird studiert, aus bestehenden Theorien übernommene und argumentativ hergeleitete Konstrukte werden in einem Modell zusammengefasst und anschließend die Hypothesen / Modelle mithilfe empirischen Materials aus quantitativen Umfragen getestet / validiert. Qualitative Daten werden in der Regel nicht zur Exploration herangezogen. Komplexe Zusammenhänge wie das Zusammenspiel zwischen Mensch und Technologie können dabei nicht umfassend abgebildet werden. Ebenso ist es möglich, dass unerwartete und gegebenenfalls wichtige Einflussgrößen übersehen werden [20]. Dabei zeigen die (wenigen) Studien, in denen die Akzeptanz mit qualitativen Methoden erforscht wurde, dass Ergebnisse möglich sind, deren Aussagen, sich außerhalb bekannter Theorien bewegen. So zeigen z. B. Ergebnisse qualitativer Studien die Bedeutung und den Einfluss der Führungskultur [39] oder von Persönlichkeitsmerkmalen [40] auf die Akzeptanz. Darüber hinaus wird deutlich, dass spezifische Faktoren, die von der Art der Technologie abhängen, oft eine wichtige Rolle für die Akzeptanz spielen (z. B. [41]). Im Umfeld wenig erforschter Softwareprodukte ist die Relevanz von Arbeiten, die empirisches Material nur zur Theorieprüfung heranziehen, daher fraglich. Ggf. finden Konstrukte, die von großer Bedeutung sind, in überwiegend argumentativ begründeten Modellen keine Berücksichtigung.

3 Darstellung eines qualitativen Vorgehens in einer explorativen Studie zur Akzeptanz von Projektmanagement-Software

3.1 Vorgehen bei der Analyse der Akzeptanz von Projektmanagement-Software

Im Rahmen eines Forschungsprojektes haben wir die Akzeptanz von Projektmanagementsoftware untersucht. Diese Softwareprodukte dienen der Unterstützung der Anwender bei der Projektabwicklung in allen Projektphasen [42]. Damit zeichnet sich Projektmanagement-Software durch einen klaren Aufgabenbezug in Szenarien mit einer potenziellen Vielzahl von Anwendern mit heterogenen Anforderungen und Vorkenntnissen aus. Dies unterscheidet sie maßgeblich von generischen Softwareprodukten. Die Übertragbarkeit der Ergebnisse der Untersuchung auf andere betriebliche Anwendungssysteme wie ERP-Systeme ist aufgrund des starken betrieblichen Kontextes und des klaren Aufgabenbezugs zu vermuten.

Im Folgenden wird das Vorgehen der qualitativen Untersuchung beschrieben, wie es in Abbildung 1 dargestellt ist.


1.	Untersuchung planen	
1a	Forschungsfrage erarbeiten	
1b	Interessengruppen identifizieren/Interviewpartner auswählen	
1c	Forschungsdesign festlegen	
2.	Untersuchung durchführen	
2a	Interviews führen	
2b	Interviews mitschneiden	
2c	Interviews transkribieren	
3.	Analyse	 mehrfacher Material- durchlauf mit Korrekturen
3a	Analyseeinheiten identifizieren	
3b	Strukturierungsdimensionen festlegen und codieren	
3c	Zuordnung der Analyseeinheiten zu Strukturierungsdimensionen	
4.	Auswertung	
4a	Relevanzskala festlegen	
4b	Relevanzwerte berechnen	
4c	Quantitative Auswertung und Interpretation	
4d	Ableitung von Handlungsempfehlungen aus den Analyseeinheiten	

Abb. 1. Vorgehen der qualitativen Untersuchung

Untersuchung planen. Die hier zu Grunde gelegte Forschungsfrage lautet: welche Merkmale der Anwender, der Software und der einsetzenden Unternehmen beeinflussen die Akzeptanz der Anwender von Projektmanagement-Software unter Berücksichtigung der Richtung der Einflussnahme?

Da der Begriff der Akzeptanz in der bisherigen Akzeptanzforschung oft nur über Hilfskonstrukte wie Nutzungsabsicht oder tatsächliche Benutzung (gemessen in der Dauer der Verwendung) erfasst wurde, war es in diesem Fall wichtig, den Begriff der Akzeptanz eindeutig zu definieren. Akzeptanz wurde in diesem Untersuchungsrahmen festgelegt als: tatsächliche Verwendung der Software unter Berücksichtigung der jeweiligen Rolle, also dem tatsächlichen Bedarf der Benutzung, durch die Anwender.

In der Untersuchung haben wir uns für ein exploratives Vorgehen entschieden. Werkzeug der Exploration in dieser Untersuchung ist ein Interviewleitfaden zur Durchführung von semi-strukturierten Interviews. Dies Verfahren grenzt sich insofern von andern qualitativen Verfahren wie beispielsweise Beobachtungen ab, als dass in dieser Befragung der Erfahrungsschatz aus einer Vielzahl von Projekten vorliegt, anders als eine Beobachtung, die meist die Betrachtung eines konkreten Anwendungsfalls im Fokus hat. Zudem wäre eine Beobachtung in diesem Fall als unangemessen aufwendig einzuschätzen, da eine PMS-Einführung sich über Jahre und über diverse Nutzergruppen hinweg erstrecken kann.

Aus der Forschungsfrage ist abzuleiten, dass mehrere Interessengruppen (Anwender, Hersteller und einsetzende Unternehmen) in der Untersuchung zu adressieren sind. Daher wurde bei der Auswahl der Interviewpartner darauf Rücksicht genommen, dass diese sich in verschiedene Perspektiven auf die Akzeptanz einfinden kön-

nen. Die Interviewpartner wurden sorgfältig gemäß ihrer Qualifikation und Erfahrung in verschiedenen Software-Einführungsprojekten ausgewählt. In der vorliegenden Studie wurden zu annähernd gleichen Teilen Interviewpartner aus dem Bereich der Anwender, Berater und Software-Hersteller befragt. Die erste Gruppe zeichnet sich durch ihren direkten Blickwinkel auf die Einführungssituation aus Anwendersicht und Sicht der einführenden Unternehmen aus. Die Hersteller der zu betrachtenden Software verfügen über Expertenwissen bezüglich Anforderungen der Anwender während aller Phasen der Einführung und der späteren fortschreitenden Verwendung und funktionalen Erweiterung der Software. Die Gruppe der Berater deckt zusammenfassend den Blickwinkel beider zuvor genannter Gruppen aus einer externen Perspektive ab. Vorteilhaft bei der Zusammensetzung der Gruppen ist, dass in dieser Situation bereits vielfältige Erfahrungen, zusammengetragen in verschiedenen Einführungsprojekten und Akzeptanzsituationen expliziert werden. Die Beteiligten sind nicht zwangsläufig Teil des Problems, sondern vielmehr aktiv an der Lösung beteiligt.

Untersuchung durchführen. Die Interviews wurden nach kurzem Vorgespräch, welches der Auflockerung der Situation dient [43], geführt. Sie wurden unter Einverständnis der Teilnehmer mitgeschnitten und anschließend transkribiert.

Analyse. Die so entstandenen Transkripte stellten die Analyseeinheit der Untersuchung dar. Gemäß dem Vorgehen der formalen Strukturierung [44] sollten die relevanten Aussagen der Interviews nach thematischen Kriterien zusammengefasst und einem theoriegeleiteten Kategoriensystem verortet werden.

Die Orientierung an Leitkategorien erlaubt den Bezug zu bestehenden Theorien. Leitkategorien, die nicht in den Analyseeinheiten bestätigt werden, können ausgeschlossen werden. Andere, zuvor nicht definierte Kategorien können aus den Texten abgeleitet werden. Dies ermöglichte es, ein inhaltsreiches System aufzustellen.

Als zu Grunde liegende Strukturierungsdimension wurde zunächst das TAM2 [45] gewählt sowie das Modell von Wixom und Todd [46] und das Task-Technology-Fit Modell [47], da diese Modelle über den größten Aufgabenbezug verfügen. Der Bedarf an einem hohen Aufgabenbezug ergibt sich aus dem Untersuchungsgegenstand, da Projektmanagement-Software an sich einen starken Aufgabenbezug aufweist. Kategorien anderer Modelle wurden sukzessive hinzugezogen.

Relevante Aussagen wurden in Kernaussagen umgewandelt. Kodierregeln vereinfachen die Zuordnung der Kernaussage zur Kategorie.

Es folgte eine wiederholte Sichtung des Materials und eine Revision der Kategorien. Textpassagen, die nicht den bisherigen Kategorien zugeordnet werden konnten, wurden neuen, induktiv entwickelten Kategorien zugeordnet. Die Schritte der Zuordnung wurden in verschiedenen Durchgängen erneut durchlaufen, bis die Kategorien, Kodierregeln und Paraphrasen in den Interviews eine kohärente Struktur aufwiesen und alle relevanten Textpassagen eindeutig den Kategorien zugewiesen wurden.

Die Revision durch verschiedene Personen erfüllt den Objektivitätsanspruch qualitativer Forschung [48]. Um sicher zu gehen, dass relevante Inhalte erfasst und verständlich zusammengefasst wurden, war eine Überprüfung durch mehrere Personen notwendig.

Auswertung. Anschließend wurden die identifizierten Faktoren sortiert und nach der Relevanz gestaffelt. Dieser Schritt erlaubte es, das Modell auf die gewünschte Größe zu bringen und so der Komplexität der Situation wie auch der anzustrebenden Einfachheit (i.S. der Parsimony“) des Modells nachzukommen. Zu diesem Zweck wurden die Kategorien gezählt (Häufigkeit) und inhaltlich bewertet (Relevanz). Insgesamt können vier Relevanz-Werte unterschieden werden: Explizite Ablehnung durch Interviewpartner = (-1), Nichtnennung im Interview = (0), Nennung durch den Interviewpartner = (1) und Nennung durch den Interviewpartner mit Vermerk auf Bedeutsamkeit = (2).

Als Indikatoren der Bedeutsamkeit wurden Ausdrücke der Heftigkeit (beispielsweise Aussagen wie „wichtig ist...“ oder „das darf man keinesfalls vergessen...“) herangezogen. Die Einstufung der Relevanz ergibt sich aus der Notwendigkeit, dass Kategorien der Akzeptanz im Interview auch genannt wurden, um sie ausdrücklich auszuschließen. Daher ist das bloße Zählen der Nennungen allein nicht zielführend.

Zudem wurde geprüft, in wie vielen Interviews die Kategorien genannt wurden (Interviewhäufigkeit). Es ist davon auszugehen, dass Kategorien, die von vielen Interviewpartnern genannt wurden, von höherer Relevanz sind.

Es konnten in den 14 Interviews 396 Aussagen zur Akzeptanz beim Einsatz von PMS extrahiert werden, das entspricht durchschnittlich 28,28 Aussagen pro Interview. Es wurden dabei 33 verschiedene Akzeptanzkategorien identifiziert. Darunter waren 15 Kategorien, die keinen direkten Bezug zu den bekannten Akzeptanztheorien aufweisen. Diese 15 Kategorien ließen sich in drei Gruppen unterteilen. Eine Übersicht ist in Tabelle 1 dargestellt.

Tabelle 1. Kategorien ohne TAM Bezug

Situations-Kategorien	Empfindungs-Kategorien	Einführungsprozess-Kategorien
wahrgenommener Projekterfolg	Vertrauen in Vorgesetzte und Kollegen	Informationsgestaltung bei der Einführung
Anpassbarkeit an Projektprozesse	Vertrauen in die Datensicherheit der Software	Grad der Beteiligung an der Einführung
PM-Erfahrung	Qualität der Schnittstellen	Art des Trainings
Aufgaben-Status	Kapitalausstattung	
Reifegrad		
Rolle im Projekt		
Orientierung an Standards		
PMS-Erfahrung		

Die dritte Gruppe bilden die Kategorien, die einen starken Prozessbezug aufweisen. Sie spielen eine wichtige Rolle für die Akzeptanz der Anwender. Sie sind jedoch aufgrund ihrer Komplexität nicht uneingeschränkt über einen quantitativen Fragebogen nach Art des TAM abfragbar. Vielmehr wurden sie in der vorliegenden Studie in Form von Schilderungen verschiedener Best Practices gesammelt, verglichen und in Handlungsempfehlungen überführt.

3.2 Diskussion des Vorgehens

Das hier beschriebene Vorgehen tritt den Defiziten der bisherigen quantitativen Softwareakzeptanzforschung entgegen. Das zu Stande gekommene Modell enthält deutlich mehr, als lediglich die im Grundmodell des TAM benannten Faktoren. Zusammenhänge zwischen den Faktoren können von den Befragten explizit dargestellt und beschrieben werden.

Die Ergebnisse zeigen, dass viele Faktoren die Akzeptanz beeinflussen, die aus der Theorie so nicht herleitbar gewesen wären. Das Heranziehen bisheriger Theorien und Konstrukte hätte nicht zu vergleichbaren Ergebnissen geführt. Daher ist das Verfahren durchaus als theoriebildend einzustufen. Das hier vorgestellte Verfahren greift nicht nur Kategorien zuvor formulierter Modelle auf, sondern ist die Basis um komplementär weitere Kategorien zu entwickeln und diese zu priorisieren. Beispielhaft sind vor allem Faktoren der emotionalen Ebene anzuführen. So spielt die Art und Weise, wie die Mitarbeiter ihre persönliche berufliche Situation bei der Einführung und beim Erlernen einer Software empfunden haben, eine große Rolle. Vertrauen in Vorgesetzte, Kollegen und die Sicherheit der Software wurden in der Untersuchung oft als wichtige Faktoren der Akzeptanz formuliert. Hinzu kommt, dass die Stärke des Einflusses dieser Faktoren je nach Position und Rolle der Anwender im Zeitverlauf variieren. Die Abbildung derart unterschiedlicher Einflussachsen wäre in einem statistischen Datenmodell nur über moderierende Effekte abzubilden, was das Modell stark verkomplizieren würde.

Andere hier identifizierte Kategorien können hinreichend beschrieben werden. Eine Reduktion zur Aufbereitung innerhalb einer quantitativen Studie hätte den Sinn der Konstrukte zu stark verkürzt. Beispielsweise wäre ein auf wenige Items beschränktes Maß für den Einführungsprozess nicht sinnvoll zu formulieren. Gleiches gilt für den Aspekt der Sicherheit.

Vorteil des qualitativen Vorgehens ist zudem, dass die Teilnehmerzahl der Exploration überschaubar bleibt. Damit kann der Forscher eine Auswahl an qualifizierten Interviewpartnern treffen. Dies macht es möglich, auch Softwareprodukte mit heterogenem Anwenderkreis hinsichtlich ihrer Akzeptanz zu betrachten. Eine Beschränkung der Untersuchung auf generische Produkte mit hohen Anwenderzahlen und homogenen Anforderungen wie in quantitativen Verfahren, bleibt aus. Zudem erfolgt keine Beschränkung auf Teilnehmer in einer exakt identischen Einführungssituation, wie sie in den meisten Akzeptanzstudien zu finden ist [50]. Vielmehr können Anwender, Berater und Hersteller aus verschiedenen Phasen von der Softwareauswahl bis hin zu dauerhaften Verwendung berichten und diese Phasen voneinander abgrenzen.

Durch die spezifische Auswahl der Interviewpartner kann in der Untersuchung genau auf das Ziel der Softwareverwendung eingegangen werden, wie es beispielsweise von Bagozzi [4] gefordert wird. Werden unterschiedliche Interessengruppen der Softwareverwendung befragt, so können mögliche Differenzen im Einsatz mit Hilfe der Forscher aufgedeckt und diskutiert werden. Wichtig ist es dazu, auch die Rolle der Anwender bei der Verwendung der Software zu berücksichtigen. Die unterschiedliche Handhabung der Anwender je nach ihrem praxisbezogenen Aufgabengebiet ist typisch für den Einsatz von Informationssystemen. Die Rolle der Unternehmenskultur und der Status, der mit der Software zu erfüllenden Aufgaben, können in der Diskussion ebenfalls berücksichtigt werden.

In Bezug auf den Informationsgewinn kann dieses Vorgehen als inhaltlich reich eingestuft werden. Die Diskussion mit den Interviewpartnern und die Möglichkeit, nachzufragen und sich Sachverhalte schildern zu lassen, ermöglichte im weiteren Forschungsverlauf eine anschließende Diskussion und Ableitung von Handlungsempfehlungen. Beispielsweise wurden konkrete Maßnahmen bezüglich der Gestaltung von Softwaretrainings und sozialen Lernformen von den Befragten formuliert. Die Handlungsempfehlungen konnten den Interessengruppen (Hersteller, Anwender und einsetzendes Unternehmen) zugeordnet werden. Dies bedeutet gegenüber quantitativen Verfahren einen deutlichen Nutzensgewinn. Bisher war es Ziel der Akzeptanzuntersuchungen Einflussfaktoren zu identifizieren. Die Erhöhung der Akzeptanz durch konkrete Maßnahmen war nicht intendiert und müsste aufgrund der unzureichenden Theoriebildung auch unvollständig bleiben.

Trotz der genannten Vorteile ist dieses Vorgehen nicht gänzlich frei von Fehlerquellen. Selbst bei sorgfältiger Herangehensweise lassen sich Fehler durch Interpretationen der auswertenden Personen nicht immer ausschließen. Die kontextfreie Reduktion kann zudem nie den Gehalt der ursprünglichen Angaben aufrechterhalten. Zudem bleibt zu diskutieren, ob die Orientierung an vorherigen Akzeptanzstudien bei der Bildung der Kategorien nicht zu einer Einschränkung des Blickfelds der Forscher führt. Wir sind jedoch der Meinung, dass explorative Forschung nicht gleichbedeutend mit einem Verzicht auf Erkenntnisse vorliegender Theorien bedeutet. Als Verteidigung dieser Maßnahme lässt sich anführen, dass die Interviews im zentralen Teil offen geführt wurden, also dass die Befragten ohne Konfrontation mit bestehen TAM-Konstrukten in ihren Aussagen eingeschränkt wurden und dass auch Faktoren, die in anderen quantitativen Verfahren bestätigt wurden, die in der vorliegenden Studie keine Bestätigung erfahren haben. Beispielsweise ist das Konstrukt der Subjective Norm des TAM2 [45] anzuführen, welches von den Interviewpartnern der beschriebenen Studie definitiv abgelehnt wurde.

4 Fazit und Ausblick

Qualitative Forschung zur Untersuchung der Akzeptanz von Software ist bisher unterrepräsentiert. Der hier aufgezeigte Ansatz tritt den Defiziten bisheriger quantitativer Forschung entgegen, bzw. bietet er komplementäre Ansätze zur Theoriegewinnung. Projektmanagement-Software-Produkte ähneln anderen Anwendungssoftware-

Systemen in Bezug auf den starken Arbeitsbezug und die betriebliche Prozessunterstützung sowie die Heterogenität der Anwender durchaus. Daher lässt sich das hier skizzierte Vorgehen auf andere Produktgruppen wie ERP-Software übertragen.

Nachteilig am beschriebenen Verfahren ist es, dass nur mit einer ausreichenden Anzahl geschulter Forscher die Objektivität der identifizierten Kategorien sichergestellt werden kann. Qualitative Verfahren setzen einen hohen Einsatz der Forscher bei der Identifikation möglicher Gesprächspartner sowie während der Interviews und Analysen voraus. Dies ist einer der Kritikpunkte qualitativer Forschung wie sie oft [43] diskutiert werden. Qualitative Verfahren bieten großen Interpretationsspielraum. Zudem können in qualitativen Verfahren nicht derart viele Anwender befragt werden, wie es mit standardisierten quantitativen Fragebögen möglich ist.

Der hier vorgestellte qualitative Ansatz stellt ein inhaltsreiches Verfahren dar, das es erlaubt, Einflussgrößen auf die Akzeptanz zu identifizieren und zu diskutieren. Die theoriebildende Herleitung der Faktoren ist eindeutig und direkt aus den Aussagen der Befragten ablesbar. Der direkte Austausch mit den Anwendern und Anwendergruppen lässt es zu, von umstrittenen Zielgrößen wie Use und Intention to Use Abstand zu nehmen. Ergebnis der Studie sind neue Kategorien, die aus bestehenden Akzeptanzstudien nicht ableitbar gewesen wären sowie die eindeutige Ablehnung bestimmter Kategorien aus TAM-Modellen wie Subjective Norm oder Image. Die Ableitung von Handlungsempfehlungen zur Erhöhung der Akzeptanz ist möglich. Verschiedene Maßnahmen können identifiziert, diskutiert und allgemein formuliert werden. Der direkte Austausch erlaubt den Befragten Rückfragen bspw. hinsichtlich Erfahrungen und Best Practices. So kommen klar verständliche Konstrukte zu Stande, die in sinnvolle Handlungsempfehlungen münden. Zudem eignet sich das vorgestellte Verfahren der Theoriebildung, um die Ergebnisse in einer sich anschließenden quantitativen Theorieprüfung zu verifizieren. Beispielsweise lassen sich die identifizierten Kategorien und ihre Zusammenhänge in Form einer quantitativen Fragebogenerhebung mit statistischen Verfahren wie LISREL oder PLS auswerten. Mit einem derartigen multimethodischen Ansatz wären den Anforderungen an Aussagefähigkeit und Generalisierbarkeit erfüllt.

Literatur

1. Ma, Q., Liu, L.: The Technology Acceptance Model: A Meta-Analysis of Empirical Findings. *Journal of Organizational & End User Computing* 16, 59–72 (2004)
2. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 319–340 (1989)
3. Lee, Y., Kozar, K.A., Larsen, K.R.T.: The Technology Acceptance Model: Past, Present, and Future. *Communications of AIS* 12, 752–780 (2003)
4. Bagozzi, R.P.: The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. *J AIS* 8, 244–254 (2007)
5. Venkatesh, V., Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences* 39, 273–315 (2008)

6. de Vreede, G.-J., Jones, N., Mgya, R.J.: Exploring the Application and Acceptance of Group Support Systems in Africa. *Journal of Management Information Systems* 15, 197–234 (1998)
7. van der Heijden, H.: User Acceptance of Hedonic Information Systems. *MIS Quarterly* 28, 695–704 (2004)
8. Sykes, T.A., Venkatesh, V., Gosain, S.: Model of Acceptance with Peer Support: A Social Network Perspective to Understand Employees' System Use. *MIS Quarterly* 33, 371–393 (2009)
9. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly* 36, 157–178 (2012)
10. Wang, W., Benbasat, I.: Trust in and Adoption of Online Recommendation Agents. *JAIS* 6, 72–101 (2005)
11. Schwarz, A., Chin, W.: Looking Forward: Toward an Understanding of the Nature and Definition of IT Acceptance. *Journal of the AIS* 8 (4), article 4 (2008)
12. Orlikowski, W.J.: The sociomateriality of organisational life: considering technology in management research. *Cambridge Journal of Economics* 34, 125-141(2009)
13. Beaudry, A., Pinsonneault, A.: Understanding User Responses to information technology: A coping model of user adaptation. *MIS Quarterly* 29 (3), 493-524 (2005)
14. Pipek, V., Wulf, V.: Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology. *Journal of the Association for Information Systems* 10, 447-473 (2009)
15. DeSanctis, G., Poole, M.S.: Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science* 5 (2), 121-147 (1994)
16. Majchrzak, A., Rice, R.E., Malhotra, A., King, N., Ba, S.: Technology adaption: the case of a computer-supported inter-organizational virtual team¹. *MIS Quarterly* 24 (4), 569-600 (2000)
17. Orlikowski, W.J.: The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science* 3 (3), 398-427 (1992)
18. Lee, A.S., Baskerville, R.L.: Generalizing Generalizability in Information Systems Research. *Information Systems Research* 14, 221–243 (2003)
19. Schepers, J., Wetzels, M.: A Meta-Analysis of the Technology Acceptance Model: Investigating Subjective Norm and Moderation Effects. *Information & Management* 44, 90–103 (2007)
20. Wu, P.F.: A Mixed Methods Approach to Technology Acceptance Research. *JAIS* 13, 172–187 (2012)
21. Orlikowski, W.J., Baroudi, J.J.: Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research* 2, 1–28 (1991)
22. Seddon, P.B.: A Respecification and Extension of the DeLone and McLean model of IS success. *IS Research* 8, 240–253 (1997)
23. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27, 425–478 (2003)
24. Hirschheim, R.: Introduction to the Special Issue on "Quo Vadis TAM - Issues and Reflections on Technology Acceptance Research". *JAIS* 8 (4), 204–205 (2007)
25. Benbasat, I., Barki, H.: Quo vadis, TAM? *JAIS* 8 (4), 212–218 (2007)
26. Goodhue, D.L.: Comment on Benbasat and Barki's "Quo Vadis TAM" article. *JAIS* 8 (4), 220–222 (2007)
27. Palvia, P., En Mao, P., Salam, A.F., Soliman, K.S.: Management Information Systems Research: What's there in an Methodology?. *Communications of AIS* 11, 289–308 (2003)

28. Lee, A.S., Baskerville, R.L.: Conceptualizing Generalizability: New Contributions and a Reply. *MIS Quarterly* 36, 749-A7 (2012)
29. Meyer, M.M.: Softwareunterstützung für das Projektmanagement. Auswahl, Einführung und Einsatz softwaregestützter Instrumente, Anforderungen und Lösungen sowie medialer Transfer. IPMI, Bremen (2007)
30. Hirschheim, R., Klein, H.K.: A Glorious and Not-So-Short History of the Information Systems Field. *Journal of the Association for Information Systems* 13, 188–235 (2012)
31. Chen, W.S., Hirschheim, R.: A Paradigmatic and Methodological Examination of Information Systems Research from 1991 to 2001. *Information Systems Journal* 14, 197–235 (2004)
32. Deng, X., Doll, W.J., Hendrickson, A.R., Scazzero, J.A.: A Multi-group Analysis of Structural Invariance: An Illustration Using the Technology Acceptance Model. *Information & Management* 42, 745–759 (2005)
33. King, W.R., He, J.: A Meta-Analysis of the Technology Acceptance Model. *Information & Management* 43, 740–755 (2006)
34. Legris, P., Ingham, J., Colletette, P.: Why Do People Use Information Technology? A Critical Review of the Technology Acceptance Model. *Information & Management* 40, 191 (2003)
35. Wu, K., Zhao, Y., Zhu, Q., Tan, X., Zheng, H.: A Meta-Analysis of the Impact of Trust on Technology Acceptance Model: Investigation of Moderating Influence of Subject and Context Type. *International Journal of Information Management* 31, 572–581 (2011)
36. Turner, M., Kitchenham, B., Brereton, P., Charters, S., Budgen, D.: Does the Technology Acceptance Model Predict Actual Use? A Systematic Literature Review. *Information & Software Technology* 52, 463–479 (2010)
37. Srite, M., Karahanna, E.: The Role of Espoused National Cultural Values in Technology Acceptance. *MIS Quarterly* 30, 679–704 (2006)
38. Turel, O., Yuan, Y., Connelly, C.E.: In Justice We Trust: Predicting User Acceptance of E-Customer Services. *Journal of Management Information Systems* 24, 123–151 (2008)
39. Kavanagh, M.H., Ashkanasy, N.M.: The Impact of Leadership and Change Management Strategy on Organizational Culture and Individual Acceptance of Change during a Merger. *British Journal of Management* 17, 81–103 (2006)
40. Ouadahi, J.: A Qualitative Analysis of Factors Associated with User Acceptance and Rejection of a New Workplace Information System in the Public Sector: A Conceptual Model. *Canadian Journal of Administrative Sciences* 25, 201–213 (2008)
41. Zoellner, J., Schweizer-Ries, P., Wemheuer, C.: Public Acceptance of Renewable Energies: Results from Case Studies in Germany. *Energy Policy* 36, 4136–4141 (2008)
42. A Guide to the Project Management Body of Knowledge (PMBOK guide) Project Management Institute, Inc., Newtown Square, Pa (2008)
43. Walsham, G.: Doing Interpretive Research. *European Journal of Information Systems* 15, 320–330 (2006)
44. Mayring, P.: *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz Pädagogik Beltz, Weinheim u.a (2010)
45. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science* 46, 186–204 (2000)
46. Wixom, B.H., Todd, P.A.: A Theoretical Integration of User Satisfaction and Technology Acceptance. *IS Research* 16, 85–102 (2005)
47. Goodhue, D.L., Thompson, R.L.: Task-Technology Fit and Individual Performance. *MIS Quarterly* 19, 213-236 (1995)

48. Flick, U., von Kardorff, E., Steinke, I.: *Qualitative Forschung. Ein Handbuch.* Rowohlt, Reinbek bei Hamburg (2007)
49. Raab-Steiner, E., Benesch, M.: *Der Fragebogen. Von der Forschungsidee zur SPSS-Auswertung.* Facultas.wuv, Wien (2012)
50. Brown, S.A., Dennis, A.R., Venkatesh, V.: Predicting Collaboration Technology Use: Integrating Technology Adoption and Collaboration Research. *Journal of Management Information Systems* 27, 9–54 (2010)

Building Taxonomies in IS and Management – A Systematic Approach Based on Content Analysis

Dennis M. Steininger, Manuel Trenz, and Daniel J. Veit

University of Mannheim, Business School, Mannheim, Germany
{steininger,trenz,veit}@bwl.uni-mannheim.de

Abstract. Classification schemes such as taxonomies are important groundwork for research on many topics in Information Systems (IS) and Management. They make investigating topics manageable by allowing researchers to delimit their work to certain taxa or types and provide a basis for generalization. Opposed to theoretically grounded typologies, taxonomies are empirically derived from entities of a phenomenon under investigation and therefore have several advantages such as more detailed and exhaustive coverage. Nevertheless, research is still missing a clear set of procedures on how to empirically build taxonomies. We tackle this topic by suggesting an inductive approach based on the procedures of content and cluster analysis. Each of the proposed six steps is amended with comprehensive state of the art guidelines, suggestions, alternatives and formative measures of reliability and validity.

Keywords: Taxonomy, Classification, Typology, Content analysis, Method

1 Introduction

Classification schemes allow the systematic ordering or sorting of phenomena into similar groups or classes. They are of fundamental importance for science and academic research in disciplines such as Information Systems (IS) and Management [1-2]. Wolf [3] emphasizes this importance by explaining the links and stating that verification of laws of science may only succeed after classification has been completed since it is “the first and last method employed by science” [3]. Hence, classification schemes such as taxonomies make investigating phenomena manageable by allowing researchers to delimit their work to certain classes (i.e. taxa or types) such as IS technologies or firms and also provide a basis for generalization. This allows building theories that apply to certain classes of the developed schemes. When classifying an area of investigation, two different approaches can be used: typologies or taxonomies. Typologies are created deductively by classifying the objects into predefined groups that are created based on intuition or previously existing knowledge and theory [4]. Especially when examining an unexplored area of research, the threat of researcher bias or general misconception is very high, since existing theory is limited. Opposed to theoretically grounded typologies, taxonomies are derived inductively from empirical data (i.e. entities of a phenomenon under investigation) and therefore have several

advantages such as more detailed and exhaustive coverage and mutual exclusiveness of classes. Nevertheless, IS research is still missing a clear set of rigorous procedures on how to empirically build taxonomies of artifacts, systems, user behavior or processes. Especially in fast moving areas such as IS, it is important to be able to describe new phenomena rigorously and quickly by applying systematic actions. Building on these thoughts outlined above we propose the following research question for this article:

How and by which procedures can mutually exclusive and collectively exhaustive taxonomies be built rigorously in the IS and Management disciplines from empirical entities?

We tackle this question by suggesting an inductive empirical approach based on the procedures of content and cluster analysis. Content analysis allows a systematic and rigorous analysis of entities under investigation to get a first grasp on their characteristics, associated manifestations and densities. Based on these results, procedures of cluster analysis can be applied to derive final classes. The remainder of this paper is structured as follows: In the second section we propose six steps to build taxonomies. Each of these steps is amended with state of the art guidelines, alternatives and formative measures of reliability and validity. Summative measures of taxonomic quality are also depicted for evaluating final taxonomic constructs. In the last section we sum up our findings, address the usefulness of taxonomic outcomes and identify interesting topics in IS that might be investigated by using the introduced method.

2 The Process of Taxonomy-Building

We introduce detailed steps and procedures to build taxonomies in IS-related phenomena using content and cluster analysis. The process is based on Steininger et al. [5] who use clustering and mainly content analysis to inductively build a taxonomic framework of Web 2.0 characteristics. This article can be seen as a working example. We added inspirations from the articles of Nag et al. [6], defining Strategic Management via content analysis and clustering and Al-Debei and Avison [7] developing a business model framework through content analysis. Content analysis is a scientific research technique to gain “replicable and valid inferences from text” [8] and thereby find trends, characteristics, patterns or densities. Materials for analysis might include written or spoken texts as transcripts. Objectivity, validity and reliability of the outcomes are obtained through rigorous rules and systematic procedures, which have been refined and adapted to the various needs of different disciplines over time [5], [9-11] and distinguish content analysis from regular critical reading. The aforementioned potential in rigorously and reliably uncovering characteristics and patterns is of high relevance for constructing taxonomies. Hence, we adapt state of the art procedures of inductive and deductive content analysis for major parts of the taxonomy-building process suggested in the remainder of this paper. The outline of our idea is to define a phenomenon of investigation and collect examples resembling the phenomenon as entities of investigation. We then inductively develop the characteristics of the phenomenon from these entities and deductively measure the manifestation of the characteristics for each entity.

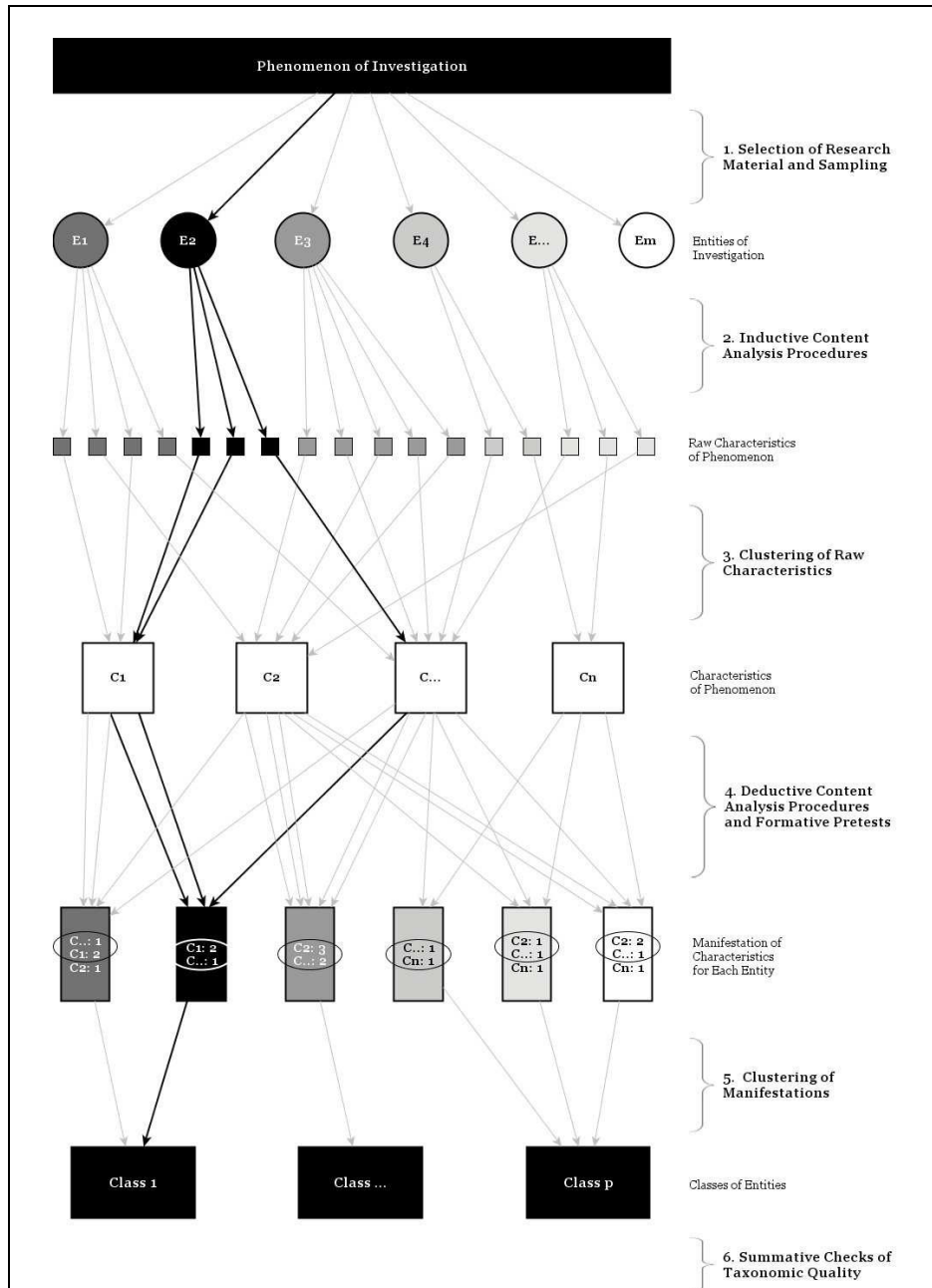


Fig. 1. Detailed Overview on the Taxonomy-Building Process

We finally propose to cluster the entities into classes (i.e. taxa) by analyzing their manifestations and densities of characteristics. The entire process is depicted in Figure 1, highlighted for one entity (marked with black ink). It starts with a definition of the phenomenon under investigation (e.g. electronic business models). This entails a clear statement of the research question (e.g. what classes of electronic business models do exist?). After these initial specifications, a set or population of entities and their textual descriptions resembling the phenomenon (e.g. examples of existing electronic business models) are required as a basis for analysis, which is addressed in our first suggested step on the selection and sampling of entities. Analyzing the manifestation of the phenomenon's characteristics for each entity is needed to proceed in building the taxonomy. Since we assume missing theoretical foundations on the characteristics of the phenomenon, we describe procedures on how to inductively derive raw characteristics from selected entities by using content analysis (step 2). Raw characteristics are subsequently suggested to be reduced to main characteristics of the phenomenon under investigation (e.g. characteristics of electronic business models) by applying cluster analysis (step 3). These two steps might be skipped if our assumption does not hold true and there are already existing and exhaustive definitions of characteristics for the phenomenon in theory which can be utilized for the fourth step. In this fourth step we suggest deductive content analysis procedures to measure the manifestations and densities of the characteristics for each entity (e.g. how often is a characteristic mentioned in the textual material for one entity). This can be reached through analyzing the entities by applying a coding scheme of characteristics, which might be constructed from the inductively developed (cf. steps 2/3) or aforementioned theoretically derived characteristics. The classes of similar entities for the taxonomy (e.g. virtual shopping malls) are then built by suggested procedures of cluster analysis on the resulting manifestations (step 5). We amend this penultimate step by propositions and guidelines on measures for taxonomic quality (e.g. mutual exclusiveness). Details and guidelines on each of our suggested steps are given in the sections below.

2.1 Selection of Research Material and Sampling

Entities of investigation (e.g. firms using an electronic business model) are needed as empirical research material to develop and retrieve characteristics, manifestations and final classes (i.e. taxa) for a phenomenon. We explain procedures for selecting and sampling these entities throughout this section and amend them with hints on data sources and data collection techniques to gain rich data on the selected entities.

A representative sampling of entities might be used but in many cases neither be manageable nor required. Instead, we propose to follow a theoretical sampling approach as suggested by Eisenhardt [12]. This means broadly choosing the entities of investigation for variation, heterogeneity (i.e. unique cases) or replication instead of random selection [13]. The availability of existing textual (e.g. case descriptions, annual or mission statements, product descriptions, websites, directories) or transcribable (e.g. interviews) descriptions for the entities might also be taken into account as a factor of selection during this sampling process. We suggest collecting descriptive data of the entities by following the sources of evidence given in Table 1.

Table 1. Possible Sources of Descriptions for Entities [13-14]

Name	Application	Advantages	Disadvantages
Documentation and Archival Records	Usually available in written form.	<ul style="list-style-type: none"> • Stable • Unobtrusive 	<ul style="list-style-type: none"> • Bias of author unknown • Retrievability and Access
Interviews	Transcription by person independent from interviewer. Final approval of transcript by interviewee.	<ul style="list-style-type: none"> • Targeted • Insightful 	<ul style="list-style-type: none"> • Poor question bias • Response bias • Reflexivity
Fieldwork	Written memos of direct or participant observation. Final check of memos by participants.	<ul style="list-style-type: none"> • Real-time coverage • Contextual • Insightful into behavior and motives 	<ul style="list-style-type: none"> • Time-consuming • Observer bias • Reflexivity
Physical Artifacts	Use of existing descriptions or composition of descriptive memos by two independent authors.	<ul style="list-style-type: none"> • Insightful into cultural features and technical operations 	<ul style="list-style-type: none"> • Selectivity • Availability • Access

It is recommendable to use similar sources of evidence for all entities. Triangulation of more than one source might enrich the descriptions and lead to more robust results, cf. [13]. We suggest listing derived entities in a longlist (*LL*). If entities are gained from different sources, this list needs to be cleaned from possible duplicates. The introduction of a selection factor (*SF*) can help to prepare the *LL* for further proceedings [5]. This selection factor might encompass extra credit points for criteria such as an entity being a unique or extreme case, certain keywords within the name of an entity for restriction to a specific area of interest or the availability of evidence for an entity. In a final step the *LL* has to be sorted in descending order by *SF*. Entities at the lower end of the list not reaching a certain selection factor might now be truncated which results in shortlist (*SL*). Different approaches to gain this shortlist might also be applied (i.e. taking a sample of entities from an existing journal paper on the phenomenon). The *SL* is to be amended with an ascending research material ID (*i*) for each entity in a finalizing step.

2.2 Inductive Content Analysis Procedures

In this second step of our suggested outline, we present a set of procedures and guidelines on how to inductively develop raw characteristics from textual descriptions of the selected entities from the preceding section.

After specification of the entities of investigation and their sampling as research material, the unit of analysis needs to be defined subsequently. This addresses the issue of “the basic unit of text [e.g. word or paragraph] to be classified” [15] into the categories of characteristics derived in succeeding steps. The configuration of this unit has a considerable impact on quality and reliability of research results. Choosing a smaller unit (e.g. word) usually leads to higher reliability and possible automation but might corrode results which focus on larger meanings than transported by single words [16]. Following Kassarian [17], the ‘theme’ is usually suggested for this type of taxonomic method ensuring the capturing of word or sentence-spanning ideas es-

pecially within the inductive phase of building raw characteristics. To stabilize the results and reliabilities, entire sentences are to be used as the operationalized coding unit, which leads to solely coding a category once within one sentence [5]. In the suggested approach the raw characteristics are to be developed inductively from the selected research material (i.e. entities of investigation). This is done to initially capture the characteristics of the phenomenon of investigation, which are needed as groundwork for further analysis.

Table 2. Units of Analysis and Implications (adapted from [17])

Unit	Description	Advantages	Disadvantages
Word	Analysis of single words such as key symbols or value-laden terms	<ul style="list-style-type: none"> • Ease of coding • Ease of automation • Highest reliabilities 	<ul style="list-style-type: none"> • Loss of context • Loss of word-spanning ideas
Sentence	Analysis of entire sentences	<ul style="list-style-type: none"> • Relative ease of coding • Clear demarcation of unit borders 	<ul style="list-style-type: none"> • Loss of sentence-spanning ideas
Theme	Analysis of single assertions about a subject	<ul style="list-style-type: none"> • Capturing of entire subjects of investigation • Very useful in most content analyses 	<ul style="list-style-type: none"> • Ambiguous unit borders • Difficult coding • Lower reliabilities
Item	Entire documents such as speeches, letters, manuals	<ul style="list-style-type: none"> • Useful in classifying entire documents 	<ul style="list-style-type: none"> • Often too gross for most research
Character	Mostly used in the analysis of streaming media or commercials to analyze heroes, bad guys etc.	<ul style="list-style-type: none"> • Useful in the analysis of behavior or communication of actors • Might be of interest to develop taxonomies of user behavior in IS 	<ul style="list-style-type: none"> • Sometimes Ambiguous unit borders • Context might not be captured
Space and Time	Analysis by column (e.g. newspaper), line, paragraph or minute	<ul style="list-style-type: none"> • Useful for historical timeline analysis and longitudinal taxonomies • Clear demarcation of unit borders 	<ul style="list-style-type: none"> • Loss of unit-spanning ideas and context

Based on raw characteristics building rules [18], the research material needs to be worked through consecutively and raw characteristics are to be defined beginning with the first selected entity of investigation. Each occurrence of a new or additional raw characteristic-building incident needs to be marked and uniquely numbered using the research material ID i (cf. Section 2.1). If the marked and colored occurrence in the text defines a nonexistent characteristic, a new and unique raw characteristic ID (r) is hyphenated (e.g. $i.1-r$). If the occurrence matches with an existing raw characteristic and only adds richness to the description of the characteristic, that existing characteristic number is to be used instead, and the mark is suggested to be set in a different color (e.g. dark blue). All raw characteristics categories are suggested to be summoned in a list (RcL). This process is to be continued until saturation is reached (i.e. no new raw characteristics can be derived from the research material entities) [19].

2.3 Clustering of Raw Characteristics

In this section we develop a set of guidelines on how to reduce and cluster the raw characteristics developed through the procedures outlined above. The goal of this step is to reach a generalizable and manageable set of main characteristics on the phenomenon of investigation, which can be used for further analysis.

As suggested by Mayring [19], the entire list of raw characteristics has to be iteratively reduced and qualitatively bundled until main characteristics emerge. We depict some of the approaches available to operationalize this task in the following. A first approach is suggested by Eisenhardt and Bourgeois [20] in iteratively comparing within-group similarities (i.e. groups of similar raw characteristics) and intergroup differences. The technique can be advanced by using matrices and introducing continuous measurement scales for comparison [20]. As an alternative approach an iterative comparison of pairs can be used by listing similarities and differences for each pair [12]. Another way to operationalize the task of grouping the raw characteristics into categories of main characteristics might be based on the approach of Steininger et al. [5]. They suggest having at least two independent researchers who are familiar with the topic judge proximities of paired raw characteristics in a matrix ranging from 100 to show perfect similarity to zero reflecting complete independence. Whichever approach is finally used, each of the resulting main characteristics is to be provided with a descriptive name, which is ideally developed inductively from associated bundles of raw characteristics [19]. From these grouped resulting main characteristics of the phenomenon under investigation, a category or coding scheme of characteristics needs to be developed. This is reached through amending each main characteristic with explanations, 'anchor examples' from the associated and coded raw characteristics and coding rules (i.e. rules on when an occurrence of a characteristic needs to be coded or excluded during analysis). For quality assurance the scheme might be tested by three or four judges following the suggestions of Moore and Benbasat [21].

2.4 Formative Pretests and Deductive Content Analysis Procedures

In the following we depict the deductive content analysis of the sampled entities based on the main characteristics coding scheme developed in the preceding steps. This is needed to ensure formative quality and reliability of the coding scheme and to find manifestations and densities of characteristics for each entity. A content analytical core component is the classification of aforementioned units of analysis into the categories of characteristics by independent researchers. This process is typically referred to as 'coding' [22] and requires the category scheme of characteristics developed above. To capture word-spanning meanings and stabilize the results and reliabilities, we suggested the theme as coding unit and entire sentences as the operationalized coding unit in this study, which leads to only coding a certain category once within one sentence [17]. The finalized category scheme of characteristics (also referred to as coding scheme) is iteratively to be used and adjusted for an extensive training of coders. At least a second independent coder needs to be employed to ensure stable results and calculate intercoder reliabilities [23]. The coder(s) need(s) to be

trained in a one day workshop using research materials from LL with the lowest SF. The coding scheme and rules have to be adjusted iteratively to sort out ambiguities through discussion of non-matching codings. The procedure is repeated with different materials until the overall agreement (reliability) of all coders is calculated above 0.8, cf. [24]. This ensures intersubjectively comprehensible results and verifies the decency of the main characteristics coding scheme. Clearly distinguishable and exclusive categories of main characteristics are thereby ensured. We suggest using Krippendorff's Alpha for a sensitive and advanced measurement or the most commonly used simple 'percent agreement' reliability measure of Holsti [25]. More details on possible measures, their mathematical references, advantages and disadvantages are given in Table 3. All calculated reliabilities, discussions and adjustments made to the coding scheme or the coding rules need to be collected and given in a transparent and comprehensive manner for reproducibility (e.g. 'If there are two occurrences of the same subcategory within one sentence, only the first occurrence is to be coded, counted and marked'). Density results of the materials used for training shall be discarded after calculation of agreements and not be used for the building of final classes.

After finishing the aforementioned amendments to the coding scheme during the training session, the main coding process for the entire research material entities is started. This is done by analyzing the entire evidence of each entity for occurrences (i.e. manifestations) of the main characteristics categories. All manifestations are to be marked and counted within the materials by category and entity. They are individually deemed as belonging to a certain category of characteristics. Finally all manifestations in the evidence of each entity are to be counted separately for every category. We suggest transforming these results into relative numbers (i.e. relative manifestations) and thereby making them comparable through dividing them by the number of averaged sentences in the sources of evidence for each entity. This number is calculated by counting the words of an entity's sources of evidence and dividing the results by 22. The number of 22 is the average of words contained within a sentence in English texts reported by Charniak [26]. For readability reasons the averaged sentences are interchangeably referred to as 'sentences' in the following. No further refinements to the coding scheme and coding rules within this main coding process are to be made. Results are not to be exchanged or discussed by the coders during this main phase [23]. It is suggested to employ coders independently from the ones used for adjusting the coding scheme if possible. After finishing the coding process of the entire research material, the summative reliabilities need to be calculated for the resulting manifestations. Pavlou and Demoka [27] suggest also calculating intracoder reliabilities by having each coder re-code a sample after a certain time. There is no common absolute number of these agreements which is found to be satisfactory in the academic discussion on reliabilities. This is due to large differences especially in the units of analysis and coding but also in category systems, complexity of the evaluated contents and coder experience on the phenomenon. Nevertheless, a reliability of at least 0.7 to 0.85 is seen as acceptable and reachable by many authors (e.g. [8], [23], [28]) for the 'theme' as the unit of analysis that we suggest for this type of study.

Table 3. Frequently Cited Measures of Inter-coder Reliability for Content Analysis

Name	Advantages	Disadvantages
Krippendorff's Alpha [29]	<ul style="list-style-type: none"> • Allows any number of coders • Takes into account agreements by chance • Takes into account low coding numbers • Takes into account number of categories • Allows binary, nominal, ordinal, interval, ratio, polar and circular data • Allows measuring of incomplete data 	<ul style="list-style-type: none"> • Complex application • Extensive details of data regarding coded occurrences needed
Holsti's Percent Agreement [25]	<ul style="list-style-type: none"> • Very facile and quick application • Basic calculations 	Does not take into account variables such as the number of categories, correct codings on incident etc.
Scott's Pi [30]	<ul style="list-style-type: none"> • Relatively facile and quick application 	<ul style="list-style-type: none"> • Only allows nominal data • Assumes same distribution of coder responses
Fleiss' Kappa [31]	<ul style="list-style-type: none"> • Relatively facile and quick application • Extends Scott's Pi by allowing multiple coders 	<ul style="list-style-type: none"> • Only allows nominal data • Assumes same distribution of coder responses
Cohen's Kappa [32]	<ul style="list-style-type: none"> • Takes into account agreements occurring by chance • Does not assume same distribution of coder responses 	<ul style="list-style-type: none"> • Sometimes considered a too conservative measure • Only allows measuring of two coders

2.5 Quantitative Clustering of Manifestations

Having verified the manifestations of the characteristics of each entity enables us to group the different entities. Thereby a set of classes (of entities) within the phenomenon of investigation can be identified. These classifications have usually been performed subjectively based on researchers' ideas or intuition. Using our empirically derived and standardized densities instead leads to more objective classifications. Following the inductive procedure, again, no classes were predefined but instead derived inductively from the data sources.

The main goal of this step is to identify classes that are mutually exclusive and collectively exhaustive. This means that there must be an appropriate class for each entity and each entity must fit into one class only [4]. Furthermore the classification should be generally applicable. The latter requirement is met by the extensive sampling method applied before which ensures that the data used appropriately represents the phenomenon. The former two requirements are addressed by cluster analysis. Cluster analysis generally aims at finding classes such that entities within the same group are similar to each other while entities in different groups are as dissimilar as possible. The five typical steps of cluster analysis are outlined based on our problem [33]: (1) Selection of a sample to be clustered, (2) Definition of a set of variables on which to measure the entities in the sample, (3) Computation of similarities among the entities, (4) Use of a cluster analysis method to create groups of similar entities, (5) Validation of the resulting cluster solution.

The first step, selecting the sample, has already taken place. Regarding the selection of the cluster variables, which usually is a very complicated procedure [34], it is

again very helpful that we have already identified and reduced the relevant characteristics in the previous qualitative steps. Therefore, we can directly create the data matrix containing the densities of the characteristics that correspond to the different entities (cf. Table 4). In the next step, the similarity calculation takes place. Due to the standardized scale of manifestations (i.e. relative manifestations), the Minkowski distance¹ can be used to calculate these values without having to compute weights for the different characteristics [35] (cf. Table 5). The elimination of potential single outliers that have a high distance to all other entities has to be checked manually by an in-depth analysis of the underlying data of this entity. Rash elimination of entities can lead to problems in the validity of the resulting taxonomy and should be avoided.

Many different cluster methods can be applied in order to derive clusters from this data. Generally, partitioning methods like K-Means [36] have been shown to be superior to hierarchical methods in this case [37]. Nevertheless, these methods need a priori information about the starting points and the number of clusters which may not be available when investigating a new phenomenon inductively. In this case, it might be useful to apply Ward's minimum variance method [38] to derive preliminary clusters. Their center can then be used in a partitioning algorithm like K-Means [37]. Common software package like SPSS or SAS can be used to process steps 3 and 4.

Table 4. The Manifestation Matrix of Entities and Characteristics (cf. [5])

	Characteristics			
Entities	C₁	C₂	...	C_n
E₁	x ₁₁	x ₁₂	...	x _{1n}
...
E_m	x _{m1}	x _{m2}	...	x _{mn}

Table 5. The Distance Matrix of Entities

Entities	E₁	E₂	E₃	...	E_{m-1}
E₁	d ₂₁ = d ₁₂				
E₂	d ₁₃	d ₂₃			
E₃	d ₁₄	d ₂₄	d ₃₄		
...	
E_m	d _{1m}	d _{(m-1)m}

Despite the importance of exhaustiveness and mutual exclusiveness, further quality indicators can be addressed. Checking the quality of classifications has been discussed in detail by Aldenderfer and Blashfield [33]. They suggest two major techniques that are relevant to our procedure: Significance tests and replication. Multivariate analysis of variance (MANOVA) or discriminant analysis can be used to check the significance of the clusters. However, this method has been criticized for indicating high significance even for very bad clusters. A solution for this problem might be

¹ $d(i,j) = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + |x_{i3} - x_{j3}|^q + \dots + |x_{im} - x_{jm}|^q)^{1/q}$, where q is natural number larger or equal to 1, describes the distance between the entities i and j . Most algorithms use Manhattan distances ($q = 1$) or Euclidian distances ($q = 2$)

the inclusion of external variables which is difficult when analyzing a new phenomenon [33]. The replication technique can be used to check for internal consistency of the classification. If the base of entities is large enough, the split-half method can be applied. Two random parts of the same are clustered independently using the same clustering method. If the same classes occur across different subsets of entities, this indicates further generalizability of the classification. Another way of replication is to use different clustering methods with the same data. If the same clusters are derived, the results indicate a high validity of the classification [33]. After having the clusters validated, the different classes have to be interpreted. For better understanding, they should also be described verbally. This usually complex task can be accomplished using the codings and descriptions of the entities within one class. The distribution of these codings already describes the characteristics of a certain class. If the number of entities in one class is very high, the naming should be based on the characteristics of the entities in the center of the class. The clusters should then be named inductively out of the names and characteristics from their associated entities [19].

2.6 Summative Checks of Taxonomic Quality

As discussed before, checking taxonomic quality is a very challenging task. Mutual exclusiveness and collective exhaustiveness are the two major quality measures that a high-quality taxonomy has to meet [4]. In order to increase and verify the validity of our method, we suggest performing an additional (optional) step to test discriminant validity of the classification (based on [21], [39]). If additional entities that have not been used for the taxonomy building are available, these entities should be combined with the entities from the sample into a common pool. The additional entities can be coded using the deductive procedure outlined before and then be sorted into the classes mathematically to also have their class affiliations for subsequent comparison. Three to four judges are given the names and verbal descriptions of the classes that have been derived in the previous steps. The judges now sort all entities from the pool into the classes. Two measures can be applied to the results of this sorting process. The first one measures the inter-judge reliability and focuses on the question of judges sorting the same entities into the same classes. We again suggest Krippendorff's Alpha [29] or Holsti's percent agreement [25] to measure the level of agreement between the judges and thereby determine whether or not the descriptions precisely define the classes. Reliabilities above 0.7 can be seen as satisfactory [8]. If this level is not reached the descriptions of the classes need to be enhanced iteratively. A lack of increased inter-judge reliability even with refined descriptions indicates a general problem regarding the mutual exclusiveness or the collective exhaustiveness. Furthermore, for each class, a cumulated overall measure of correctly placed entities can be calculated.² This differs from the previous measure since it challenges the strength of the different classes separately. No description of a reasonable score for this measure is

² The overall measure for the quality of the class is defined as $B(i) = \frac{\#E_c}{\#E} \in (0,1)$, where $\#E_c$ is the number of correctly selected entities into class i by all judges and $\#E$ describes the number of entities supposed to be sorted into this class.

described in literature. As a rule of thumb, the interval between 0.7 and 0.85 discussed above [23], cf. [28], [40] can also be applied as a good indicator for this measure. A high value points to high construct validity and reliability of the class. This method can also be used rather qualitatively to identify critical class definitions and borders between two classes that should be refined.

2.7 Limitations of the Method

Potential limitations regarding the procedures introduced throughout this article should be taken into account. They are given below and if counter measures do exist, they are also depicted in the following. Overall, we have tried to keep the complexity of the process low. Nevertheless, it might inhibit broader use. The process of inductively constructing raw characteristics from the entities is continued until saturation. This allows gaining real knowledge and deep insights on classes. Nevertheless, theoretical saturation is critical to identify. This might lead to missing definitions of characteristics threatening the collective exhaustiveness. The probability seems low since we suggested measures to objectify significant saturation within the inductive process. Inductively built categories might also be biased by a coder's world views or insights on the phenomenon. Lowering the likelihood of such a bias might be reached through introducing more than one coder for inductively building the raw characteristics. Construction of main characteristics from raw characteristics might also be subject to coder's bias since they are qualitatively clustered. Improvement within this area might be reached by applying large proximity matrices judged by more than one person and statistical cluster analysis for their entire set.

The method of using averaged sentences for comparability reasons might lead to excessive numbers of coded sentences since figures or tables within the sources of evidence might be handled as text. This is additionally fostered by the assumption during calculations that all sentences only contain one code, which must not hold true since the rules allow coding a sentence twice with two different categories. One major critique regarding cluster analysis is that it lacks theoretical foundation. Therefore the identified clusters may simply be statistical artifacts that capitalize on random numerical variation across entities [41]. Furthermore, cluster analysis might also find classes in situations where no clusters exist, e.g. [33]. Our approach tries to invalidate the criticisms partly because the clusters are directly named and described based on the densities of their characteristics and are therefore no artificial constructs [19]. Another main critique of cluster analysis is the potential multicollinearity among characteristics that may lead to overweighting of certain aspects [42]. Using more advanced distance measures like the Mahalanobis distance might solve this issue [43], but this measure is supported neither by Ward's minimum variance method [38] nor by software like SPSS and SAS. However, our approach addresses this issue early in the research process. Since the characteristics of the topic are derived from the raw categories inductively and by controlling for weakness of the single characteristics [28], the risk of multicollinearity issues is reduced.

3 Conclusion

We outlined and developed a method of building taxonomic classification schemes for the IS and Management disciplines throughout this papers. Although the importance of such classifications is seen as very high in the research community [1-3], [44], these classifications have usually been performed subjectively based on researchers' ideas or intuition. The delineated approach enables researchers to derive classifications empirically leading to more objective classifications [4]. In essence we proposed six subsequent steps relying on content and cluster analysis. Especially the use of content analysis in this context enhances the available set of techniques within our field. The first step begins with the sampling of entities and their sources of evidence as instantiations or examples of the topic. Since our method is focusing on new and unexplored topics of investigation, we assumed no theoretical basis of the topic to be available. Accordingly, the second and third step proposed to develop the characteristics of the topic from selected entities by using inductive content analysis procedures. Based on these results we proposed a fourth step of deductive content analysis to find manifestations and densities of the derived characteristics for each entity. Cluster analysis is then applied to identify specific classes in the research material, leading to a taxonomic classification scheme. Formative state of the art procedures for quality assurance were suggested throughout all steps of the method. Additionally, summative measures of taxonomic quality for the resulting constructs are outlined. We conclude with an extensive discussion of potential limitations of our method. We believe that our results will help academics to develop empirically grounded rigorous taxonomies in their fields of research by applying our suggestions, guidelines and depicted alternatives. Taxonomies are important vehicles in IS and Management research since they allow limiting investigations on a topic to certain subclasses or taxa, which makes research projects more manageable. Lastly, they are of high value for intra- and inter-class generalization, enabling the development of theories through analysis of these classes and their generalizations. There are innumerable applications of our method in the field of IS research. New and upcoming phenomena such as cloud computing applications, crowdsourcing services might need taxonomic classification, but also long standing non-empirically grounded typologies in areas such as outsourcing, operational application software systems or electronic business model research might be revisited and updated by applying our method to the topic.

References

1. Kantor, J.R.: *The Logic of Modern Science*. University of Akron Press, Akron (1953)
2. Kemeny, J.G.: *A Philosopher Looks at Science*. Van Nostrand, Princeton (1959)
3. Wolf, A.: *Essentials of Scientific Method*. Macmillan, New York (1926)
4. Bailey, K.D.: *Typologies and Taxonomies: An Introduction to Classification Techniques*. Sage Publications, Thousand Oaks (1994)
5. Steininger, D.M., Huntgeburth, J.C., Veit, D.J.: A Systemizing Research Framework for Web 2.0. In: *ECIS 2011 Proceedings*, pp. 1–13. AISeL, Helsinki, Finland (2011)

6. Nag, R., Hambrick, D.C., Chen, M.-J.: What is Strategic Management, Really? Inductive Derivation of a Consensus Definition of the Field. *Strat. Mgmt. J.* 28, 935–955 (2007)
7. Al-Debei, M.M., Avison, D.: Developing a Unified Framework of the Business Model Concept. *European Journal of Information Systems* 19, 359–376 (2010)
8. Krippendorff, K.: *Content Analysis: An Introduction to its Methodology*. Sage Publications, Thousand Oaks (2004)
9. Abbasi, A., Chen, H.: CyberGate: A Design Framework and System for Text Analysis of Computer-mediated Communication. *MIS Quarterly* 32, 811–837 (2008)
10. Angelmar, R., Stern, L.W.: Development of a Content Analytic System for Analysis of Bargaining Communication in Marketing. *Journal of Marketing Research* 15, 93–102 (1978)
11. Steininger, D.M., Huntgeburth, J.C., Veit, D.J.: Conceptualizing Business Models for Competitive Advantage Research by Integrating the Resource and Market-Based Views. In: *AMCIS 2011 Proceedings*, Detroit, USA (2011)
12. Eisenhardt, K.M.: Building Theories from Case Study Research. *Academy of Management Review* 14, 532–550 (1989)
13. Yin, R.K.: *Case Study Research: Design and Methods*. Sage Publications, Inc, Thousand Oaks (2009)
14. Myers, M.D.: *Qualitative Research in Business & Management*. Sage Publications Ltd. (2008)
15. Insch, G.S., Moore, J.E., Murphy, L.D.: Content Analysis in Leadership Research: Examples, Procedures, and Suggestions for Future Use. *The Leadership Quarterly* 8 (1), 1–25 (1997)
16. Saris-Gallhofer, I.N., Saris, W.E., Morton, E.L.: A Validation Study of Holsti's Content Analysis Procedure. *Quality and Quantity* 12 (2), 131–145 (1978)
17. Kassirjian, H.: Content Analysis in Consumer Research. *Journal of Consumer Research* 4, 8–18 (1977)
18. Mayring, P.: *Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken* (in German). Beltz, Weinheim (2002)
19. Mayring, P.: *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (in German). Beltz, Weinheim (2008)
20. Eisenhardt, K.M., Bourgeois, L.J.: Politics of Strategic Decision Making in High-Velocity Environments: Toward a Midrange Theory. *Academy of Management Journal* 31, 737–770 (1988)
21. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research* 2, 192–222 (1991)
22. Scott, W.: Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly* 19, 321–325 (1955)
23. Mayring, P.: Qualitative Content Analysis. *Forum: Qualitative Social Research* 1, 1–10 (2000)
24. Moore, J.E.: One Road to Turnover: An Examination of Work Exhaustion in Technology Professionals. *MIS Quarterly* 24, 141–168 (2000)
25. Holsti, O.R.: *Content Analysis for the Social Sciences and Humanities*. Addison-Wesley, Reading (1969)
26. Charniak, E.: Tree-Bank Grammars. In: *Proceedings of the National Conference on Artificial Intelligence*, pp. 1031–1036 (1996)

27. Pavlou, P.A., Dimoka, A.: The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation. *Information Systems Research* 17, 392–414 (2006)
28. Frueh, W.: *Inhaltsanalyse: Theorie und Praxis* (in German). UVK, Konstanz (2007)
29. Hayes, A.F., Krippendorff, K.: Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1, 77–89 (2007)
30. Scott, W.: Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly* 19, 321–325 (1955)
31. Fleiss, J.L.: Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin* 76, 378–382 (1971)
32. Cohen, J.: A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 37–46 (1960)
33. Aldenderfer, M., Blashfield, R.: *Cluster Analysis*. Beverly Hills: Sage University Paper (1984)
34. Fowlkes, E.B., Gnanadesikan, R., Kettenring, J.R.: Variable Selection in Clustering. *Journal of Classification* 5, 205–228 (1988)
35. Kaufman, L., Rousseeuw, P.J.: *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Online Library (1990)
36. Howard, N., Harris, B.: A Hierarchical Grouping Routine, IBM 360/65 Fortran IV program. University of Pennsylvania Computer Center (1966)
37. Punj, G., Stewart, D.W.: Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research* 20, 134–148 (1983)
38. Ward, J.H.: Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244 (1963)
39. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 319–340 (1989)
40. Krippendorff, K.: Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research* 30, 411–433 (2004)
41. Thomas, H., Venkatraman, N.: Research on Strategic Groups: Progress and Prognosis. *Journal of Management Studies* 25, 537–555 (1988)
42. Ketchen, D.J., Shook, C.L.: The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strategic Management Journal* 17, 441–458 (1996)
43. Hair, J.F., Black, B., Babin, B.: *Multivariate Data Analysis*. Prentice Hall (2005)
44. Lambert, S.: Do We Need a “Real” Taxonomy of e-Business Models?. In: *ACIS 2006 Proceedings, 17th Australasian Conference on Information Systems* (2006)

Qualitative Comparative Analysis in Information Systems and Wirtschaftsinformatik

Roy Wendler¹, Helena Bukvova¹, and Sandra Leupold²

¹ Technische Universität Dresden, Dresden, Germany
{roy.wendler, helena.bukvova}@tu-dresden.de

² T-Systems Multimedia Solutions GmbH, Dresden, Germany
Sandra.Leupold@t-systems.com

Abstract. The application of scientific methods is an essential element when conducting research. They ensure reproducible results and improve the overall quality of the research projects. The aim of this paper is to introduce a method called Qualitative Comparative Analysis, which is currently nearly unrecognized within our discipline. Being neither a pure quantitative nor qualitative method, it yields potential benefits arising from both research streams. It accommodates answering research questions that simultaneously demand a deep understanding of complex relationships and also require the analysis of more than just a few single cases. The paper gives an introduction to this method and demonstrates its usefulness on the basis of two recently carried out research projects. In the end, the Qualitative Comparative Analysis proves to be a valuable addition to the canon of research methods and enriches the applied character of our discipline by contributing to the improvement of both rigor and relevance.

Keywords: Qualitative Comparative Analysis (QCA), research methods, case study research, design-oriented research, empirical research

1 Introduction

This article introduces a research method called Qualitative Comparative Analysis (QCA) [1], a method which is nearly unrecognized within our discipline to date and possesses many potential benefits for design-oriented as well as empirical research projects. Furthermore, the method fills in current gaps within and between different existent research paradigms.

Within the last years it became clear that there exist two main research paradigms in Information Systems (IS): the behavioral or positivist paradigm, which focuses on the creation and testing of theories, and the design-oriented paradigm, which emphasizes the development of IT-related artifacts and the design of the organizational environment [2–4]. Today, the behavioral approach clearly dominates Anglo-American research (respectively IS), whereas the design-oriented approaches are prevalent in European research, especially in the German-speaking “Wirtschaftsinformatik” (WI). Despite the ongoing development of these research streams, the last years showed some dis-

satisfaction with both [4-5]. On the one hand, the behavioral approach faces increasing discussion about its identity, the application of qualitative vs. quantitative methods, and the relevance and limitations of the prevailing research methods [4], [6-7]. On the other hand, the design-oriented approach struggles with the availability of a rigorous set of research methods and a lack of international recognition [4], [8].

Consequently, a call for approaches arises which may overcome the current limitations in the research fields of both IS and WI and bridge the gaps between existing paradigms. Frank claims that “due to the diversity of research topics and objectives in ISR [Information Systems Research], mono-paradigmatic research is not sufficient” [4]. Indeed, there seems to be huge potential within the convergence of the behavioral and the design-oriented approach. Lyytinen & King stated, with regard to the legitimacy of IS research: “Theory has value *only* in reference to praxis” [7]. However, limiting attention to the IT artifact alone would be misleading, too, because “artifacts never deliver value in their own right. They are complementary assets in production, and their value cannot be understood without the context of their application” [7].

One essential element to ensure (scientific) rigor as well as (practical) relevance is the selection of suitable research methods that take into account the characteristics of the research questions in IS/WI [9]. Different research questions ask for different research methods, and the chosen method should always be determined by the questions asked. However, if the dichotomy of the two named paradigms remains, there is the danger that scientists in IS/WI ask their questions in a way that fits predefined methods instead. This would be a clear threat to the improvement of both rigor and relevance.

Such progress could already be observed in Management Science and Sociology. Researchers were stuck within a dispute between qualitative case-oriented (more practical) approaches and quantitative variable-oriented (more theoretical) approaches. It has been discovered that the prevalent methods have limitations in answering specific research questions, despite focusing on measurement, sampling, analysis, and so on [10]. In the end, the quality of research suffered, because “often, the desire to use these [multivariate statistical] techniques shapes the way social scientists ask their questions” [1].

Hence, to support the call for a pluralistic pool of scientific methods which are equally acknowledged among the research community of IS/WI [4], [7-8], [11], we are introducing QCA as a potential method. To support our argumentation, the paper is structured as follows: Section 2 gives a short introduction to QCA and explains basic concepts on the basis of a simple example. Section 3 discusses the potential benefits of QCA in IS/WI and highlights possible usage in design-oriented as well as behavioral research approaches. The argumentation is underlined by two recently conducted research projects. Finally, section 4 points out possible limitations and summarizes the article.

2 A Short Introduction to Qualitative Comparative Analysis

2.1 Origin and Classification

Qualitative Comparative Analysis is a research method whose roots date back to the mid of the 19th century. In this period John Stuart Mill established the fundamentals in

his essay “Of the Four Methods of Experimental Inquiry”, especially the “Method of Agreement” and the “Method of Difference”, which deal with a systematic comparison of cases to search for common causal relationships [12-13].

QCA in its current form emerged from the work of Charles C. Ragin in the 1980s [1]. QCA has been known in Comparative Politics for a time, but soon it was recognized as a potential method for Sociology, too [14-15]. Today, QCA has been applied also in Management Science, Business and Economics, Health Research, Legal Studies, and International Relations, among others. An overview of current application areas, software tools, and further references can be found at the website compass.org (COMPARative Methods for Systematic cross-caSe analySis) [16].

In fact, the term “qualitative” may be misleading. The method can be classified as neither a qualitative nor a quantitative approach, but shares key elements of both. Strictly speaking, QCA can generally be seen as a more quantitative method which operates on qualitatively coded data and allows a systematic comparison of cases [17]. Therefore, in recent years one can observe a tendency to subsume QCA and related methods, like multi-value QCA (*mvQCA*) and fuzzy set QCA (*fsQCA*), into a family of so called “Configurational Comparative Methods” [18].

This ambiguity reflects itself also from an epistemological point of view. As a method, QCA cannot be categorized to one distinct paradigm. Although the logical foundation of QCA (see section 2.2) conveys a rather positivist character of the method, it can be rarely established that way [12-13]. Furthermore, the results obtained by utilizing QCA heavily rely on the interpretation and the decisions of the researcher. These are characteristics of an interpretative paradigm [18]. However, supporting Mingers’ view that methods may be used “critically and knowledgeably” within different paradigmatic assumptions [11], QCA does not exclude itself from one or the other paradigm. It is flexibly applicable depending on the researcher’s goals and individual positioning.

2.2 Basic Concepts

This section aims at introducing some basic concepts necessary for conducting a QCA in a very brief manner. Due to restrictions in space, we cannot give a comprehensive description. Therefore, interested readers may be referred to the publications of Ragin [1], Schneider & Wagemann [14], and Rihoux & De Meur [19] which serve as basis and orientation for the following explanations within this section, too.

Furthermore, the given explanations, mainly taken and modified from [1], are accompanied by an example which we kept simple for illustration purposes. Let a research project deal with the success (S) of reorganization projects. The researcher analyzes the effect of the three variables: Top Management Support (T); Early Employee Involvement (E); and Investments Spent (I), over eight independent cases. In QCA terms, the variables are called *Conditions*, whereas the success is called *Outcome*.

First of all, QCA builds upon the use of Boolean Algebra. That means variables have to be coded dichotomously into 0 (false, absent) or 1 (true, present), and cases are represented as sets according to set theory. For our example, the outcome would be 0 for an unsuccessful project (written lower case: s) and 1 for a successful project (written upper

case: S). Analogously, the conditions are coded into Top Management Support was absent (t) or present (T), and Early Employee Involvement was absent (e) or present (E). The third condition, Investments Spent, is an interval-scaled variable. Therefore the researcher needs to identify what constitutes a high (I) or low (i) investment.

The collected data is then presented in a *Truth Table*. A hypothetical truth table for our example is given in Table 1. Every row represents a possible *Configuration*. That means every theoretically possible combination of conditions (excluding outcome). For our example, there exist $2^3 = 8$ rows at maximum. Our example is chosen in such a way that every case represents a possibly configuration. Of course, this is unlikely to happen. So there may be less than 8, if some configurations were not observed, and simultaneously, one row may represent more than one case. Both situations would not affect the application of the standard QCA procedure.

Table 1. Hypothetical truth table

Configura- tion ID	Conditions			Outcome
	Top Management Support (T)	Early Employee Involvement (E)	Investments spent (I)	Project Success (S)
1	0	0	0	0
2	0	0	1	0
3	0	1	0	1
4	1	0	0	0
5	0	1	1	0
6	1	0	1	1
7	1	1	0	1
8	1	1	1	1

For instance, row (configuration) 5 is characterized as follows: The project had no top management support (0), but an early employee involvement took place (1). A high amount of investments was spent (1), but the project was not successful in the end (0).

The next step is to find out which combination of conditions leads to successful or unsuccessful projects. For that, QCA utilizes combinatorial logic and Boolean minimization. Therefore, it is important to understand Boolean addition and multiplication. A Boolean sum represents the logical OR, meaning at least one of the parts of the sum has to be true. A Boolean product, however, is a combination of conditions and is equivalent to the logical AND. Every configuration now can be written in a Boolean product, whereby uppercase letters indicate presence (1), and lowercase letters indicate absence (0). For instance, configuration 5 can be written as “tEI”, whereas configuration 7 would read “TEi”. Now, we are already able to write down a primitive Boolean term for both outcomes success (S) and no success (s) in the form of Boolean sums combining the corresponding configurations that are given in Table 2.

Table 2. Primitive Boolean sums-of-products

Sum-of-products	Corresponding configurations
$s = tei + teI + tEI + Tei$	unsuccessful projects 1, 2, 4, and 5
$S = tEi + TeI + TEi + TEI$	successful projects 3, 6, 7, and 8

However, such a representation is not satisfying. So, QCA continues with a pairwise comparison of every configuration with the aim of minimizing complexity and simul-

taneously maintaining causality. The fundamental rule of minimization states: “If two Boolean expressions differ in only one causal condition yet produce the same outcome, then the causal condition that distinguishes the two expressions can be considered irrelevant and can be removed to create a simpler, combined expression.” [1]. Looking, for instance, at configurations 7 and 8 in Table 1, both differ only in the investments spent. Nevertheless, both are successful. Therefore, the investments spent are irrelevant for this (and only this!) pair of configurations. They can be reduced to:

$$S = TE_i + TEI = TE \quad (1)$$

This procedure is carried on for every pairwise comparison and is repeated with the reduced configurations (like TE in the example above) until no further minimization is achievable. However, only those configurations are compared to each other, which had the same outcome. The analysis of successful and unsuccessful projects results in two independent runs of QCA. The pairwise comparison of all primitive expressions for project success (S) in our example results in the following minimized Boolean expression that cannot be reduced any further:

$$S = E_i + TI + TE \quad (2)$$

The last step in QCA is the deduction of *Prime Implicants*. The Boolean concept of implication states that one expression implies another, if the second is a subset of the first. The goal of this step is to produce a final expression with a logically minimal number of prime implicants. For this task, the *Prime Implicant Chart*, illustrated in Table 3, is a helpful tool. It maps the minimized against the original primitive expressions, whereby an X states that the minimized expression implies the primitive one.

Table 3. Prime implicant chart

	tEi	TeI	TEi	TEI
Ei	X		X	
TI		X		X
TE			X	X

The requirement for a logical minimal expression is that every column of the chart has at least one X. This means that every primitive expression is covered by a minimized one. As easily becomes clear, the third row (TE) is redundant and can be omitted without changing the result. Another advantage of this chart is that primitive expressions not covered by any of the minimized expressions become visible. They have to be included as prime implicants, too (this is not the case in our example). The final solution, then, for the project’s success is:

$$S = E_i + TI \quad (3)$$

This is read as follows: Projects are either successful when employees are involved early (E) AND the spent investments are low (i) OR when top management support was given (T) AND the spent investments are high (I). It becomes clear that QCA keeps causality and analyzes the effect of conditions, always in observance of the

presence or absence of other causally relevant conditions. The researcher may now interpret this result further.

Additionally, Boolean minimization allows distinguishing between *Necessary* and *Sufficient Conditions* for specific outcomes, which in turn allows a deeper interpretation. Further important concepts are *Logical Remainders*, configurations that are logically conceivable but missing in the data set, and *Contradictions*, cases with identical configurations but different outcomes. The interested reader may refer to basic literature (for instance [1], [14], [20]) for further explanation of these issues.

3 QCA in IS and WI: Potential and Examples

As a scientific discipline, IS/WI is rooted between Business Administration and Economics, as a part of Social Sciences on the one side, and Computer Science with Engineering foundations on the other side [21-22]; hence, it continually seeks to unite two different research traditions. Furthermore, it inherits an applied character from both disciplines, and thus the need to make its findings available to practice [8]. This constellation poses a considerable challenge to establishing generally accepted research approaches. This conflict has been made visible by the discussion about the role of behavioral and design-oriented research methods in IS/WI (see section 1) [3], [8]. Within WI, researchers are now leaning towards finding a consensual middle way between relevant, application oriented research and rigorous, systematic research procedures, strongly soliciting the design-oriented approach [8]. It is our aim in this article to support this development by presenting QCA as a suitable method for increasing the rigor, without an unnecessary detachment from the practice or reduction of complex practical cases.

The nature of QCA as a systematic and configurational comparative method yields a number of benefits for researchers of IS/WI. First of all, the QCA stands outside the two main fields of qualitative and quantitative methods but combines some of the key strengths of both [1], [23-24]. Therefore, QCA is suitable for application in our discipline for two reasons: it supports the systematic comparison of multiple cases with limited information loss and it helps to establish causal relationships based on factor configuration while maintaining case-inherent complexity [24]. Due to the applied character of the IS/WI discipline, combined with the frequently used design-oriented approach, researchers often deal with complex case studies. While this approach allows an in-depth understanding of the empirical context, systematic and reproducible comparison and generalization of findings are often difficult [25].

By using the QCA instead, context and findings of the case studies can be simplified into sets of different outcomes and hence made comparable. The calibration of the data – unlike simple quantification – allows the retention of the context through set description and the definition of set membership. Furthermore, the configuration, as means of expressing causal relationships, results in meaningful descriptions of possible case variations.

While quantitative methods have been used to describe causal relationships in IS, their interpretation and application in practice appears difficult. Furthermore, these

methods imply the isolation of factors (variables), and therefore suffer limitations regarding the complexity of underlying theoretical models [1], [6], [15], [24]. QCA does not view causes in isolation but always within the context of other relevant conditions. Every case is analyzed as a whole regarding absence or presence of dependent and related conditions [1]. Moreover, combinations of factors are easier to interpret and provide suitable models as a foundation for analysis and configuration of cases in practice.

In addition, the QCA offers some new possibilities in the way that the variables are determined. Variables are not seen as pure quantitative data rather than interpreted as qualitative construct. Although the dichotomization yields some limitations in itself, this treatment allows a direct coding of a qualitative construct instead of technically driven operationalization via dummy variables [17].

When compared to qualitative methods, QCA can also offer a valid alternative. While it cannot provide the same deep data emergence as for example Grounded Theory [26], its structured process helps the researchers explicate their decisions and reasoning. The reliance on set-theory rather than statistical methods emulates the causal relations from the real world and thus helps to maintain the applicability of the results. Qualitative research in IS/WI is often based on single or multiple case studies [25], where the QCA can be suitable for abstracting and comparing the findings.

In summary, both behavioral as well as design-oriented research in IS and WI can benefit from the application of QCA. The method can be applied in behavioral studies where the research is based on a moderate number of case studies and is concerned with the description of causal relationships and the identification of patterns. In design-oriented research, QCA can be helpful to support different research steps to deduce artifacts and useful IT solutions: during the analysis and exploration of the problem as well as during evaluation of the designed artifact.

QCA expands the set of useable methods apart from surveys, single case studies, interviews, pilots, simulations, etc. Therefore, it offers a possibility that does not force the researcher to decide between understanding complexity and gaining generalized insights rather than allows asking research questions combining both aspects [1], [27].

In the following section, we present and briefly discuss two examples of the application of QCA in IS/WI research. The first example is concerned with the identification of success factors in agile project management and the second with the search-behavior on online profiles belonging to scientists. Both examples are meant to show possible uses of the method in two different settings – not to serve as references of proper implementation of the method. On the contrary, we use the experience from the two research projects to point out and discuss difficulties and risks in the use of QCA in IS/WI research.

3.1 Identification of Success Factors in Agile Project Management

The analysis shown in this subsection was accomplished within a research project in cooperation with an international communication and web 2.0 company. The company's main focus lies in developing and providing software solutions. Since 2008, project managers in the company have used agile methods like Scrum to accomplish software and organizational projects. Originally, the purpose was to save time and costs due to increased transparency and shorter development cycles. However, it emerged that most

Scrum projects were not more but less successful than projects conducted with traditional methods. Interviews indicated that missing basic Scrum conditions could be the reason for the lack of success. Therefore, a total of 19 Scrum projects were analyzed using QCA. The aim of the study was to determine on which basic Scrum conditions, or combinations of these, the success of Scrum projects depended.

The most agile project management methods are based on the “Manifesto for agile Software Development” from 2001. Agile methods contain four principles: they are based on individuals and interactions; the relationship with the customer; an efficient solution; and a fast reaction to changed requirements [28]. Therefore, the purpose of agile project management is to develop pieces of the solution in short time, under permanent contact to the customer and changing requirements. Within this environment, Scrum was developed in 2001. To ensure short development times, Scrum uses the term “Sprint”. A Sprint is usually two to four weeks long and begins with a discussion between the Scrum Team and the requirements manager (Product Owner). This role is responsible for the customer communication during a Sprint, in case of changed requirements. Simultaneously, the Scrum Team develops customer requirements in self-controlled environments and presents the solution to the Product Owner for approval at the end of a Sprint. The Scrum Master is responsible for the motivation of the Scrum Team only by using specialized motivation methods. Hence, no disciplinary methods are necessary. Additionally, Scrum requires several basic conditions to unroll its advantages [29–31]. During a theoretical analysis, seven basic Scrum conditions were discovered: Training and Coaching, Internal Communication, Allocation of Tasks, Role Understanding, Problem Solving, Requirements Management, and Meetings.

Due to the fact that, in summary, 19 projects could be analyzed, the csQCA proved suitable regarding sample size and the research aim. A questionnaire was created to collect data. It was based on the identified Scrum basic conditions and the success of the project, which were operationalized by several questions. Afterwards, the Scrum Master, one project employee, and the customer of every project were given the questionnaire. Every answer in the questionnaire received a score which depended on how good the answer matched with the Scrum theory. Thereafter, the sum of the achieved score points were divided through the possible maximum score. The outcome was a “degree of fulfillment” of the basic conditions and the success per project.

Table 4 shows an example for this procedure. The basic condition of Problem Solving contained three questions. Every project could gain 0, 0.5, or 1 point per question. The maximum score was 3 points. Therefore, project 1 for instance, achieved a degree of fulfillment of 83.33%.

Table 4. Example for data collection and calculation of degree of fulfillment

Project	Q. 1	Q. 2	Q. 3	Achieved score	Maximum score	Degree of fulfillment
1	1	0.5	1	2.5	3	83.33%
2	0	0.5	0	0.5	3	16.67%
3	0	1	0	1	3	33.33%

Afterwards, it was necessary to dichotomize the data regarding the calculated degrees of fulfillment. This was done with the help of thresholds for every basic condition as

well as the success. The exact threshold was determined according to two restrictions. On the one hand, it was important not to separate degrees of fulfillment which were close together. On the other hand, both produced sub-groups should be of similar size. In some cases, it was not possible to accommodate these restrictions. Hence, a single case analysis was accomplished to create clearness of doubtful project classifications. The result of this step was a truth table which was analyzed by the computer program TOSMANA (see [16] for further information).

The analysis delivered the following basic conditions which were responsible for positive project success: the presence of Internal Communication, Training and Coaching, and the Task Allocation. The results for negative project success were: the absence of Internal Communication, Problem Solving, Task Allocation, Requirements Management, and Role Understanding. The results again underline the qualitative nature of the QCA approach, because they show that the conditions for unsuccessful projects are not simply the opposite of the ones affecting successful projects. Both analysis were conducted independently, and thus together they deliver a more detailed picture. For some projects it was necessary to analyze them separately. For example, one project had the basic condition "Internal Communication" unfulfilled (absence), despite being successful, and thus was investigated and evaluated separately in the form of a single case analysis. Afterwards, the identified factors were interpreted further, because the company needed deep information for every positive or negative success factor. Based on these results, recommendations to improve future projects based on Scrum were deduced and given to the company.

The QCA proved itself to be a very suitable method to investigate the success of Scrum projects because of the following reasons: First, due to time and resource constraints, it was not possible to investigate and compare every one of the 19 projects using single case analysis. The QCA, however, offered a standardized procedure to cope with this amount of cases. In addition, it was still possible to investigate conditions in more detail for selected projects. This was necessary with regard to content. One result of the analysis was that a successful project had to show presence of Internal Communication OR a number of other AND-connected conditions. But as mentioned above, one single project did not fulfill this rule, despite being successful. The QCA allowed identifying this project by treating every case as one whole unit and conducting an additional single case analysis. If regression analysis or a similar multivariate method had been used, this additional insight would not have been gained.

At the same time, this example illustrates another important issue: in every step of QCA, the researcher has a relatively wide range of intervening actions (selection of cases and variables, dichotomization or categorization of data, interpretation of results). It is crucial that the method is not applied in a mechanical way because this may lead to misinterpretation of findings [1], [23]. This underlines the qualitative side of the method, but inherits the same pros and cons of other qualitative methods like transparency of decisions and reproducibility of findings. Furthermore, the applicability within design-oriented research can be illustrated through this example. The developed guideline, with recommendations for further Scrum-based projects, can be seen as artifact [32], and the accompanying research process went through the proposed phases of analysis, design, evaluation, and diffusion [8]. The QCA was utilized

especially in the analysis phase to create a substantial basis for deducing the recommendation guidelines for the company. The use of QCA within the evaluation phase is conceivable, too, but was not carried out through this project.

3.2 Analysis of Search-Behavior Patterns on Scientists' Online Profiles

The study was carried out in context of a larger research project on the self-presentation of scientists on the Internet [33]. Scientists from different disciplines increasingly use the Internet to present themselves. Besides the traditional institution or private HTML webpages, scientists can now also use Web 2.0 tools, such as social networking services, blogs, or microblogs [34-35]. The growing number of self-presentation options on the Web has led to discussions of the influence of online self-presentation on the reputation of scientists [36]. However, while the self-presentation of scientists in online profiles has been studied, there is little research on the impact and use of the profiles by others. Thus, the purpose of the study was to analyze the search behavior of European scientists on the Internet profiles of their peers, providing a foundation for the potential of online self-presentation for scientists. It was not an aim of the study to provide representative statements about the general behavior of all European scientists, but rather to detect behavioral patterns. The search behavior was thus analyzed using QCA. Furthermore, quantitative association measures were used to provide a direct comparison.

The study was based on the theory of social networks, assuming that individuals are connected by social ties of different strengths [37-38]. These ties can be strong, with close relationships and intensive exchange; weak, with lesser exchange intensity and little shared resources; latent, founded on organizational structures rather than social contact; or fully absent [38-39]. According to Haythornthwaite [38], the acceptance and adoption of technology for the purpose of communication depends on the strength of the existing tie between two communicators. Hence the study viewed search patterns in connection to existing social ties. To collect data, an online questionnaire based on the critical incident technique [40] was sent to a stratified clustered sample of 1008 European scientists, delivering 123 usable answers. The study looked for search patterns in several different areas, but as it is not our aim to describe the entire study [41] here, we will demonstrate the proceeding on one of them: the access way to the online profiles.

The analysis of the data was carried out using a csQCA [1], [42], in combination with quantitative association measures. The collected data were first coded for quantitative analysis, coding tie strength as an ordinal variable and labeling the different access ways as a nominal variable. These data were then used to generate measures of association. The existence of a relationship of the tie strength and the access ways in general, as well as each access way separately, were then determined using quantitative association measures (Decady-Thomas corrected chi-squared test [43-44], Perason's chi-squared test, Gamma, and Cramer's V). However, this procedure only tested the general existence of a relationship between tie strength and access ways and described associations between single access ways and tie strength. To derive more complex access-way patterns, a csQCA was used. fsQCA and mvQCA were not applied, because tie strength was interpreted as the existence of four distinct outcomes

rather than of increments, where a distinct combination of access ways (condition) was sought for each tie-strength set. The data was recoded for the csQCA: tie strength was interpreted as four separate sets (StrongTie, WeakTie, LatentTie, and AbsentTie), and a separate set was created for each access way. The data were then calibrated according to their membership in each set. To depict the relationship between the tie strength and the access ways, the combinations of access-way set-memberships were viewed as potential predictors of the tie-strength sets:

$$\text{TieSet} = f(\text{AccessWay}_1, \dots, \text{AccessWay}_n) \quad (4)$$

Table 5. Resulting access-way patterns

		Direct	Looked up page	Looked up person	Search for person	Search for topic	Used link	Other	Raw coverage	Unique coverage	Consistency	Solution coverage	Solution consistency
Latent	b1	X	X	O	X	X	X	O	0,01	0,01	1,00	0,62	0,84
		X	O	O	X		X	X	0,19	0,00	0,89		

The QCA was carried out with the fs/QCA software (see [16] for further information). The derived combinations for each tie set were grouped according to common features, identifying core and periphery elements of the patterns [45]. Table 5 shows as an example two (out of seven) patterns describing the access-way combination for scientists with latent ties. The conditions, i.e., the used access ways, that are present in the combination are marked with an X, and the conditions that are absent, i.e., the unused access ways, are marked with an O. One condition is unmarked, as its presence or absence has no influence on the outcome. The two patterns in Table 5 were grouped as being similar, where the conditions describing their similarities were considered core conditions (marked bold), while the others were considered peripheral. The derived combinations, although not describing clear, general patterns, showed visible trends in the way scientists access the online profiles of their peers [41].

Unlike in the previous example, this study can be seen as a part of a theory-building research design. It shows that the QCA can prove beneficial in areas that cannot be easily satisfied with traditional quantitative methods. The identification of patterns and configurations plays an important role in IS research and practice, but the factors of interest often defy meaningful quantification. Hence the available quantitative analyses offer but a simplified picture, such as in this study. The QCA offered more complex and holistic results. At the same time, the results of the study also show pitfalls of the QCA application: this high number of conditions allowed only recognition of trends. In order to obtain better details, the number of responses for each tie set would have had to be higher. At the same time, the increase in the number of cases in itself does not necessarily guarantee better results, as the data may not cover the possible combinations. Selective sampling is helpful in order to cover different combination possibilities.

4 Conclusion

Our aim in this article was to present the QCA method and discuss its potential for IS and WI research. To this end, we have briefly introduced the method and its application, discussed its usability in IS/WI, and presented two studies that have actively applied QCA: one related to design research process and one related to theory building research process. From these, we have concluded that the use of QCA in IS/WI appears suitable and offers a number of advantages. These include a structured method for the comparison of several case studies, simplification and variable calibration without full context loss, focus on causal relationships, and the derivation of understandable and applicable configurations. The QCA also offers new, creative ways of doing research, as it is based on different principles than traditional quantitative methods. However, there are limitations and potential pitfalls that have to be taken into account when using QCA. These will be discussed in the following.

Firstly, the move from quantitative coding towards an interpretative calibration is helpful when dealing with qualitative data, but it relies strongly on the researchers' interpretation. While some variables can be easily calibrated, as they are already binary or nominal, in other cases (e.g., the project section 3.1), the researcher must decide upon suitable threshold values. In IS/WI, this can be of advantage: as an applied discipline IS/WI research often studies in-depth cases and the researchers thus possess good case knowledge necessary for calibration. At the same time, the focus on data calibration and use of a standardized data manipulation method can lead to the loss of the "big picture". While well structured, QCA relies on qualitative means and requires researchers' attention to the actual meaning of the data. Losing the overview of the context for the benefit of high granularity and standardization of the sets can lead to overcomplicated results with essentially trivial outcomes.

Secondly, although QCA uses calculation to assess the causal relationships, it is not well suited for exploratory analysis based on quantitative measures (i.e., hunting for strong correlations), as it is sensitive to both coding and measurement errors [46]. This aspect makes the method dependent on a sound theoretical foundation and good case-level knowledge, as these determine the adequate specification of the causal conditions for inclusion in the truth table. Unlike conventional quantitative approaches, the QCA does not offer a simple mechanism for excluding 'nonsignificant' variables. Typically, most if not all of the causal conditions that are considered while constructing the truth table appear in the statement of causal combinations that conclude as a result [27], [47]. While a disregard of this explicit reliance on researchers' knowledge and interpretation can lead to mistakes, in IS/WI it may actually help bridging the mentioned gap between behavioral and design approach. QCA explicitly demands the combination of existing theories (coming from behavioral research) with in-depth case studies (in design research). The method can thus help to generate results both grounded in empirical data as well as with a sound theoretical foundation.

Thirdly, the results do not always paint a clear picture. The derived configurations can exhibit traits such as of equifinality (different configurations leading to the same solutions) and natural permutations (interaction stable core conditions and changing peripheral conditions) [48]. As QCA is not a quantitative method, it does not offer

definite analytical measures, and sometimes even disregards traditional measures (e.g., frequency). Similarly, QCA is not a tool designed for assessing net effects, as the underlying logic is that causes combine, not that they are in competition with each other. QCA is still in development, and further measures may appear, but provided the character of the method, the effort connected to evaluation and generalization of the results must lie largely with the researchers. Hence, the real test of the value of an application of QCA is not a summary statistic but researchers' assessments of how well the results help them understand their cases. Again, this explicit reliance on the applicability of the results appears particularly relevant for IS/WI research.

In summary, QCA is a well-structured but highly interpretative method, suitable for dealing with a limited number of complex cases. It focuses on the description of causal relationships using a configurational approach and Boolean algebra. We believe that QCA can be used in IS and WI in areas where it is necessary to generate descriptions and factor configurations from a limited number of cases. As such, it offers potential of increasing the rigor of empirical research without sacrificing contextual richness, and the practical applicability of the findings.

References

1. Ragin, C.C.: *The Comparative Method - Moving Beyond Qualitative and Quantitative Strategies*. University of California Press, Berkeley (1987)
2. March, S.T., Smith, G.F.: Design and natural science research on information technology. *Decision Support Systems* 15, 251–266 (1995)
3. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. *MIS Quarterly* 28, 75–105 (2004)
4. Frank, U.: *Towards a Pluralistic Conception of Research Methods in Information Systems Research*. ICB-Research Report 7, University Duisburg-Essen (2006)
5. Niehaves, B.: Epistemological Perspectives on Multi-Method Information Systems Research. In: *ECIS 2005 Proceedings*, Paper 120 (2005)
6. Fichman, R.G.: Going Beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods. *Journal of the AIS* 5, 314–355 (2004)
7. Lyytinen, K., King, J.L.: Nothing At The Center?: Academic Legitimacy in the Information Systems Field. *Journal of the AIS* 5, 220–246 (2004)
8. Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Kremar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J.: Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. In: Österle, H., Winter, R., Brenner, W. (eds.): *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*. Infowerk, Nürnberg (2010)
9. Frank, U.: Zur methodischen Fundierung der Forschung in der Wirtschaftsinformatik. In: Österle, H., Winter, R., Brenner, W. (eds.): *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*. Infowerk, Nürnberg (2010)
10. Ketchen, D.J., Boyd, B.K., Bergh, D.D.: Research Methodology in Strategic Management: Past Accomplishments and Future Challenges. *Organizational Research Methods* 11, 643–658 (2007)
11. Mingers, J.: Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research* 12, 240–259 (2001)

12. Mill, J.S.: A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. In: Robson, J.M. (ed.): *Collected Works of John Stuart Mill*. The University of Toronto Press, Toronto (2007)
13. Berg-Schlosser, D., De Meur, G., Rihoux, B., Ragin, C.C.: Qualitative Comparative Analysis (QCA) as an Approach. In: Rihoux, B., Ragin, C.C. (eds.): *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. SAGE Publications, Thousand Oaks, CA (2009)
14. Schneider, C.Q., Wagemann, C.: *Qualitative Comparative Analysis (QCA) und Fuzzy Sets: Ein Lehrbuch für Anwender und jene, die es werden wollen*. Barbara Budrich, Opladen (2007)
15. Greckhamer, T., Misangyi, V.F., Elms, H., Lacey, R.: Using Qualitative Comparative Analysis in Strategic Management Research: An Examination of Combinations of Industry, Corporate, and Business-Unit Effects. *Organizational Research Methods* 11, 695–726 (2007)
16. compasss, <http://www.compasss.org/>
17. Cronqvist, L.: Konfigurationelle Analyse mit Multi-Value QCA als Methode der Vergleichenden Politikwissenschaft mit einem Fallbeispiel aus der Vergleichenden Parteienforschung (Erfolg Grüner Parteien in den achtziger Jahren) (2007)
18. Niehaves, B., Stahl, B.C.: Criticality, Epistemology, and Behaviour vs. Design - Information Systems Research Across Different Sets of Paradigms. In: *ECIS 2006 Proceedings*, pp. 1–12 (2006)
19. Rihoux, B., De Meur, G.: Crisp-Set Qualitative Comparative Analysis (csQCA). In: Rihoux, B., Ragin, C.C. (eds.): *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. SAGE Publications, Thousand Oaks, CA (2009)
20. Rihoux, B., Ragin, C.C. (eds.): *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. SAGE Publications, Thousand Oaks, CA (2009)
21. Hess, T.: Erkenntnisgegenstand der (gestaltungsorientierten) Wirtschaftsinformatik. In: Österle, H., Winter, R., and Brenner, W. (eds.): *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*. Infowerk (2010)
22. Heinrich, L.J.: *Wirtschaftsinformatik: Einführung und Grundlegung*. Springer, Heidelberg (2004)
23. Rihoux, B.: Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research. *International Sociology* 21, 679–706 (2006)
24. Rihoux, B.: Bridging the Gap between the Qualitative and Quantitative Worlds? A Retrospective and Prospective View on Qualitative Comparative Analysis. *Field Methods* 15, 351–365 (2003)
25. Yin, R.K.: *Case Study Research: Design and Methods*. Sage Publications, Thousand Oaks, CA (2009)
26. Glaser, B.G., Strauss, A.: *The Discovery of Grounded Theory: Strategies for qualitative research*. Sociology Press, Mill Valley, CA, USA (1967)
27. Ragin, C.C., Shulman, D., Weinberg, A., Gran, B.: Complexity, Generality, and Qualitative Comparative Analysis. *Field Methods* 15 (4), 323–340 (2003)
28. Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J., Thomas, D.: Manifesto for Agile Software Development, <http://agilemanifesto.org/>

29. Bunning, R.: Kicking ScrumBut, <http://www.scrumalliance.org/resources/1122>
30. Wirdemann, R.: Scrum mit User Stories. Carl Hanser, München (2009)
31. Gloger, B.: Scrum: Produkte zuverlässig und schnell entwickeln. Carl Hanser, München (2008)
32. Becker, J.: Prozess der gestaltungsorientierten Wirtschaftsinformatik. In: Österle, H., Winter, R., Brenner, W. (eds.): Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz. Infowerk, Nürnberg (2010)
33. Lovasz-Bukvova, H.: Scientists' self-presentation on the Internet, <http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-87592>
34. Ferguson, R., Clough, G., Hosein, A.: Shifting themes, shifting roles: the development of research blogs. In: Proceedings of the 17th Association for Learning Technology Conference, pp. 111–117 (2010)
35. Nentwich, M.: Web 2.0 and academia. In: Proceedings of the 9th Annual IAS-STS Conference "Critical Issues in Science and Technology Studies", pp. 66–78 (2010)
36. Heimeriks, G., Vasileiadou, E.: Changes or transition? Analysing the use of ICTs in the sciences. *Social Science Information* 47, 5–29 (2008)
37. Granovetter, M.S.: The strength of weak ties. *American Journal of Sociology* 78, 1360–1380 (1973)
38. Haythornthwaite, C.: Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society* 18, 385–401 (2002)
39. Marsden, P.V., Campbell, K.E.: Measuring tie strength. *Social Forces* 63, 482–501 (1984)
40. Tenopir, C., King, D.W., Edwards, S., Wu, L.: Electronic journals and changes in scholarly article seeking and reading patterns. *Aslib Proceedings* 61, 5–32 (2009)
41. Bukvova, H.: Information Demand on Scientists' Internet Profiles, <http://sprouts.aisnet.org/11-173/>
42. Ragin, C.C.: *Redesigning Social Inquiry - Fuzzy Sets and Beyond*. University of Chicago Press, Chicago, IL, USA (2008)
43. Thomas, D.R., Decady, Y.J.: Testing for Association Using Multiple Response Survey Data: Approximate procedure Based on the Rao-Scott Approach. *International Journal of Testing* 4, 43–59 (2004)
44. Decady, Y.J., Thomas, D.R.: A simple test of association for contingency tables with multiple column responses. *Biometrics* 56, 893–6 (2000)
45. Fiss, P.C.: Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal* 54, (2011)
46. Hug, S.: An MCA of QCA, <http://www.unige.ch/ses/spo/static/simonhug/amoq/amoq.pdf>
47. Amenta, E., Poulsen, J.D.: Where to Begin: A Survey of Five Approaches to Selecting Independent Variables for Qualitative Comparative Analysis. *Sociological Methods & Research* 23, 22–53 (1994)
48. Fiss, P.C.: A set-theoretic approach to organizational configurations. *Academy of Management Review* 32, 1180–1198 (2007)

General Track - Visionary and Cross-Disciplinary Issues

Jörn Altmann¹ and Stefan Kirn²

¹ Seoul National University, Seoul, South Korea

² University Hohenheim, Stuttgart, Germany

Track 11, the track on visionary and cross-disciplinary issues, has been newly introduced to WI this year. The aim of this track has been not to address any specific topic but rather to allow for new ideas to emerge within WI. In particular, this track was looking for novel analysis ideas and novel applications of IS. The high number of 49 high quality paper submissions showed that this track had been a valuable addition to WI.

Each submission has been reviewed at least three times by an international program committee. Our final program consists of 5 sessions:

- Session 1: Internet Services
- Session 2: Adoption
- Session 3: Management
- Session 4: Privacy
- Session 5: Resource Allocation

As the five session titles suggest, the track brings together contributions from different research areas in IS, namely services, adoption, management, privacy, and resource allocation. In total, there are 15 contributions that have been selected from 49 submitted papers, resulting in an acceptance rate of 30,6%.

Session 1 on Internet services comprises three papers. The first paper by C. Kummer on “Student’s Intentions to Use Wikis in Higher Education” adapts the decomposed theory of planned behavior to examine the situation in higher education classrooms and tests the model with data collected from a survey. Their results show the importance of an intrinsic and extrinsic motivation construct to explain students’ wiki use. The paper entitled “Envy on Facebook: A Hidden Threat to Users’ Life Satisfaction?” by H. Krasnova, H. Wenninger, T. Widjaja, and P. Buxmann investigates the impact of envy through passive following in social networking systems based on two studies with 584 Facebook users. They find that users frequently perceive Facebook as a stressful environment. The final contribution in this session, entitled “Effects of the Network Structure on the Dynamics of Viral Marketing”, by M. Opuszko and J. Ruhland performs simulations on the diffusion of information. Based on the results, which show that diffusion in networks depends on initial settings and network structure, the authors derive recommendations for marketers.

Session 2 on adoption starts with a contribution from D. Grgecic entitled “A Structural Perspective on Belief Formation”, focusing on how IT-related factors influence the formation of object-based beliefs. The findings of a survey of 183 students in information system suggest that functionalities provided by an IT system positively affect beliefs about information and system quality. The second contribu-

tion entitled “A Comprehensive Analysis of E-Government Adoption in the German Household” has been authored by C. Akkaya, P. Wolf, and H. Krcmar. The results, which are based on a survey of 1000 households on e-filings, highlight that compatibility, relative advantage, and perceived risk are the main antecedent of e-filing adoption in Germany. The final contribution, which is provided by J. Putzke, K. Fischbach, D. Schoder, and P. Gloor, is entitled “The Coevolution of Network Structure and Perceived Ease of Use”. The authors introduce the concept of perceived ease of use as a network construct. A result shows that a person's perception of ease of use is directly linked to the frequency of being asked for advice.

Session 3 on management starts with a contribution entitled “Resistant Use of Project Management Methodologies - Using Psychology to Rethink the Influence of Methodology Attributes” by K. Mohan and F. Ahlemann. The authors investigate the way how project management methodologies are used. They survey 2645 subjects to examine the impact on a user's resistant use behavior. Their findings suggest that the need for cognition moderates the impact of project management methodologies on resistant use behavior. M. Böhm, C. Stolze, and O. Thomas in their paper entitled “Teaching the Chief Information Officers: An Assessment of the Interrelations within their Skill Set” investigate skill sets of CIOs. Their results identified a set of skills and how they relate. For their analysis, they collected data from a workshop with 21 CIOs. Finally, T. Wirtky, S. Laumer, A. Eckhardt, and T. Weitzel discuss social software features in e-learning on user motivation in a business context in their paper entitled “Towards Understanding Social Software and its Impact on Corporate e-Learning Motivation”. The results, which are based on 39 interviews, show that the impact of those features depends on the learner's needs.

Session 4 on privacy comprises the paper by A. Novotny and S. Spiekermann entitled “A Three-Tier Model for Personal Information Markets and Privacy”, which introduces a new conceptual model that allows managing businesses' demand for personal information and users' privacy. The authors also show how such a model can be introduced in the current environment. The second paper, which is entitled “Predicting the Disclosure of Personal Information on Social Networks: an Empirical Investigation” and has been authored by T. Buckel and F. Thiesse, addresses the question to what extent self-disclosure decisions and subsequent behavior can be explained by existing theories. Their result, which is based on survey data and data from Facebook, shows that simple regressions can predict behavior better than existing theories. The final contribution in this section comes from K. Steininger and D. Rückel and is entitled “Legal Literacy and Users' Awareness of Privacy, Data Protection and Copyright Legislation in the Web 2.0 Era”. This paper focuses on German and Austrian law on content production in the Internet. To understand users' lawful behavior, the authors analyzed a survey of 1134 students. The results indicate the impact of legal literacy, legal awareness and lawful attitude, among which legal awareness has the strongest impact.

Session 5 on resource allocation includes the final three contributions. The first contribution focuses on explaining the benefits of vehicle to grid technology and has been entitled “A Household-Oriented Approach to the Benefits of Vehicle-to-Grid-capable Electric Vehicles” by T. Brandt, S. Feuerriegel, and D. Neumann. With the help of a proof-of-concept case study, the authors show that simple management strategies can lead to additional revenues for a household and reduce the peak load on the electricity grid. The second contribution entitled “Simplified Bid Languages - a Remedy to Efficiency Losses in Large Spectrum Auctions” by S. K. Mayer and P. Shabalin investigates the combinatorial clock auction. By simplifying the bidding language, the authors could reduce the bidder’s complexity without losing efficiency. The final contribution of this session introduces a tool for evaluating wind energy projects and has been entitled “A Decision Support Tool for the Risk Management of Offshore Wind Energy Projects”. The authors, A. Koukal and M. H. Breitner, evaluate their tool through a simulation of an offshore wind park in the German North Sea.

Finally, we would like to thank all reviewers and program committee members for completing their reviews on time and for giving useful and valuable feedback to the authors. We would also like to extend our thanks to the conference chairs of WI2013, Rainer Alt und Bogdan Franczyk, for their excellent organization of the reviewing process. In addition to this, we would like to thank Christian Hrach for his permanent availability, immediate responses to our emails, and his support in keeping the deadlines.

Program Committee

Christian Anhalt, Bundesagentur für Arbeit
Byungjoon Yoo, Seoul National University
Byungtae Lee, Korea Advanced Institut of Science and Technology
Costas Courcoubetis, Athens University of Economics and Business
Andreas Dietrich, GfK-Nürnberg e.V.
Felix Hampe, University Koblenz-Landau
Knut Hinkelmann, Univ. of Applied Sciences and Arts Northwestern Switzerland
Kibae Kim, Technical University of Braunschweig
Jörg Leukel, University Hohenheim
Ashraf Bany Mohammed, University of Ha‘il
Claus D. Müller-Hengstenberg, University Hohenheim
Jan Muntermann, University Göttingen
Martin Sedlmayr, University Erlangen-Nürnberg
Ingo Timm, University Trier
Robert Woitsch, University Wien

Envy on Facebook: A Hidden Threat to Users' Life Satisfaction?

Hanna Krasnova¹, Helena Wenninger², Thomas Widjaja², and Peter Buxmann²

¹ Humboldt-Universität zu Berlin, Institute of Information Systems, Berlin, Germany
krasnovh@wiwi.hu-berlin.de

² TU Darmstadt, Chair of Software Business & Information Management, Darmstadt, Germany
{wenninger, widjaja, buxmann}@is.tu-darmstadt.de

Abstract. The wealth of social information presented on Facebook is astounding. While these affordances allow users to keep up-to-date, they also produce a basis for social comparison and envy on an unprecedented scale. Even though envy may endanger users' life satisfaction and lead to platform avoidance, no study exists uncovering this dynamics. To close this gap, we build on responses of 584 Facebook users collected as part of two independent studies. In study 1, we explore the scale, scope, and nature of envy incidents triggered by Facebook. In study 2, the role of envy feelings is examined as a mediator between intensity of passive following on Facebook and users' life satisfaction. Confirming full mediation, we demonstrate that passive following exacerbates envy feelings, which decrease life satisfaction. From a provider's perspective, our findings signal that users frequently perceive Facebook as a stressful environment, which may, in the long-run, endanger platform sustainability.

Keywords: Envy, Facebook, Passive Following, Life Satisfaction, Mediation

1 Introduction

With users sharing a whopping 30 billion pieces of content each month, Facebook (FB) represents the largest database of social information the world has ever witnessed [1]. By sharing their updates, users keep in touch and broaden their horizons [2]. As users learn more about each other, bonding and bridging social capital can be created [3-4]. Despite these benefits, opponents warn against negative repercussions these developments bring along. For example, past research has linked consumption of social information on FB to such undesirable outcomes as jealousy [5], increase in social tension [6], social overload [7], isolation [4] and even depression [8].

While findings on the negative effects of social information consumption are alarming, the underlying logic of this dynamic is little understood. In this regard, first reports underscore the proliferation of upward social comparison among members of Social Networking Sites (SNSs), suggesting that envy could be one of the most common negative consequences of following information of others on these platforms [9-10]. Indeed, defined as "*an unpleasant and often painful blend of feelings [...] caused*

by a comparison with a person or group of persons who possess something we desire" [11], envy can be triggered in a multitude of ways in the SNS environment. First, unprecedented scale of information sharing registered on SNSs naturally provides a ground for envy, which is typically induced when new information is learned about the other [11]. Second, SNSs offer users easy and transparent means to compare and "benchmark" themselves against their peers, inducing them to engage in social comparison. Moreover, an asynchronous and "controllable" way of communicating on SNSs creates vast possibilities for impression management, with SNS members often over-emphasizing their achievements [12].

Triggered by over-exposure to social information on a SNS, envy feelings can cause significant damage to users' well-being and impact their life satisfaction [11]. Indeed, past research from social psychology reveals that envy may lead to frustration [13], mental suffering [14] and even depression [15-16]. To limit their contact with envy-inducing information, users may consciously reduce platform use, as described for other contexts by past research on organizational psychology [17]. This, however, is undesirable for SNS providers, who face significant pressures to maintain a stable user base [18]. Overall, it appears that envy can indeed be the missing link explaining the negative effect of social information consumption on users' emotional states and loyalty behavior. However, despite potential seriousness of these effects, no study to the best of our knowledge directly investigates the phenomenon of envy in the SNS context.

To fill this gap, we adopt a two-stage approach. In study 1, we explore the scale, scope, and nature of envy-inducing incidents triggered by FB use. In study 2, the role of envy feelings is examined as a mediator between FB passive use and users' life satisfaction - a critical indicator of value engagement with SNSs brings along.

2 Theoretical Background

Indeed, outcomes of SNS use are tightly coupled with passive following of information others share on the platform – a behavior also synonymously referred to as "*content consumption*" [4]; "*social searching*" and "*social browsing*" [19]. Passive following takes place when users browse their News Feed, click on 'stories', follow communication of their friends, or proactively examine profiles of others. State-of-the-art research recognizes the importance of studying consequences of passive consumption of information since it represents the dominant activity on SNSs [19].

Apparently, widespread engagement in this activity signals a positive affect users experience in the process of social browsing [19]. Following details of others' lives may also have positive cognitive effects since it helps to reduce uncertainty, thereby providing basis for social trust, civic engagement and political participation [20]. Koroleva et al. [2] show that by passively following others on FB, users broaden their horizons and build a sense of connectedness. This helps them realize an array of tangible networking benefits. On the other hand, a growing body of research warns against this one-sided positive view. Indeed, most recent evidence suggests that continuous engagement in passive following may lead to feelings of exhaustion, annoy-

ance, irritability and overload [7], [21]. In a seminal study, Burke et al. [4] uncover a significant link between social content consumption and perceptions of loneliness. Altogether, these findings are alarming, since they signal a potential of SNSs to intervene with users' well-being and life satisfaction and, over long-term, impact platform sustainability. Nonetheless, little research exists uncovering the roots of the observed dynamics.

The findings of Jordan et al. [22] suggest that upward social comparison and envy can be rampant in a peer-dominated SNS environment, which can provide explanation to negative outcomes passive following was shown to produce. Specifically, in a non-SNS related study, the authors show that people tend to underestimate negative experiences of others and overestimate their positive experiences, which causes negative emotions to proliferate. SNS environment is particularly likely to exacerbate envy feelings, since it promotes narcissistic behavior, with most users sharing only positive things about themselves [12]. For example, Chou and Edge [9] find that respondents actively using FB were more likely to agree that "*others had better lives than themselves*". Moreover, friend lists typically consist of individuals with a high degree of similarity to the profile owner [23], which is particularly conducive for the proliferation of envy feelings [11].

However, despite the obvious potential of SNSs to promote envy and social comparison, no study to the best of our knowledge has investigated the scale and consequences of this phenomenon in the SNS context. Recognizing this lack of studies, we draw on research from social psychology and organizational science to discuss potential effects of this emotional state. In this research, envy is typically described as a painful emotion that emerge as a result of upward comparison to advantaged others, who possess something, that one covets but lacks [11]. Envy is an unalienable part of social interaction, with people experiencing this feeling in private and workplace settings, or any other environment, where inter-personal interactions take place [24]. On the positive side, benign envy was shown to lead to learning, motivation, better performance, and achievement [25]. On the negative side, malicious envy leads to desire to harm the envied object and breeds hostility [24]. Endured over longer time periods, envy can damage one's sense of self-worth, result in group dissatisfaction and withdrawal, lead to depressive tendencies, reduce perceptions of well-being, and poor mental health [11]. Considering these detrimental effects of envy, in this study, we explore the dynamics of envy processes in the SNS context. Specifically, in the next step, we examine the scale, scope, and nature of envy-inducing incidents triggered by FB. Building on our findings, we then explore whether envy feelings can explain a negative impact of passive consumption on individual well-being.

3 Study 1: Exploring the Scale, Scope, and Nature of Envy on SNSs

Considering the lack of studies directly addressing the phenomenon of envy in the SNS context, the scale, scope, and nature of envy phenomenon on SNSs have been studied first in an explorative manner. To do so, a short online survey including a

mixture of open- and close-ended questions has been conducted. Responses to open-ended questions have been always content-analyzed in two steps. First, authors have identified relevant categories. Next, two independent coders were trained to code the data. This allowed us to calculate Inter-Coder Reliability (IRR) measured by Cohen's kappa to ensure the quality and validity of the coding procedure. In case of disagreement, a final decision on the attribution of a code was achieved via consensus by authors. We refrain from describing questions in a separate part of this paper due to space limitations, but rather present them in the course of our analysis below.

The survey was advertised using a mailing list of Humboldt-Universität zu Berlin and was positioned as a study of “emotions of Facebook users” to avoid priming. A lottery of 15 € Amazon.de gift cards was offered as an incentive to take part in the study. 357 respondents answered most parts of the survey (the answers were not forced). 34.2% (65.8%) of respondents were male (female). The median age comprised 24 years. 93.8% stated Germany as the country, where they have spent most time of their life. 89.9% (3.9%) of respondents were students (employed). 25% of respondents studied language / culture studies, 7.6% studied law, with the rest studying a great variety of other disciplines. The median number of FB friends reached 169. 50.0% of respondents stated to spend between 5 and 30 minutes daily on FB.

Question 1: Emotional Outcomes of Facebook Use. To avoid priming, a general open-ended question was asked first: “Please think about the last time you used Facebook. What did you feel afterwards? Which emotions have you experienced?” 347 (97.1%) of respondents gave a short answer to this question resulting in a data corpus of 3167 words. Initial analysis and subsequent coding of responses reveal a plethora of emotional outcomes the use of FB can produce (see Table 1).

Table 1. Emotional outcomes of most recent Facebook use.

Emotional Outcomes	Share of respondents	Emotional Outcomes	Share of respondents
<i>Positive (at least one)</i>	43.8%	<i>Negative (at least one)</i>	36.9%
• joyful / fun	28.8%	• bored	13.8%
• satisfied	9.5%	• angry	9.2%
• informed	7.8%	• frustrated	8.9%
• excited	6.1%	• guilty	3.5%
• relaxed	4.3%	• tired	2.9%
<i>Neutral</i>	28.8%	• sad	2.6%
<i>Other</i>	4.3%	• lonely	1.4%
		• envious	1.2%

Specifically, five positive and eight negative emotional outcomes emerged from the data varying in their degree of intensity. 29.7% of respondents reported two or more emotional states, with the rest reporting only one emotional outcome. IRR was high reaching 0.733 ($p = 0.000$), which shows that our data is suitable for further analysis. We find that 43.8% of respondents report at least one positive emotional outcome following their Facebook use. At least one negative emotional state is reported by the lower share of respondents (36.9%). Among positive outcomes, “joy and fun” feelings were the most prevalent with 28.8% of respondents reporting these experiences. The-

se were followed by feelings of “*satisfaction*”, feeling of being “*informed*”, “*excitement*” and “*relaxation*”. Among negative outcomes, 13.8% of respondents reported feeling “*bored*”, 9.2% admitted “*anger*” and 8.9% reported “*frustration*”. “*Envy*” – a subcategory in the focus of our study – was only mentioned by 4 respondents in our sample. The seeming unimportance of envy revealed in this analysis may be rooted in respondents’ reluctance to directly admit to experiencing envy, rather reporting such general outcomes as feelings of anger, exhaustion, frustration, and irritation [11], [26]. Admitting to these feelings is more socially acceptable, since they can be equally caused by information overload [21] or social conflict [6], which carry less social stigma. Whether this explanation holds, was tested in a follow-up question.

Question 2: Causes of Frustration with Facebook. The second question was not directed at the respondent but rather targeted emotional outcomes of “others”. Specifically, respondents were asked: “*Many users report feeling frustrated and exhausted after using Facebook. What do you think causes these feelings?*” Projective techniques are often used in survey design, as they help to elicit honest responses [27]. 307 (86.0%) respondents answered this question resulting in a data corpus of 5831 words. As a result of initial content analysis, 13 sub-categories have been identified, as summarized in Table 2. IRR was high reaching 0.735 ($p = 0.000$). Most respondents mentioned only one reason for frustration, 17.3% mentioned two, and 2.0% mentioned three reasons. We find that “*envy*” emerges as the category of the highest importance with 29.6% of respondents mentioning it as a major reason behind frustration and exhaustion of “others”. Feelings of envy by far surpass such causes, as “*lack of attention*” (19.5%), “*loneliness*” (10.4%), and “*time loss*” (13.7%). This outcome suggests that even though respondents do not admit feeling envy when asked directly (question 1), they readily relate this emotion to frustration resulting from FB use.

Table 2. Reasons for “others” being frustrated after Facebook use.

Reasons	Explanation	Share of Respondents
Social Causes		
• lack of attention	lack of comments, likes, feedback	19.5%
• social problems	conflicts, tension, quarrel	2.0%
• loneliness	social isolation, no face-to-face contact	10.4%
• having missed smth	missing events, concerts, not invited	5.5%
• jealousy	jealous of one’s (ex-)partner, friend	2.0%
• envy	envy, social upward comparison	29.6%
Informational Causes		
• content of news	sad news, frustrating news	10.1%
• uselessness of news	information overload, bad content	7.2%
• boring news	boredom, no news, uninteresting news	5.5%
Functionality	long loading times, missing functionality	4.2%
Time Loss	losing time	13.7%
Other Causes	other non-classified causes	1.3%
Unidentified Causes	not being able to identify a cause	5.9%

Moreover, when subsequently asked “How often do you experience feelings of frustration and exhaustion after using Facebook?” with pre-specified answers on a 7-point scale: 1 = (almost) never; 4 = sometimes; 7 = (almost) every time, 36.4% of respondents reported feeling frustrated and exhausted at least sometimes or more often. Among those respondents, 29.2% mentioned envy as a major cause of frustration behind FB use, thereby indirectly admitting to envy. This outcome leads us to conclude that even though respondents are unwilling to admit to it, envy is a common consequence of SNS use.

Question 3: Triggers of Envy and Role of Facebook. The scope and nature of the envy phenomenon on SNSs has been further explored by asking respondents about the context of their most recent envy experience: “Please think about the last time you envied someone. Where did you experience this feeling?”. Pre-specified options included “personal encounter”, “Facebook”, “Xing” and “other”. We find that 71.5% (238) of recent envy-inducing incidents are still experienced *offline*. Nonetheless, Facebook is responsible for causing 21.3% (71) of most recent envy cases, with the rest being triggered by “Xing” or “Other” settings. This magnitude of envy incidents taking place on FB alone is astounding, providing evidence that FB offers a breeding ground for invidious feelings. To better understand in which areas incidents of envy are common and whether FB provokes incidents of different nature, the following question was posed: “What have you envied that time?”. This autobiographical narrative methodology has been successfully used in eliciting envy-related and other secretive experiences [28].

Table 3. Envy-inducing incidents in offline settings and on Facebook.

Envy-inducing incidents	Share of Respondents		Mann-Whitney Test p-value
	Offline	FB	
<i>N sample</i>	238	71	
Travel and Leisure	19.3%	56.3%	.000
Social Interaction	6.3%	14.1%	.035
Love, Family and Relationship	5.9%	4.2%	.592
Success			
• money / material possessions	8.0%	1.4%	.049
• success in studies	8.4%	2.8%	.109
• abilities	11.8%	2.8%	.026
• success in job	15.5%	2.8%	.005
• success in general	10.5%	5.6%	.218
Appearance	7.1%	2.8%	.184
Personality	6.3%	1.4%	.103
Happiness	6.3%	7.0%	.824
Other	2.9%	5.6%	.283
No Envy	2.5%	1.4%	.581

333 obtained responses resulted in a data corpus of 2834 words. In most cases only *one* subject of envy has been stated, with 9.6% indicating *two* subjects. In total, 13 categories have been identified and subsequently coded, with IRR reaching an appropriate level of 0.706 ($p = 0.000$). Table 3 summarizes the differences between envy subjects for *offline* and *FB* contexts. We find that envy about “*travel and leisure*”, “*social interactions*” and “*happiness*” belong to the three most frequently mentioned causes of envy triggered by *Facebook* use. In daily encounters, however, “*travel and leisure*”, “*success in job*” and “*abilities*” of another person represent the three most common categories. Overall, the fact that “*travel and leisure*” account for a whopping 56.3% of all envy incidents triggered by FB is interesting. The reasons for this are likely rooted in a high share of travel photos posted by FB users. Indeed, while sharing content directly depicting expensive material possession might be seen as bragging by others; posting photos from vacations has long established itself as a norm on SNSs [29]. As a result, by sharing this type of content respondents do not risk to be accused of engaging in the outright self-promotion, while still, in a way, doing so.

4 Study 2: The Role of Facebook Envy in Users’ Life Satisfaction

In line with study 1, we found that envy feelings are often triggered by following information of others on FB. Experienced over a long time period, these invidious emotions can lead to frustration and exhaustion, damaging individual life satisfaction – a critical indicator of users’ well-being. To investigate whether these effects take place, a follow-up study investigating the role of envy feelings in the relationship between passive following and life satisfaction – was undertaken.

4.1 Research Hypotheses

Most studies consistently find a positive link between *active* communication or general FB activity and such desirable outcomes as life satisfaction [30], social capital [20], and emotional support [2]. The evidence, however, is mixed when it comes to the impact of *passive* use of SNSs, suggesting that directionality of FB impact depends on a specific type of user activity. Since passive use is central for the emergence of envy feelings, we concentrate on this type of user behavior in our model.

Using a convenience student sample and self-report measures, Koroleva et al. [2] find a positive link between intensity of information consumption and such favorable outcomes as horizon-broadening and networking value. Nonetheless, a number of recent studies voice an array of concerns regarding negative effects of passive following, including feelings of exhaustion [7], information overload [21], and social tension [6]. Apparently, while passive following satisfies users’ needs for novelty and cognitive stimulation, thereby resulting in a positive affect in the short-run [19], [31], in the long-term excessive consumption of social information leads to depletion and exhaustion. Confirming these effects, Burke et al. [4] find a positive link between intensity of passive following and perceptions of loneliness. Based on FB log data as

a measure of user activity, the findings of this study appear to provide the most reliable snap-shot of long-term consequences of FB usage available so far. Therefore, considering that global life satisfaction and loneliness are related components and are both parts of the concept of well-being [32], we cautiously hypothesize that: **Intensity of passive following on FB is negatively associated with Life Satisfaction (H1).**

With an average user having 130 friends in her profile, SNSs offer a fertile ground for episodic envy feelings to proliferate. These feelings can be triggered as users encounter positive information of others and engage in social comparison while browsing the News Feed [9]. Presence of friends with similar background is likely to magnify these effects, since demographic similarity is often used to identify a suitable reference group one attempts to match with [11]. Moreover, proliferation of narcissistic self-presentation, well documented by past research [12], makes activation of invidious emotions hard to avoid. In fact, both men and women find themselves under high pressure to communicate their best sides to their peers, even though in different areas. For example, male users have been shown to post more self-promotional content in “About Me” and “Notes” sections on FB, as they attempt to communicate their accomplishments and establish social standing [12]. Women, on the other hand, stress their physical attractiveness and sociability [33]. Overall, however, shared content does not have to be “explicitly boastful” for envy feelings to emerge. In fact, a lonely user might envy numerous birthday wishes his more sociable peer receives on his FB Wall [4]. Equally, a friend’s change in the relationship status from “single” to “in a relationship” might cause emotional havoc for someone undergoing a painful breakup. Against this background, we hypothesize that: **Intensity of passive following on FB is positively associated with Envy on FB (H2).**

Social and organizational psychology research provides a well-documented body of scientific studies on the negative consequences of invidious emotions [11], [24]. In work settings, job-related envy was shown to lead to extreme general, job and group dissatisfaction, and even health problems [34]. Similar effects have been also registered in other settings. For example, Salovey and Rodin [35] show that unfavorable social-comparison on self-relevant dimensions elevates anxiety in college undergraduates. Moreover, experienced over longer time periods, feelings of envy were shown to cause inferiority attitudes, which lead to depression and poor mental health [11]. Equally, we expect Facebook-induced envy to have a negative impact on personal well-being. We, therefore, hypothesize that: **Envy on FB is negatively associated with Life Satisfaction (H3).**

Central for our work, Hypothesis 4 disentangles the relationships described above (H1-H3). Indeed, a number of studies provide evidence for the link between passive following on FB and such outcomes as feelings of exhaustion [7], social tension [6], and loneliness [4]. Considering the central role that *Envy* plays in the interpretation and internalization of information about others [26], [36], it is plausible to hypothesize that these negative outcomes are mainly enacted by the *Envy* processes it triggers. In other words, in the context of our study it is not (only) passive following by itself that reduces individual life satisfaction, but also (and rather) these are envy feelings, passive following is likely to produce, which lead to this negative outcome. We hy-

pothesize that: **Envy mediates a relationship between Intensity of Passive Following on FB and Life Satisfaction (H4).**

4.2 Control Variables

Past research offers significant support that SNS-enabled outcomes are contingent on user demographic characteristics. For example, women were found to emphasize relational uses of SNSs more than men, which make women particularly likely to engage in upward social comparison [37]. Hence, we include *gender* as a control variable. Additionally, age differences may also intervene with SNS-enabled outcomes and envy processes. This is because older people are likely to have more experience in coping with envy-inducing incidents, which helps them minimize the impact of invidious emotions [11]. Hence, we also integrate *age* as another control variable. Further, size, and structure of one's friend list has often been used as a predictor of SNS-enabled outcomes [4]. Therefore, we integrate *number of friends* as the third control variable into our model. Finally, users who actively participate on a SNS by sharing status updates, links, and comments may simultaneously be eager consumers of social information. In this case, both active and passive participation can impact the reported level of life satisfaction. To discern the confounding influence of these variables, *active participation* was additionally controlled for.

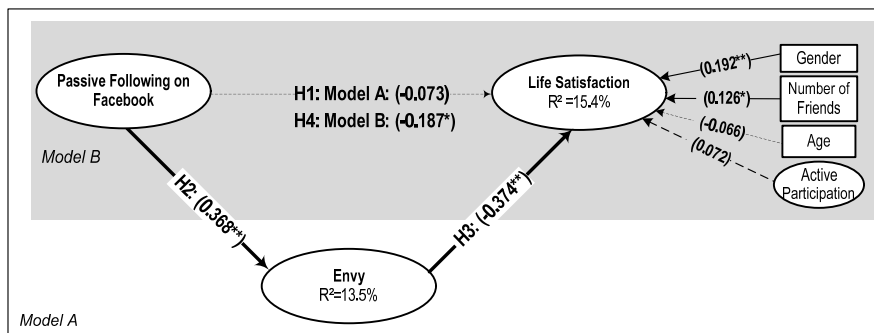
4.3 Survey Design and Sampling

To validate our research model a follow-up online survey was conducted by distributing the invitation via a mailing list similar to study 1. Participation in a raffle of 40 €5 Amazon.de gift certificates was offered as an incentive to take part in our survey. While we relied on the pre-tested measures where possible, some scales had to be developed anew or slightly modified to fit to the context of our study. A 9-item **Envy** scale reflecting the FB context was self-developed. Examples include: “*When on Facebook, I catch myself envying: (1)...how much of the world others have already seen; (2)...how successful others are; (3)...what abilities others have; (4)...how happy others are.* The 7-point scale used ranged from 1 = “(almost) never” to 7 = “(almost) always”. Four items for the **Life Satisfaction** Scale were borrowed from Diener et al. [32] and measured on a 7-point Likert scale. Examples include: (1) *In most ways my life is close to my ideal;* or (2) *I am satisfied with my life.* The measure for **Passive Following** included items: “*On Facebook, how often do you? (1)...look through the News Feed; (2) ...look through the conversations your friends are having; (3)...browse the profiles of others*”, which were largely based on Koroleva et al. [2]. A 7-point scale was used: 1=never; 7=several times a day. All constructs were measured as reflective. **Active Participation** was measured based on Koroleva et al. (2011). 227 usable observations were obtained. 31.7% (67.8%) of respondents were male (female). The median age comprised 24.3 years. 94.3% stated Germany as their country of residence. 86.3% (4.4%) of respondents were students (employed). 25.1% of respondents studied language / culture studies, 10.1% studied social sciences, and 7.5% business and economics. The median number of FB friends reached 179. 5.7%

of respondents stated to spend less than 5 minutes per day on FB; 45.8% spend between 5 and 30 minutes; 22.0% spend between 31 minutes and 1 hour; and 26.4% spend more than 1 hour on FB.

4.4 Evaluation of the Research Model

Exploratory nature of our study as well as a non-normal distribution of our data dictated the choice of Partial Least Squares (PLS) approach to test our hypotheses using SmartPLS 2.0.M3 [38]. Model A was evaluated to test H1, H2 and H3 (see Figure 1). Following Baron and Kenny [39], model B – only including a direct causal link from *Passive Following* to *Life Satisfaction* – was additionally assessed to test the mediation effect of *FB Envy* (H4) (see Figure 1).



Significance: * at 5%; ** at 1% or lower; R² values reported for Model A.

Fig. 1. Results of the Structural Model

Assessment of Models A and B included two steps: First Measurement Model (MM) and then Structural Model (SM) were evaluated. MM was evaluated by verifying the criteria for Convergent Validity (CV) and Discriminant Validity (DV). To ensure CV, parameters for Indicator Reliability (IR), Composite Reliability (CR) and Average Variance Extracted (AVE) were assessed. For IR, constructs should explain at least 50% of the variance of their respective indicators. Items with factor loadings below 0.4 should be removed from the model [40]. The majority of item loadings in both models satisfied the former strict criteria, exceeding the level of 0.7 [41]. Two items in Model A and two items in Model B were only slightly below 0.7. Taken together, IR was assured. Further, CR values for all constructs in both models were higher than the required level of 0.7 [41]. The AVE values for all measured constructs by far surpassed the threshold level of 0.5 [42]. Finally, Cronbach's Alpha (CA), a measure of Internal Consistency of construct scales, reached 0.92 for *Envy*, 0.84 for *Life Satisfaction*, 0.68 for *Passive Following*, and 0.81 for *Active Participation*. Taken together, CV can be assumed. Next, DV was assessed by ensuring that the square root of AVE for each construct was higher than the correlation between this construct and any other construct in the model [41]. This requirement was fulfilled for all constructs in both models. Moreover, the correlation between *Passive Following* and *Active Partic-*

ipation comprised only 0.505, providing evidence that two types of participation are related, yet different constructs. Altogether, MMs for both Model A and B are well-specified (all item and construct-level data is available from authors upon request due to space limitations).

Next, the Structural Model (SM) of Model A was assessed (see Figure 1). With *Envy* included, we find no link between intensity of *Passive Following* on FB and users' *Life Satisfaction* (*H1 rejected*). At the same time, we find a strong positive link between *Passive Following* and *Envy* (*H2 supported*) and a negative link between *Envy* and *Life Satisfaction* (*H3 supported*). Interestingly, being a woman and having more FB friends increases one's *Life Satisfaction*. Altogether, our model A explains 15.4% of variance in *Life Satisfaction*. Considering that a multitude of other factors can be responsible for individual perceptions of *Life Satisfaction* – social status, family situation, culture, just to name a few – this degree of explanatory power is astounding. Finally, the presence of the mediation effect was evaluated by assessing *Model B*. According to Baron and Kenny [39], full mediation can be observed when a path significant in a direct model (*Model B*) becomes insignificant when the mediator variable – *Envy* – is present (*Model A*), assuming that relationships between independent variable and mediator; and mediator and dependent variable are significant, which was confirmed in *Model A*. For our case, removal of *Envy* construct revealed a significant negative path between *Passive Following* on FB and *Life Satisfaction*. The Sobel Test statistic used to test the mediation effect in *Model A* was also significant ($p=0.000$). Taken together, we conclude that *Envy* fully mediates the relationship between *Passive Following* on FB and *Life Satisfaction* (*H4 supported*).

5 Theoretical and Managerial Implications

Our study is the first step in understanding envy dynamics on SNSs in general and FB in particular. We find that envy feelings are common on FB, with 20.3% of all recent envy incidents being triggered by FB use. The subjects of envy are contingent on the content users provide on these platforms with “*travel and leisure*”, “*social interactions*” of others, and “*happiness*” landing on the top of the list. This is in some contrast to offline encounters, where “*travel and leisure*”, “*success in job*”, and “*abilities*” are envied more often.

By confirming the mediating role of envy in the relationship between passive following on FB and individual life satisfaction, our findings significantly enrich existing literature on the role of SNSs in defining users' well-being. We show that intensity of passive following is likely to reduce users' life satisfaction in the long-run, as it triggers upward social comparison and invidious emotions. We expect these findings to hold across cultures, since envy feelings are a ubiquitous phenomenon, representing an important building block of evolutionary processes [11]. Our findings complement the findings by Burke et al. [4], who were first to link intensity of information consumption on SNSs and loneliness, but left explanation of this phenomenon for future studies. Discovered in our data, a full mediation role of envy provides rationale for the dynamics Burke et al. [4] reveal.

In addition, our findings suggest that outcomes of SNS participation are a function of user behavior. While directed communication has been shown to lead to positive outcomes [3], effects of passive following are more complex. As a result, we recommend future studies to avoid combining different types of user activity into one “general” variable. Instead different types of user behavior should be integrated as independent constructs.

The fact that envy feelings are rampant in SNS environment and can be intensified by passive following – two important outcomes of our study – should be worrisome for providers. Indeed, Tai et al. [24] argue that whether malicious or benign – envy is an experience characterized by pain. To deal with it, people are likely to seek ways to minimize or avoid these experiences. In the SNS context possible strategies may include (1) avoiding adding friends one feels particularly envious about or (2) unfriending these people. These strategies, however, are unlikely to be popular among users, since they contradict social norms present on SNSs and can lead to social tension [23]. Further, as part of their envy-coping plan, some users may (3) engage in even greater self-promotion and impression management. After all, overstatement of personal accomplishment is a common reaction to envy feelings [44]. This behavior can trigger the phenomenon we denoted as the *self-promotion – envy spiral*, with users reacting with even more self-promotional content to the self-promotion of others. As a result, envy-ridden character of the platform climate can become even more pronounced. Finally, the most straightforward way to cope with envy can be (4) to hide posts from friends one feels particularly envious about or (5) to partly or even fully refrain from passive following on SNS. While hiding posts is likely to go unnoticed to the other person, not following social content can lead user to miss out on some relevant information about events or other occurrences, thereby harming a user. Moreover, these strategies are particularly likely to undermine platform sustainability in the long-run, since passive following represents an integral part of SNS participation. Hence, addressing this threat should be seen as priority by providers.

While providers cannot impact general propensity to envy, which is a deeply-rooted personality characteristic, they can reduce users’ exposure to particular content, thereby reducing “*episodic envy*” [25], especially for those who are particularly likely to experience invidious emotions. For example, past research shows that users tend to envy those who are *similar* to them in terms of such characteristics as gender, age, cultural background and social status, since they provide a suitable reference group for self-evaluation [11]. Hence, demographics can be used as criteria for information filtering by providers. Second, a series of experiments conducted by Hill et al. [36] show that people tend to study information of others more attentively when envy is involved. Hence, analysis of log files may provide relevant insights into which friends may be triggering invidious reactions, which again can be used for filtering information presented to a user. Third, Smith and Kim [11] suggest that only domains of comparison a subject considers *relevant* can trigger envy. To identify these domains, past research can be drawn upon. For example, people in mid-thirties are more likely to envy family happiness, while teenagers will be indifferent to this information. Similarly, women are more likely to envy physical attractiveness [36]. Integrating these simple heuristics as part of information filtering mechanism can signifi-

cantly reduce users' exposure to envy-inducing information, leading them to perceive a SNS as a stress-free environment. Finally, users with more favorable self-evaluations have been found to react less negatively when envious, while the opposite was true for people with low self-esteem, low self-efficacy, high neuroticism, and high external locus of control [24]. To elicit these personality characteristics, providers can offer users participation in "Personality Quizzes" highly popular on SNSs. Responses to these tests can then be used to identify a group of users particularly sensitive to envy and use special filtering techniques to limit their exposure to envy-inducing content. Overall, regularly measuring the levels of envy on a SNS, and even including an "envy barometer" as a key performance indicator appears to be a rational choice in the light of destructive effects envy feelings can produce in the long-run.

6 Concluding Remarks and Future Research

This study uncovers a rampant nature of envy on SNSs. According to our findings, passive following triggers invidious emotions, with users mainly envying happiness of others, the way others spend their vacations; and socialize. The spread and ubiquitous presence of envy on SNSs is shown to undermine users' life satisfaction – a major contribution of study. In fact, feelings of envy mediate the relationship between passive consumption of information on SNSs and life satisfaction, which provides rationale for relationships identified but not explained in previous research [4]. Finally, our results offer an explanation to the ever increasing wave of self-presentation and narcissism behavior witnessed on SNSs – a phenomenon we refer to as the *self-promotion - envy spiral*. On the limitations side we note that most respondents in both studies were German students. While envy is a worldwide phenomenon, German users might have a distinct sharing behavior, making specific envy subjects particularly salient. Future studies may reveal these exciting particularities.

References

1. Pring, C.: <http://thesocialskinny.com/100-social-media-statistics-for-2012/>
2. Koroleva, K., Krasnova, H., Veltri, N.F., Günther, O.: It's All About Networking! Empirical Investigation of Social Capital Formation on Social Network Sites. In: ICIS 2011 Proceedings (2011)
3. Burke, M., Kraut, R., Marlow, C.: Social capital on Facebook: Differentiating uses and users. In: Conference on Human Factors in Computing Systems, pp. 571-580. ACM, (2011)
4. Burke, M., Marlow, C., Lento, T.: Social network activity and social well-being. In: Conference on Human Factors in Computing Systems, pp. 1909-1912. ACM (2010)
5. Muise, A., Christofides, E., Desmarais, S.: More Information than You Ever Wanted: Does Facebook Bring Out the Green-Eyed Monster of Jealousy?. *CyberPsychology & Behavior* 12 (2009)
6. Boyd, D.: Friends, friendsters, and myspace top 8: Writing community into being on social network sites (2006)

7. Maier, C., Laumer, S., Eckhardt, A., Weitzel, T.: When Social Networking Turns to Social Overload: Explaining the Stress, Emotional Exhaustion, and Quitting Behavior from Social Network Sites' Users. In: ECIS 2012 Proceedings (2012)
8. O'Keeffe, G.S., Clarke-Pearson, K.: The impact of social media on children, adolescents, and families. *Pediatrics* 127, 800-804 (2011)
9. Chou, H.T.G., Edge, N.: "They Are Happier and Having Better Lives than I Am": The Impact of Using Facebook on Perceptions of Others' Lives. *Cyberpsychology, Behavior, and Social Networking* 15 (2012)
10. News, C.: http://www.cbsnews.com/2100-502303_162-7309480.html
11. Smith, R.H., Kim, S.H.: Comprehending envy. *Psychological Bulletin* 133 (1), 46 (2007)
12. Mehdizadeh, S.: Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking* 13, 357-364 (2010)
13. Maurya, A.: *Running Lean: Iterate from Plan a to a Plan That Works*. O'Reilly & Associates Inc (2012)
14. Piskorski, M.J.: Social Strategies That Work. *Harvard Business Review* 89, 116-122 (2011)
15. Etchegoyen, R.H., Nemas, C.R.: Salieri's Dilemma: A counterpoint between envy and appreciation. *The International Journal of Psychoanalysis* 84, 45-58 (2003)
16. Malone, T., Weill, P., Lai, R., D'Urso, V., Herman, G., Apel, T., Woerner, S.: Do some business models perform better than others? MIT Sloan Research Paper No. 4615-06 (2006)
17. Vecchio, R.P.: Negative emotion in the workplace: Employee jealousy and envy. *International Journal of Stress Management* 7, 161-179 (2000)
18. Krasnova, H., Hildebrand, T., Guenther, O., Kovrigin, A., Nowobilska, A.: Why Participate in an Online Social Network? An Empirical Analysis. In: 16th European Conference on Information Systems (ECIS) (2008)
19. Wise, K., Alhabash, S., Park, H.: Emotional responses during social information seeking on Facebook. *Cyberpsychology, Behavior, and Social Networking* 13, 555-562 (2010)
20. Valenzuela, S., Park, N., Kee, K.F.: Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation I. *Journal of Computer-Mediated Communication* 14, 875-901 (2009)
21. Koroleva, K., Krasnova, H., Günther, O.: STOP SPAMMING ME! - Exploring Information Overload on Facebook. In: AMCIS 2010 Proceedings (2010)
22. Jordan, A.H., Monin, B., Dweck, C.S., Lovett, B.J., John, O.P., Gross, J.J.: Misery has more company than people think: Underestimating the prevalence of others' negative emotions. *Personality and Social Psychology Bulletin* 37, 120-135 (2011)
23. Krasnova, H., Koroleva, K., Veltri, N.F.: Investigation Of The Network Construction Behavior On Social Networking Sites. In: ICIS 2010 Proceedings (2010)
24. Tai, K., Narayanan, J., McAllister, D.J.: Envy As Pain: Rethinking the Nature of Envy and Its Implications for Employees and Organizations. *The Academy of Management Review (AMR)* 37, 107-129 (2012)
25. Cohen-Charash, Y.: Episodic envy. *Journal of Applied Social Psychology* 39, 2128-2173 (2009)
26. Foster, G.: The anatomy of envy. *Current Anthropology* 13, 165-202 (1972)
27. Donoghue, S.: Projective techniques in consumer research. *Journal of Family Ecology and Consumer Sciences/Tydskrif vir Gesinsekologie en Verbruikerswetenskappe* 28, (2010)
28. DelPriore, D.J., Hill, S.E., Buss, D.M.: Envy: Functional specificity and sex-differentiated design features. *Personality and Individual Differences* 53, 317-322 (2012)

29. Strano, M.M.: User Descriptions and Interpretations of Self-Presentation through Facebook Profile Images. *Journal of Psychosocial Research on Cyberspace* 2 (2008)
30. Ellison, N., Steinfield, C., Lampe, C.: The benefits on Facebook "Friends": Social Capital and College Students Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12 (2007)
31. Hui, K.L., Tan, B.C.Y., Goh, C.Y.: Online Information Disclosure: Motivators and Measurements. *ACM Transactions on Internet Technology* 6, 415-441 (2006)
32. Diener, E., Emmons, R.A., Larsen, R.J., Griffin, S.: The satisfaction with life scale. *Journal of personality assessment* 49, 71-75 (1985)
33. Manago, A.M., Graham, M.B., Greenfield, P.M., Salimkhan, G.: Self-presentation and gender on MySpace. *Journal of Applied Developmental Psychology* 29, 446-458 (2008)
34. Vecchio, R.: Explorations in employee envy: Feeling envious and feeling envied. *Cognition & Emotion* 19, 69-81 (2005)
35. Salovey, P., Rodin, J.: Some antecedents and consequences of social-comparison jealousy. *Journal of Personality and Social Psychology* 47, 780 (1984)
36. Hill, S.E., DelPriore, D.J., Vaughan, P.W.: The cognitive consequences of envy: Attention, memory, and self-regulatory depletion. *Journal of Personality and Social Psychology* 101, 653-666 (2011)
37. Goudreau, J.: <http://www.forbes.com/2010/04/26/popular-social-networking-sites-forbes-woman-time-facebook-twitter.html>.
38. Ringle, C.M., Wende, S., Will, A.: *SmartPLS 2.0 (beta)* (2005)
39. Baron, R.M., Kenny, D.A.: The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51, 1173 (1986)
40. Homburg, C., Giering, A.: Konzeptualisierung und Operationalisierung komplexer Konstrukte. *Marketing-Zeitschrift für Forschung und Praxis* 18, 5-24 (1996)
41. Hulland, J.: Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal* 20, 195-204 (1999)
42. Fornell, C., Larcker, D.F.: Structural Equation Models with Unobservable Variables and Measurement Errors. *Journal of Marketing Research* 18, 39-50 (1981)
43. Hulland, J.: Use of partial least-squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal* 20, 195-204 (1999)
44. Gino, F., Pierce, L.: Dishonesty in the name of equity. *Psychological Science* 20, 1153-1160 (2009)

Students' Intentions to Use Wikis in Higher Education

Christian Kummer

Technische Universität Dresden, Dresden, Germany
christian.kummer@tu-dresden.de

Abstract. Although wikis have gained considerable attention in higher education, students are often reluctant to use wikis in formal learning processes. Unlike company employees, students are not often rewarded for their participation in wiki-based assignments. Therefore, students seem to be opportunistic and decide to adopt wikis if they fit their current situation and preferences. This paper adapts the Decomposed Theory of Planned Behaviour to examine the situation in higher education classrooms. To better understand students' decision to use wikis, we introduced an intrinsic and extrinsic motivation construct. A survey was conducted with 133 first semester students to test the proposed model. The results provide support for the importance of an intrinsic and extrinsic motivation construct to explain influence on students' wiki use.

Keywords: technology acceptance, wiki, intrinsic motivation, extrinsic motivation, decomposed theory of planned behaviour

1 Introduction

The Web 2.0 has attracted considerable attention over the past few years. Social networking sites, blogs, wikis, podcasts, and more have changed the way that people search for, obtain, and share information. This change has had a substantial impact not only on our private lives but also on higher education [13]. Many educators have started to adapt and incorporate technology into their classrooms. Wikis especially have become popular with the development of the Web 2.0 and have gained reasonable attention in higher education. A wiki is a “freely expandable collection of inter-linked web pages, a hypertext system for storing and modifying information – a database where each page is easily editable by any user” [25]. Wikis have been used to support collaborative learning [44], collaborative writing [24], and student engagement [28]. Although wikis are not a new phenomenon in higher education [19], there is still uncertainty about how to integrate wikis into classroom efficiently. As a consequence, instructors are struggling with students that are reluctant to use wikis [11], [14]. At the same time, there is a lack of empirically tested research [18], [27] about the students' perception of using wikis in formal learning processes within higher education.

The goal of this paper is to address this research gap by exploring factors that influence students' decision to adopt wikis within formal learning processes in higher

education. Our research is distinctive for three reasons: first, we use the Decomposed Theory of Planned Behaviour (DTPB) [37] as a theoretical background for this study, which has not been previously tested in this context. Prior research only investigated students' intentions towards the whole software category, Web 2.0 [21], and is therefore limited. Students' decisions to adopt particular software depend on the context: e.g. the intention to voluntarily use social bookmarks in a classroom is different from the intention to write a graded assignment in a wiki. Second, we adapt the DTBP to suit the specific context of using wikis within formal learning processes in higher education. This is necessary because the motivation to use wikis is different from other contexts. Therefore, we included two constructs that represent the influence of intrinsic and extrinsic motivation on students' intention to use wikis. Based on literature, we integrate perceived enjoyment [22] and anticipated rewards [6]. In combination, these constructs allow us to understand the different motivations that are crucial for the adoption of a particular technology. Third, by modelling perceived usefulness as an influence on behavioural intention and by removing attitude, we take previously reported substitution effects from studies [38], [42] into account for the DTPB. The revised model was tested using a survey of first semester students in an introductory course on information systems. We then examined the proposed hypotheses using the partial least square approach to data analysis.

The paper is structured as follows: in the second section, we propose an adapted model of technology acceptance of wikis in higher education. Within the third section, we explain the methodology of our study. The results of the study are presented in the fourth chapter. The fifth chapter discusses the results and shows how instructors could benefit from these results. Finally, the next research steps are outlined.

2 Theoretical Model

The theoretical framework used in this study is based on the Decomposed Theory of Planned Behaviour [37]. The DTPB has its origin in the Theory of Planned Behaviour [2]. The TPB asserts that an individuals' usage behaviour is a direct function of perceived behavioural usefulness and behavioural intention that in turn is a function of attitude, subjective norm, and perceived behavioural control. The DTPB extends the TPB by adding further influence factors on attitude and perceived behavioural control, resulting in more explanatory power [37]. As this study did not test a particular wiki implementation, we omitted usage behaviour because it cannot be measured without using a wiki. Therefore, behavioural intention is used as the strongest predictor of actual use. For this reason, we favoured the DTPB over the Technology Acceptance Model (TAM) [12] because the DTPB allows a better prediction of the behavioural intention than the TAM [37]. The proposed model (see Figure 1) can be used to test influences on wiki adoption without relying on a specific wiki implementation.

As a consequence, we removed perceived ease of use as an influence on perceived usefulness for three reasons. First, we did not test a particular wiki implementation. Although different wiki implementations have basis functionality in common, ease of

use is dependent on a particular wiki implementation. Second, today's students are comfortable with nearly every form of technology [30]. Perceived ease of use is moderated by experience. Therefore, perceived ease of use is not that important in forming behavioural intention if the users are already familiar with the technology [40], [43]. Third, a wiki is web-based application. Perceived ease of use did not appear as a significant determinant when access to a system is provided by a graphical front-end and a browser [1]. Therefore, we removed perceived ease of use in order to get a model that is as parsimonious as possible, but facilitates the understanding of students' decision to adopt wikis [3], [37].

In the following, we introduce an adapted DTPB that is used to study factors influencing students' behavioural intention to use wikis if they are provided within formal learning processes in higher education classrooms.

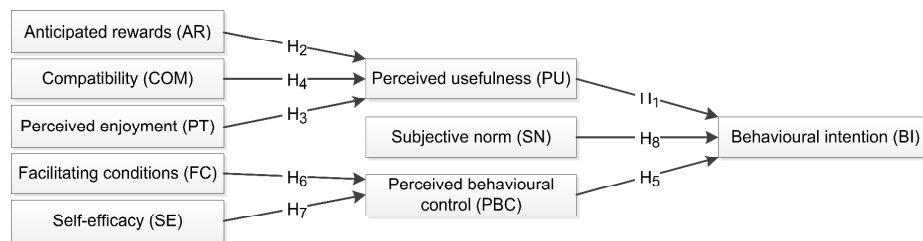


Fig. 1. Research model

2.1 Perceived Usefulness

In contrast to the DTPB, we replaced attitude with perceived usefulness because attitude is not always a reliable predictor of behavioural intention. Attitude is not significant if constructs related to performance and effort expectancies are included in the model [42]. Research results suggest that attitude can be substituted with perceived usefulness [38], [42].

Perceived usefulness (PU) is defined as the “prospective user's subjective probability that using a specific application system will increase his or her job performance within an organizational context” [12]. Previous studies have shown strong empirical support that perceived usefulness positively influences behavioural intention [37], [43]. In the context of wikis in formal learning processes, perceived usefulness is viewed as the degree to which students believe that using wikis will help them learn better. Hence, perceived usefulness is hypothesised to positively influence the behavioural intention.

Hypothesis 1: Perceived usefulness will positively influence students' intentions to use wikis.

The situation in higher education classrooms is different than the situation in small and medium-sized companies, for example. Unlike company employees, students are often not rewarded for their participation in wiki-based assignments if they are not mandatory [14]. Therefore, students seem to be opportunistic, and they will decide to adopt wikis in formal learning processes only if they fit their current situation and

preferences. Students choose an approach to learning by interpreting the “teaching context in the light of their own preconceptions and motivations” [5]. Although the DTPB already takes individual preconceptions (e.g. perceived usefulness) and context (e.g. facilitating conditions) into account, motivations to adopt a particular technology are disregarded. The motivation to use a particular technology depends on their application; e.g. a student’s motivation when using wikis for a graded assignment is probably different from their motivation when commenting on a lecture via a social network service.

Taking this situation into account, we integrated anticipated rewards [6] and perceived enjoyment [22] as new constructs that explain students’ extrinsic and intrinsic motivation. Extrinsic motivation refers to something that is done because it results in a nameable outcome, whereas intrinsic motivation refers to something that is done because it is “inherently interesting or enjoyable” [35].

Anticipated Rewards. Different studies have shown that students do not use wikis unless it is mandatory or if they are rewarded for their work [11], [14]. Hence, we argue that *anticipated extrinsic rewards* (AR) like graded assignments cause a higher degree of perceived usefulness of wikis [45]. Anticipated rewards are defined as the “degree to which one believes that one will receive extrinsic incentives” [6] for using wikis within formal learning processes. Therefore, anticipated rewards reflect the reward-based superior influence on students’ intentions.

Hypothesis 2: Anticipated rewards will positively influence perceived usefulness.

Perceived Enjoyment. The variety of motives to contribute to Wikipedia shows [29] that people do not only participate if they are rewarded. Therefore, there are students who are intrinsically motivated to use wikis. In formal learning processes, these students enjoy using wikis because every participant can read their contributions. This gave them the “feeling of teaching other users and sharing knowledge” [33]. These students “enjoy the process and do not perceive it as being effortful compared to those who have less intrinsic motivation” [41]. Due to its facilitating effect, we included *perceived enjoyment* (PT) as an intrinsic motivator construct that positively influences perceived usefulness.

Hypothesis 3: Perceived enjoyment will positively influence perceived usefulness.

Compatibility. Conformant to the DTPB, we included the *compatibility* (COM) of an information system in the theoretical model. Compatibility is generally regarded as the degree to which an information system is congruous with the potential user’s existing values, previous experiences, and current needs [37]. In this study, we used Chen’s definition of educational compatibility as the degree to which a wiki “complies with the overall learning expectancy of students, including the current learning situation, the learning style, and the preference of conducting learning activities” [8]. Hence, it is expected that an increasing compatibility of the learning style and learning situation with wikis will positively influence the perceived usefulness of wikis in formal learning processes.

Hypothesis 4: Compatibility will positively influence perceived usefulness.

2.2 Perceived Behavioural Control

Individuals do not have complete control over their behaviour in some circumstances. *Perceived behavioural control* (PBC) reflects the level of control individuals feel they have over their own behaviour. It is an important determinant of intention because individuals' behavioural intention is strongly influenced by their perception of the ability to perform it [2], [37]. Regarding students' use of wikis in formal learning processes, perceived behavioural control reflects the students' feeling about their confidence in using wikis and the availability of resources needed to use them.

Hypothesis 5: Perceived behavioural control over wikis will positively influence students' intentions to use wikis.

Facilitating Conditions. The first component influencing perceived behavioural control describes the necessary resources to engage in a behaviour [2], [37]. These *facilitating conditions* (FC) influence the behavioural intentions and the actual use of the technology. In our context, missing resources such as time and technology hinder the use of wikis for formal learning processes.

Hypothesis 6: The availability of facilitating conditions will positively influence perceived behavioural control.

Self-efficacy. *Self-efficacy* (SE) beliefs can influence individuals' behavioural intentions and therefore their actions. This describes an individual's confidence in the ability to perform a behaviour [2], [4]. In the context of wikis in formal learning processes, self-efficacy defines the students' perception of their abilities to use a wiki in higher education classrooms.

Hypothesis 7: Self-efficacy will positively influence perceived behavioural control.

2.3 Subjective Norm

Subjective norm (SN) is defined as the degree to which an "individual perceives that most people who are important to him think he should or should not use the system" [43]. With regard to wikis in formal learning processes, the students' perception of the use of wikis can be influenced by persons important to them. In contrast to the DTPB, we have not distinguished between peer and superior influence because previous studies showed no significant influence of superior influence on students' intentions and perceptions [32], [36].

Hypothesis 8: Subjective norm will positively influence students' intentions to use wikis.

3 Method

To test the model, a survey was conducted to investigate students' intentions to use wikis in higher education classrooms. The survey consisted of a questionnaire developed from material discussed and tested previously [6], [8], [22-23], [37], [39], [43]; see Table 1 for a list of the items. Because this study did not focus on a particular wiki implementation, the students were told to imagine a learning scenario where they collaboratively contribute information to a wiki with their fellow students. The items were slightly modified to suit the context and translated into German. We gave the German items to a peer for back-translation into English to check whether they result in items similar to the originals. Afterwards, wording and translation changes were included in the questionnaire. Each construct was measured using a five-point Likert scale, ranging from "strongly disagree" to "strongly agree".

The study was conducted January 2012 among first semester students enrolled in an introductory course in information systems at a large German university. Participation in the survey was completely voluntary. 425 questionnaires were handed out at the beginning of the lecture and were collected afterwards. 245 questionnaires were returned, at least partly filled out. 133 data records remained after discarding incomplete questionnaires. The participants included 76 males (57.1%) and 57 females (42.9%). Most of the participants were between 18 and 23 years of age ($n = 124$; 93.2%), and the remaining students ($n = 9$; 6.8%) were under 29 years of age. The participants were students of economics with a focus on business administration ($n = 89$; 66.9%), business informatics ($n = 20$; 15%), engineering management ($n = 23$; 17.3%), or business economics and education ($n = 1$; 0.8%).

The collected data was analysed using the partial least squares (PLS) path modelling to assess scales validity and test the hypotheses. This was done using SmartPLS software [34]. PLS is a component-based structural equation modelling technique that has minimal demands on measurements scales, sample size, and residual distributions [9]. We choose PLS because of its minimal requirements regarding sample size and prediction capability [17]. However, the "10 times rule" is fulfilled, specifying the minimum sample size as "10 times the largest number of predictors for any dependent variable in the model" [17]. As the "10 times rule" is only a minimum requirement, we calculated the statistical power according to Cohen using G*Power [10], [15]. The number of cases is sufficient to detect relationships of a medium effect size with a power of 95% ($n = 119$). This statistical power is regarded as sufficient because the hypotheses were previously tested in other contexts. It is therefore likely that small effect sizes were discovered.

Table 1. Summary of items and factor loadings

<i>Construct</i>	<i>Item</i>	<i>Loading</i>	
Anticipated rewards [6]	AR1	I will receive a better grade in return for my participation in wikis.	.93
	AR2	I will receive a chance for a better grading in return for my participation in wikis.	.94
Behavioural intention [43]	BI1	Assuming I had access to wikis, I intend to use them.	.94
	BI2	Given that I had access to wikis, I predict that I would use them.	.97
	BI3	I plan to use wikis in my studies if they are provided.	.97
Compatibility [8]	COM1	Using wikis is compatible with all aspects of my learning.	.90
	COM2	Using wikis is completely compatible with my current learning situation.	.93
	COM3	I think using wikis fits well with the way I like to conduct learning activities.	.93
	COM4	Using wikis fits into my learning style.	.89
Facilitating conditions [39]	FC1	When I need help to use wikis, guidance is available to me.	.92
	FC2	When I need help to use wikis, specialized instruction is available to help me.	.89
	FC3	When I need help to use wikis, a specific person is available to provide assistance.	.91
Perceived behavioural control [37]	PBC1	I would be able to use wikis.	.93
	PBC2	Using wikis is entirely within my control.	.91
	PBC3	I have the resources and the knowledge and the ability to make use of wikis.	.94
Perceived enjoyment [22]	PT1	While participating in wikis, I experienced pleasure.	.91
	PT2	The process of participating in wikis is enjoyable.	.86
	PT3	I have fun using wikis.	.88
Perceived usefulness [43]	PU1	Using wikis improves my learning efficiency/performance.	.88
	PU2	Using wikis in increases my productivity.	.89
	PU3	Using wikis enhances my effectiveness.	.93
	PU4	I find the system to be useful in my tasks.	.82
Self-efficacy [23]	SE1	I have confidence in my ability to provide knowledge that other students consider valuable.	.92
	SE2	I have the expertise needed to provide valuable knowledge for other students.	.94
	SE3	Most fellow students think that the knowledge I transfer is valuable to them.	.86
Subjective norm [43]	SN1	People who are important to me think that I should participate in wikis.	.96
	SN2	People who influence my behaviour encourage me to participate in wikis.	.96

4 Results

PLS path models are interpreted in two-steps process, encompassing (1) the assessment of the reliability and the validity of the measurement model, and afterwards, (2) the evaluation of the structural model that explains hypothesized causal paths.

4.1 Measurement Model

Each construct was measured using reflective indicators. To evaluate the reliability and validity of the measurement model, we assessed the convergent validity and the discriminant validity of the scale items.

Convergent validity was assessed using three criteria: internal consistency, indicator reliability, and average variance extracted (AVE). To ensure internal consistency, Cronbach's alpha α_c and internal composite reliability ρ_c should be greater than .70 [9]; both thresholds were exceeded for all constructs, see Table 2 for ρ_c . Table 1 lists the constructs, the related items, and the factor loadings. Indicator reliability can be assumed because each indicator loads high ($> .80$) on the related construct [9]. Finally, every AVE exceeded the suggested threshold of .50 [9].

Table 2. Descriptive statistics, correlation of constructs, composite reliability ρ_c , and AVE

	<i>M</i>	<i>SD</i>	ρ_c	<i>Inter-construct correlations</i>									
				AR	BI	COM	FC	PBC	PT	PU	SE	SN	
AR	2.46	1.16	.93	.87 (.93)									
BI	3.36	1.18	.97	.22	.92 (.96)								
COM	3.17	1.04	.95	.37	.72	.83 (.91)							
FC	2.77	1.04	.94	.45	.43	.56	.83 (.91)						
PBC	3.47	1.15	.95	.27	.67	.74	.51	.86 (.93)					
PT	2.55	1.04	.92	.37	.36	.51	.53	.41	.78 (.89)				
PU	3.14	1.00	.93	.52	.70	.80	.59	.68	.66	.78 (.88)			
SE	2.68	1.06	.93	.50	.39	.47	.49	.52	.57	.56	.82 (.91)		
SN	2.40	1.15	.96	.46	.29	.48	.50	.34	.53	.51	.44	.92 (.96)	

Note: Diagonal elements are the average variance extracted (AVE) and, in parenthesis, the square root of the AVE.

Discriminant validity is achieved if the conceptually different constructs exhibit sufficient difference. Therefore, the factor loading of each indicator is expected to be greater than all of its cross loadings [9], and the AVE of a construct should be higher than the constructs' highest squared correlation with any other construct [16], as shown in Table 2. Both criteria satisfactorily fulfilled the requirements, demonstrating discriminant validity.

The results of the model evaluation demonstrated that satisfactory reliability, convergent validity, and discriminant validity were achieved. Therefore, all scales in this study sufficiently measured the related constructs.

Finally, we addressed concerns regarding common method bias by using a statistical approach suggested by Podsakoff et al. and by following a method proposed by Liang et al. [26], [31]. The average of indicator variance caused by substantive constructs (0.85) was substantially greater than the method-based variance (0.01). Additionally, most of the method factor loadings were insignificant. Common method bias is thus not a serious concern of this study.

4.2 Structural Model

The hypotheses were tested with SmartPLS [34]. We used the bootstrapping method to determine the significance of the paths among the constructs. As recommended, we used the number of valid observations ($n = 133$) as the number of bootstrap cases, 5,000 bootstrap samples, and selected the individual sign changes option [20]. Therefore, we derived significance for item loadings and path coefficients β by using the t -statistic.

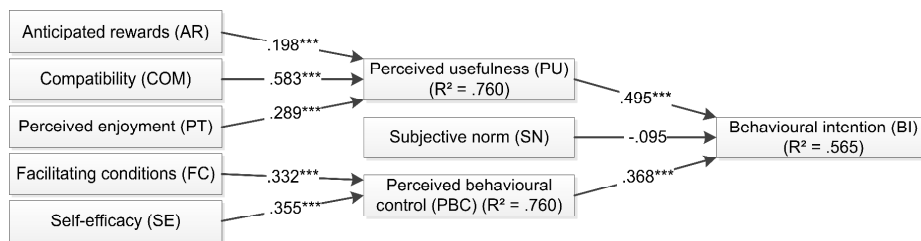


Fig. 2. PLS path analysis model (***) $p < .001$

The R^2 values of the endogenous constructs indicate the percentage of variance explained by the model and therefore give information about the explanatory power of the structural model. According to the thresholds denoted by [9], the R^2 of behavioural intention ($R^2 = .565$) and perceived behavioural control ($R^2 = .352$) is moderate, though perceived usefulness ($R^2 = .760$) had a substantial level [9]. All structural paths were found to be significant except one. The R^2 values of the endogenous variables and the significance of the modelled paths are depicted in Figure 2. Additionally, we calculated the effect size f^2 , which can be explored to see the impact of an exogenous variable on an endogenous variable. The impact at the structural model can be considered a small ($f^2 = .02$), medium ($f^2 = .15$), or large ($f^2 = .35$) effect [9]; for effect sizes of the paths see Table 3. The predictive capabilities of the proposed model were tested using cross-validated redundancy measure Q^2 . Each Q^2 value was greater than zero; therefore, the model can be seen to have predictive relevance [9].

As expected, perceived usefulness had a significant and positive influence on the behavioural intention to use a wiki in formal learning processes. The effect size on the intention was medium ($f^2 = .202$). Therefore, hypothesis H_1 ($\beta = .495$; $p < .001$) was supported. The proposed positive influence of anticipated rewards on perceived usefulness (H_2 , $\beta = .198$; $p < .001$) was significant, the effect was on a good small level ($f^2 = .114$), and therefore the hypothesis was supported. The proposed positive influ-

ence of perceived enjoyment on perceived usefulness (H_3 , $\beta = .289$; $p < .001$) was significant and had a medium effect ($f^2 = .195$). Thus, hypothesis H_3 was supported. The positive influence of compatibility on perceived usefulness (H_4 , $\beta = .583$; $p < .001$) was supported significantly. Hence, a high compatibility with the students' learning situation increases the perceived usefulness. The effect size had a large effect ($f^2 = .497$) as well as the highest influence on perceived usefulness.

Table 3. Path coefficients β , t-statistic, and effect size f^2

<i>Hypotheses</i>		β	<i>t-statistic</i>	f^2
H ₁	PU → BI	.495***	5.586	.202
H ₂	AR → PU	.198***	3.393	.114
H ₃	PT → PU	.289***	5.006	.195
H ₄	COM → PU	.583***	10.406	.497
H ₅	PBC → BI	.368***	4.072	.166
H ₆	FC → PBC	.332***	4.162	.114
H ₇	SE → PBC	.355***	4.223	.128
H ₈	SN → BI	-.095	1.606	

Note: * $p < .05$ ** $p < .01$ *** $p < .001$.

Consistent with the DTPB, hypotheses H_5 , H_6 , and H_7 were supported. The influence of perceived behavioural control on behavioural intention is positive, significant (H_5 , $\beta = .368$; $p < .001$), and had a medium effect ($f^2 = .166$) on behavioural intention. The hypothesised positive influence of facilitating conditions on perceived behavioural control was significant (H_6 , $\beta = .332$; $p < .001$). However, the effect size was only small ($f^2 = .114$). Finally, the proposed positive influence of self-efficacy on perceived behavioural control was also significant (H_7 , $\beta = .355$; $p < .001$), but has only a small effect ($f^2 = .128$). Unexpectedly, subjective norm had no significant influence on behavioural intention. Hence, hypothesis H_8 was not supported.

To ensure that no significant paths have been left out of the model, we compared it to the saturated model. A saturated model connects all exogenous variables with the endogenous variable, whereas the theoretical model only includes the hypothesised paths. Both models were compared with each other in order to verify “(1) that the significant paths in the theoretical model also remain significant in the saturated model, and (2) that adding the paths via the saturated model does not significantly increase the f^2 ” [17]. We compared the adjusted R^2 of the proposed theoretical model (adjusted $R^2 = .555$) with the saturated one (adjusted $R^2 = .606$). All relationships stayed significant, and the effect size of the additional paths was small ($f^2 = .129$). Therefore, it is unlikely that any significant path has been left out in the theoretical model.

5 Discussion

In this study, we adapted the DTPB to explore the factors that influence students' decision to adopt wikis within formal learning processes in higher education. Consistent with previous research, perceived usefulness (e.g. [12]) and perceived behavioural control (e.g. [37]) had significant influence on behavioural intention. In conjunction with subjective norm, perceived usefulness and perceived behavioural control explain 56% of the variability of behavioural intention. In contrast to previous studies (e.g. [43]), subjective norm had no significant influence on behavioural intention. This finding is counterintuitive as students are spending a lot of time together while studying. Accordingly, it seems reasonable to expect them to influence each other's attitude towards technology for learning. However, the argument may not be relevant for first semester students, as there may not have formed personal relationships and learning habits yet. The lack of mutual influence could explain why subjective norm had no influence on students' behavioural intention. Accordingly, instructors' influence on first semesters should also be limited.

By integrating anticipated rewards and perceived enjoyment as constructs that represent intrinsic and extrinsic motivation, the proposed model considers the specific situation in higher education classrooms. The results show that both constructs were antecedents of perceived usefulness. In conjunction with compatibility, anticipated rewards and perceived enjoyment explain 76% of the variability of perceived usefulness. As hypothesised, anticipated rewards and perceived enjoyment had a positive influence on perceived usefulness. Therefore, students perceive wiki assignments as more useful if they are rewarded (e.g. with grades) or if they enjoy working with wikis. Compatibility had a large and significant effect on perceived usefulness. This is consistent with findings by Chen [8], who shows that educational compatibility is more important than the perceptions regarding technology usage. Therefore, the larger the compatibility with the learning situation, the larger is the perceived usefulness.

Consistent with the DTPB [37], facilitating conditions and self-efficacy both had a small, significant influence on perceived behavioural control.

5.1 Practical Implications

The underlying question of this study is what can we do as instructors to explain our students' lack of motivation? What can we do to engage our students and encourage them to use wikis? In the following, we outline some implications of the results.

Anticipated rewards, perceived enjoyment, and compatibility were strong determinants of perceived usefulness. The large influence of compatibility calls attention to students' learning preferences and learning style. Instructors have to think carefully about how wikis can support students' learning styles rather than changing them. While someone's learning style does not change in a short time, instructors can accommodate students with a course assignment that fits their learning situation. Students use wikis if they fit their goals, whether they are intrinsically or extrinsically

motivated. But in either case, a wiki assignment must fit the learning situation of the student. Therefore, wikis should be an integral part of the course assignment rather than just an additional duty. Thereby, it is more likely that students will engage with a wiki if the course assignment necessitates it. As a consequence, instructors should ask themselves if a wiki is beneficial for a certain task assignment – and opt for an alternative if the task does not benefit from using a wiki.

While instructors cannot change their students' learning style, they can abet wiki adoption by targeting students' motivation. One precondition for the students to use wikis in class is that wiki assignments are integrated into the course in a reasonable and rewarding way. Otherwise, students will abstain from using the wiki [11], [14]. Through mandatory and assessed wiki work, extrinsic rewards can be set easily. Alternatively, instructors can try to motivate students by conferring certificates after taking part in a wiki-based course, instead of forcing them into wiki use by rewarding them with a grade [7].

Ideally, students are motivated extrinsically and intrinsically. Perceived enjoyment refers to one's intrinsic motivation. As perceived enjoyment had a significant influence on perceived usefulness, students seem to like the idea of working together within a wiki. This is consistent with previous qualitative research about the use of wikis in higher education. Students like the feeling of teaching others and sharing their knowledge [33]. And although intrinsic motivation cannot be built by instructors, they can design task assignments that stimulate intrinsic motivation. For example, working for a greater audience can motivate students: like writing a textbook together that will be published as an open educational resource.

Independent of students' source of motivation, instructors should consider students' learning preferences and learning styles by seamlessly integrating wikis into the course. Thereby, instructors would not only lower barriers of adoption and invite students' participation and engagement in the wiki, but would also provide facilitating conditions. As facilitating conditions and self-efficacy had a significant influence on perceived behavioural control, accompanying wiki training and continuous support to lower technology barriers is recommended.

5.2 Limitations and Further Research Directions

From a research perspective, the study results indicate the suitability of the proposed model to explain the influences on the use of wikis in formal learning processes within higher education by students. However, this study has some limitations.

First, we did not have the opportunity to collect data from a random sample of students. A convenience sample was used to test the model (first semester students in an introductory course in information systems). In order to generalize the findings, this study should be replicated with students from different study courses and semesters.

Second, the explanatory power of this model regarding the influence of subjective norm on behavioural intention has to be further examined. Although subjective norm has been proven as a reliable influence in various acceptance models [37], [43], this was not the case in this study. Perhaps this is due to the fact that the subjective norm was measured with only two reflective indicators and without distinguishing between

different influence groups (e.g. peer influence, superior influence). An alternative explanation is that peers do have little influence on students' intentions and superior influence was substituted by anticipated rewards. Therefore, more qualitative and quantitative research is necessary to explore further social influences on technology adoption in the context of higher education.

Third, we refrained from including wiki characteristics in the model, but provided contextual information as a preamble to the questionnaire. While some will see this as a limitation, it facilitates the comparison of different educational technologies by shifting technology-specific characteristics into the background and focusing on the adoption of a form of technology in a particular setting. Further research should therefore investigate whether the proposed model yields different results with other Web 2.0 applications (e.g. social networking services, weblogs).

6 Conclusion

The results of this study provide a foundation for future research about factors that influence student use of wikis in higher education. We show that intrinsic and extrinsic motivation plays an important role in students' decision to participate and engage in wiki assignments. Based on these findings, researchers should examine methods to foster support for student use of wikis, as well as for other Web 2.0 applications. This would enable instructors to better address students' needs and preferences.

Acknowledgements. We would like to thank Richard Ostertag who developed a theoretical model to examine students' intentions to use Web 2.0 tools within his diploma thesis, which we used as the foundation of the model outlined above. Furthermore, we would like to thank our colleagues for providing inspiration and helpful advice; esp. Helena Bukvova and Hendrik Kalb who commented on this paper.

References

1. Agarwal, R., Prasad, J.: The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences* 28 (3), 557–582 (1997)
2. Ajzen, I.: The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50 (2), 179–211 (1991)
3. Bagozzi, R.P.: The self-regulation of attitudes, intentions, and behavior. *Social Psychology Quarterly* 55 (2), 178–204 (1992)
4. Bandura, A.: Self-efficacy mechanism in human agency. *American Psychologist* 37 (2), 122–147 (1982)
5. Biggs, J.: Approaches to the enhancement of tertiary teaching. *Higher Education Research & Development* 8 (1), 7–25 (1989)
6. Bock, G.-W. et al.: Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly* 29 (1), 87–111 (2005)
7. Bonk, C.J. et al.: The tensions of transformation in three cross-institutional wikibook projects. *Internet and Higher Education* 12 (3-4), 126–135 (2009)

8. Chen, J.-L.: The effects of education compatibility and technological expectancy on e-learning acceptance. *Computers & Education* 57 (2), 1501–1511 (2011)
9. Chin, W.W.: The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (ed.): *Modern methods for business research*. Lawrence Erlbaum, Mahwah, NJ (1998)
10. Cohen, J.: *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale (1988)
11. Cole, M.: Using wiki technology to support student engagement: Lessons from the trenches. *Computers & Education* 52 (1), 141–146 (2009)
12. Davis, F. et al.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science* 35 (8), 982–1003 (1989)
13. Downes, S.: E-Learning 2.0, <http://elearnmag.acm.org/featured.cfm?aid=1104968>
14. Ebner, M. et al.: Utilizing wiki-systems in higher education classes: A chance for universal access?. *Universal Access in the Information Society* 7 (4), 199–207 (2008)
15. Faul, F. et al.: Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41 (4), 1149–1160 (2009)
16. Fornell, C., Larcker, D.F.: Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research* 18 (3), 382–388 (1981)
17. Gefen, D., Rigdon, E.E.: An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly* 35 (2), iii–xiv (2011)
18. Guo, Z., Stevens, K.J.: Factors influencing perceived usefulness of wikis for group collaborative learning by first year students. *Australasian Journal of Educational Technology* 27 (2), 221–242 (2011)
19. Guzdial, M. et al.: Beyond adoption to invention: Teacher-created collaborative activities in higher education. *Journal of the Learning Sciences* 10 (3), 265–279 (2001)
20. Hair, J.F. et al.: An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science* 40 (3), 414–433 (2011)
21. Hartshorne, R., Ajjan, H.: Examining student decisions to adopt Web 2.0 technologies: Theory and empirical tests. *Journal of Computing in Higher Education* 21 (3), 183–198 (2009)
22. Hsu, C.-L., Lin, J.C.-C.: Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management* 45 (1), 65–74 (2008)
23. Kang, M. et al.: Identifying different antecedents for closed vs open knowledge transfer. *Journal of Information Science* 36 (5), 585–602 (2010)
24. Kessler, G.: Student-initiated attention to form in wiki-based collaborative writing. *Language Learning & Technology* 13 (1), 79–95 (2009)
25. Leuf, B., Cunningham, W.: *The Wiki Way: Quick collaboration on the web*. Addison-Wesley, Upper Saddle River, NJ (2001)
26. Liang, H. et al.: Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly* 31 (1), 59–87 (2007)
27. Liu, X.: Empirical testing of a theoretical extension of the technology acceptance model: An exploratory study of educational Wikis. *Commun. Education* 59 (1), 52–69 (2010)
28. Neumann, D.L., Hood, M.: The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course. *Australasian Journal of Educational Technology* 25 (3), 382–398 (2009)

29. Nov, O.: What motivates Wikipedians? *Communications of the ACM*. 50 (11), 60–64 (2007)
30. Pence, H.E.: Preparing for the real web generation. *Journal of Educational Technology Systems* 35 (3), 347–356 (2007)
31. Podsakoff, P.M. et al.: Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88 (5), 879–903 (2003)
32. van Raaij, E.M., Schepers, J.J.L.: The acceptance and use of a virtual learning environment in China. *Comput. Educ.* 50 (3), 838–852 (2008)
33. Ravid, G. et al.: Wikibooks in higher education: Empowerment through online distributed collaboration. *Comput. Hum. Behav.* 24 (5), 1913–1928 (2008)
34. Ringle, C.M. et al.: SmartPLS 2.0 (beta), <http://www.smartpls.de> (2005)
35. Ryan, R.M., Deci, E.L.: Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25 (1), 54–67 (2000)
36. Sun, P.-C. et al.: What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education* 50 (4), 1183–1202 (2008)
37. Taylor, S., Todd, P.A.: Understanding information technology usage: A test of competing models. *Information Systems Research* 6 (2), 144–176 (1995)
38. Teo, T.: Is there an attitude problem? Reconsidering the role of attitude in the TAM. *British Journal of Educational Technology* 40 (6), 1139–1141 (2009)
39. Teo, T.: Modeling the determinants of pre-service teachers' perceived usefulness of e-learning. *Campus-Wide Information Systems* 28 (2), 124–140 (2011)
40. Venkatesh, V.: Determinants of perceived ease of use: Integrating perceived behavioral control, computer anxiety and enjoyment into the technology acceptance model. *Information Systems Research* 11 (4), 342–365 (2000)
41. Venkatesh, V. et al.: User acceptance enablers in individual decision making about technology: Toward an integrated model. *Decision Sciences* 33 (2), 297–316 (2002)
42. Venkatesh, V. et al.: User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27 (3), 425–478 (2003)
43. Venkatesh, V., Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences* 39 (2), 273–315 (2008)
44. Wheeler, S. et al.: The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology* 39 (6), 987–995 (2008)
45. Witney, D., Smallbone, T.: Wiki work: Can using wikis enhance student collaboration for group assignment tasks?. *Innovations in Education and Teaching International* 48 (1), 101–110 (2011)

Effects of the Network Structure on the Dynamics of Viral Marketing

Marek Opuszko and Johannes Ruhland

Friedrich-Schiller-University, Jena, Germany
{marek.opuszko, johannes.ruhland}@uni-jena.de

Abstract. We present an analysis of how the structure of a social network influences the diffusion of information in a viral marketing context. We performed diffusion simulations on a large number of real world and artificially generated network datasets. We analyze how the characteristics of a network and parameter settings like the selection of initial start nodes influences the diffusion. The results indicate that the network structure has a significant effect on the diffusion. Extreme cases show a difference in the diffusion of over 65%. Our investigation also proves that a viral marketing diffusion may be predicted without the knowledge of the whole network. We further provide useful recommendations for marketers which could be taken into consideration when marketing campaigns are conducted.

Keywords: viral marketing, information diffusion, social networks

1 Introduction

The success of virtual social communities on the web and the increasing resistance and avoidance of customers towards traditional forms of advertising [1] led marketers to turn to new forms of marketing such as social media marketing and viral marketing as a special type of electronic word-of-mouth marketing (WOM). Customers nowadays can easily share information including product information with their friends online, leading to a change in the information diffusion process within social peer groups. Solomon [2] states that WOM effects play an important role in the formation of judgments and attitudes towards innovations. Traditional marketing forms, on the other hand, are increasingly considered irrelevant by consumers [3]. Another important aspect of WOM and, in particular, viral marketing is the fact that the diffusion process may operate at very low costs, using an underlying social network and the participation of the network nodes themselves to actively spread an information artifact. With the increasing success of online communities, we can observe numerous examples of “viral” products, websites or user generated content. One of the most famous examples is the viral marketing campaign for the movie *The Blair Witch Pro-*

ject. The monetary success of nearly 250 million US dollars exceeded the production costs of 60,000\$ US by the factor of 4166¹.

Since viral marketing promises great effects at low costs, marketers and researchers showed interest in the understanding of the underlying processes. A main goal was to examine whether an information or innovation spreading dies out quickly or diffuses significantly into the population. Unfortunately these effects are difficult to observe in real world settings, as they usually appear spontaneous [4]. Following Bampo et al. [5], three main components determine the viral process: (1) the behavioral characteristics of the members of the network, (2) the seeding strategy and (3) the structure of the social network. Hinz et al. [6] added a fourth critical factor, (4) the attractiveness of the content supposed to be transported. Since factor (2), the seeding strategy, is the only parameter that is largely under the control of the campaign initiator, this issue has received much attention in different research fields. Seeding strategies usually aim to identify influential nodes as the initial set of nodes in a diffusion process.

Kempe and colleagues theoretically formalized this NP-Complete problem as the diffusion maximization problem [7]. Numerous strategies have been introduced and examined, mainly in the computer science and marketing community [6-10], still leaving some controversy and contradictory findings about the impact of seeding and viral marketing [6], [11]. According to a study among marketers in 2007, the major perceived problem of viral marketing is still the missing experience and a lack of measurability of the advertising effect [12]. Many factors of viral marketing diffusion still remain imprecise and vague. Leskovec et al. showed in their analysis that viral marketing is not as epidemic as hoped. They further stated that the topology of the addressed networks (factor (3)) should be analyzed in more detail [13]. Beside the conventional wisdom that the network structure influences the diffusion of information, its role has not been deeply investigated. Watts and Strogatz [14] showed that networks comprising high clustering and small path lengths, what they called “Small World Networks”, facilitate an epidemic spread. Bampo et al. found with a simulation of a real viral marketing campaign in generated networks, that the structure of a social network has a significant impact on a viral marketing campaign performance [5]. Interestingly they found that scale-free networks, first introduced by Barabási and Albert [15], are very efficient for viral campaigns and small world networks generally temper the spread of information. Shakarian and Paulo investigated on viral marketing diffusion simulations in numerous networks and observed three different types of networks, each type showing a very distinct diffusion pattern [10]. To date, no research offers extensive investigations on what network characteristics exactly influence the information diffusion in social networks, and if, to what extent. Moreover, the majority of the research on seeding is based on investigations on very few network datasets only [6-8], [16]. The findings are therefore based on a specific network structure and may not be transferable to any other viral marketing application in networks with a different structure.

Another problem is that for all the diffusion maximization approaches the whole network structure must be known *ex ante*, including every connection between any

¹ <http://boxofficemojo.com/movies/?id=blairwitchproject.htm>

two nodes. Obviously, this is not the case in many real world marketing applications. Usually only few companies, mainly the service providers themselves like Facebook or Google, have access to all information about the underlying network. Nevertheless, some information is available and assumptions about the network structure can be made. We know for instance that in 2011 an average Facebook user has 130 friends and posted 90 pieces of content every month². We further know that a network like Facebook comprises more than 150 million users in the US³ in 2012 and that the network in 2011 showed a degree of separation of only 3.74⁴.

To address the open questions we will undertake a simulation analysis on a set of real-world and artificially generated networks using different types of seeding methods and two state-of-the-art diffusion models. We will introduce a set of metrics that have been calculated to characterize every network. These metrics will be combined with the diffusion simulation results to answer the question what network characteristics influence the message diffusion. This also includes the interaction between the seeding method and the network characteristics. We will further conduct a decision tree analysis to answer the question whether the diffusion can be predicted based on the characteristics about the network structure or not. We conclude our work with a discussion of the results, managerial implications and future work perspectives.

2 Theoretical Framework

The research of diffusion of information in large online communities mainly derived from the studies of epidemics [17]. In other contexts, early efforts have been made in the understanding of the adoption of innovations in the medical and agricultural sector [18-19] or the success of product innovations [20-22]. To describe the diffusion, several models have been introduced in the past. Our investigation will rely on two standard models in this research field, the *independent cascade model* and the *linear threshold model* which reflect a different transmission behavior of the network nodes. These models are well understood and have been used in various works studying the diffusion of information in social networks [7], [13], [23-24].

The linear threshold model, as first proposed by Granovetter [25], is based on the assumption of node specific thresholds. We speak of *active* nodes if a network node has adopted an innovation or received information, e.g. a marketing message. According to the model, a node is activated if the number of active nodes in the neighborhood reaches a threshold. A real world example would be the adoption of fashion or technology trends. In the context of virtual communities like Facebook, we can observe similar aspects. In Facebook, for instance, little messages like “*Your friend XYZ is now using ...*” are presented on the starting screen after a user logs in, highlighting friends using apps or games. One could make the assumption that some users might

² <http://www.kenburbary.com/2011/03/facebook-demographics-revisited-2011-statistics-2/>

³ <http://www.socialbakers.com/facebook-statistics/>

⁴ <http://www.bbc.co.uk/news/technology-15844230>. The degree of separation refers to the average number of friends in chain from one user to any other user in the network, excluding the end nodes.

also start using a certain app (usually a game) after a critical amount of friends did it before⁵. According to Kempe [7] the process works as follows: an inactive node v is activated if the total weight b_w of any of its active neighbors w reaches the threshold θ_v (in the interval $[0,1]$):

$$\sum_w b_w \geq \theta_v \quad (1)$$

The diffusion process unfolds as a deterministic process in discrete steps t , with all nodes active in $t-1$ still active in t , given an initial set of active nodes A_0 at the beginning of the diffusion process. Each node uniformly chooses a threshold θ_v or the threshold is hard-wired for every node. The process will stop, if no further activations have been made.

Based on the idea of interacting particle systems but later used in the context of marketing [26], the independent cascade model can also be used to model diffusion processes in networks. It starts with a set of initial nodes A_0 and unfolds in discrete time steps t . Unlike the linear threshold model, an active node v is given a single chance to activate any inactive neighbor node w in the time step succeeding its activation. Whether or not an activation attempt is successful, node v cannot make any further attempts to activate w in future steps. The activation is based on an activation probability $p_{v,w}$ defined for every node prior to the diffusion start. The process runs until no further activations are possible. In the context of virtual online communities the independent cascade model could be compared with posting, commenting or sharing content on a friends pin wall in online social networks like Facebook or sending information artifacts to friends per email. If a friend is “activated”, the message might be forwarded to further people, etc. Moreover, platforms like Facebook reinforce this process by automatically⁶ sharing every status update a user makes with every friend (neighbor) of the user. These status updates include pin wall posts, using the *like* or *recommend* function or using games and apps.

3 Method

To evaluate the influence of the network structure on the information diffusion we ran simulations on networks using the independent cascade and the linear threshold model with different parameter settings. The simulation process was conducted as follows. For every network dataset 200 runs of simulations with randomly chosen start parameters were conducted. For every run the number of start nodes was set at random in the interval $[1,50]$. For the independent cascade model the activation probability for all nodes was set randomly in the interval $[0,1]$. For the linear threshold model the node threshold has been set equivalently. Furthermore, the criterion for choosing the start nodes (described in the next subsection) was selected randomly and the nodes have been selected accordingly by calculating the metrics for all network nodes. For

⁵ There are plenty of examples like the very popular farmville game on facebook, having millions of active users: <http://mashable.com/2010/02/20/farmville-80-million-users/>

⁶ A user on Facebook may change this functionality by using individual privacy settings.

every run, 50 single simulations were accomplished. The resulting diffusion was averaged and the resulting diffusion mean and the diffusion variance were stored including all used parameters and the network metrics as one case. All in all 720,000 simulations were conducted. For each diffusion model a dataset with 7200 cases each comprising the network characteristics, the simulation parameters and the results was created as a base for the analysis.

3.1 Seeding Criteria for the Selection of Initial Start Nodes

Several criteria for selecting the initial active nodes A_0 exists. The most simple would be by randomly activating a set of nodes. A common method is to evaluate the centrality of every node in the network based on different centrality measures and to choose the most *central* nodes as those are supposed to be influential [23]. We will evaluate the following centrality measures as a criterion for choosing initial nodes based on a given graph $G = (V, E)$ comprising network nodes (vertices) V and edges E . All centrality metrics have been calculated according to Wasserman and Faust [27] as well as Newman [28]. We refer to those publications for further details on the calculations.

- *Degree Centrality*, one of the most common centrality measures. Degree centrality reflects the number of ties (aka neighbors or friends in the context of online social networks) of a node.
- *Betweenness Centrality*, this centrality reflects the probability of a node to lie on a shortest path between two randomly chosen nodes.
- *Closeness Centrality*, this reflects to the inverse farness of a node v to any other node in the network.
- *Eigenvector Centrality*, a natural extension of the degree centrality. The difference is that nodes also award “points” for the degree centrality of their neighbors. A node is central if it is connected to other *important* nodes.
- *Node Clustering Coefficient*, sometimes also referred to as transitivity. The clustering coefficient of a node is the relation of the number of pairs of neighbors that are connected to the number of pairs of neighbors. In online social networks this reflects to the connection among a users’ friends.
- *PageRank Coefficient*, extension of the eigenvector centrality used by Google to rank the centrality of web pages [29]. The difference is that the centrality of a node is further divided by the out-degree of a node.
- *Random*, the initial start nodes are chosen randomly.

3.2 Network Metrics

To describe a network’s structure and characteristics and to later evaluate diffusion predictors, we calculated several metrics to describe a network.

- *Number of network nodes*, usually the number of users.
- *Number of network edges*, the number of connection between the nodes.

- *Network density*, the density of a network is the relation between existing and possible edges.
- *Connected graph* (yes | no), a graph is connected if all nodes belong to one (giant) component and no individual clusters exist.
- *Average path length*, the average of all shortest paths between any two nodes of the network.
- *Number of components/clusters*, number of isolated components comprising at least two nodes.
- *Network/Graph diameter*, the diameter of a network is the length of the longest shortest path between two arbitrary nodes in the network.
- *Average node degree*, a normalized $([0,1])$ metric of the degree centrality of all nodes.
- *Average node betweenness*, a normalized $([0,1])$ metric of the betweenness centrality of all nodes.
- *Average node closeness*, a normalized $([0,1])$ metric of the closeness centrality of all nodes.
- *Average node eigenvector*, a normalized $([0,1])$ metric of the eigenvector centrality of all nodes.
- *Average clustering coefficient*, a normalized $([0,1])$ coefficient of the clustering coefficient of all nodes.
- *Number of network communities*, communities are sub graphs or dense groups of nodes within a network which are sparsely connected to other groups. In opposite to components, these groups are not isolated from each other. We used the leading eigenvector community detection algorithm according to Newman [30].
- *Degree distribution power law fit*, since the degree distributions of network nodes often show a power law distribution, we fitted a power-law distribution with maximum likelihood methods as recommended by Newman against the degree distribution of each network [31].

3.3 Network Datasets

We used both real world and artificially generated networks as a source for the simulations. Table 1 shows the used networks for the simulation. The real world networks include some of the most common datasets used in social network analysis. One exception is the dataset Student Network. This network has been extracted in a former analysis. It comprises a Facebook friendship network of university freshmen after their first semester of study. It should be noted that prior to the analysis, all isolated nodes have been deleted from the graphs. To generate the artificial networks, three state-of-the-art algorithms have been used: Erdős-Renyi game [41], Watts-Strogatz game [14] and Barabási-Albert game [15]. The artificial networks have been generated in order to represent the characteristics of the real world datasets in terms of node and edge count. All calculations have been done using the igraph [42] package in the R software [43].

Table 1. Networks used in the analysis

Name	Nodes	Edges
Social network of Dolphin interactions [32]	62	159
Coappearance of Les Miserables (novel) characters [33]	77	254
Topology of US Western States power grid [14]	4,941	6,594
Facebook university freshmen (Student Network)	471	926
Snapshot of Gnutella peer-to-peer network 2008 [34]	6,301	20,777
Social network of the University of California (OCLinks) [35]	1,899	20,297
Coauthorship network of network scientists (NetScience) [36]	1,461	2,742
A snapshot of the structure of the Internet ⁷	22,963	48,436
High-Energy Theory coauthorship network (Hep-th) [37]	7,610	15,751
Condensed matter E-Print coauthorship network 2003 [37]	30,460	120,029
Condensed matter E-Print coauthorship network 2005 [37]	39,577	175,639
Erdős collaboration graph ⁸	6,927	11,850
Astrophysics E-Print coauthorship network [37]	16,046	121,251
Network of Email interchanges [38]	1,133	5,451
Network of Jazz musicians [39]	198	2,742
Network of users of the Pretty-Good-Privacy algorithm [40]	10,680	24,316
Network created according to Barabási-Albert 1 [15]	60	177
Network created according to Barabási Albert 2 [15]	80	237
Network created according to Barabási Albert 3 [15]	1,400	2,798
Network created according to Barabási Albert 4 [15]	20,000	39,998
Network created according to Barabási Albert 5 [15]	30,000	119,996
Network created according to Erdős-Rényi 1 [41]	868	1,040
Network created according to Erdős-Rényi 2 [41]	914	12,683
Network created according to Erdős-Rényi 3 [41]	1,000	14,902
Network created according to Erdős-Rényi 4 [41]	1,000	25,362
Network created according to Erdős-Rényi 5 [41]	6,290	19,996
Network created according to Erdős-Rényi 6 [41]	6,917	23,767
Network created according to Watts-Strogatz 1 [14]	60	177
Network created according to Watts-Strogatz 2 [14]	80	240
Network created according to Watts-Strogatz 3 [14]	60	177
Network created according to Watts-Strogatz 4 [14]	1,400	2,800
Network created according to Watts-Strogatz 5 [14]	20,000	40,000
Network created according to Watts-Strogatz 6 [14]	30,000	120,000
Network created according to Watts-Strogatz 7 [14]	6,927	14,994

⁷ <http://www-personal.umich.edu/~mejn/netdata/>⁸ <http://vlado.fmf.uni-lj.si/pub/networks/data/>

4 Analysis

4.1 Descriptive Analysis.

Figure 1 shows the resulting diffusion means of all networks in relation to the activation probability and threshold. The plot highlights a high variance in the diffusion curve depending on the underlying network. As some diffusion curves seem to follow the typical S-shape using the independent cascade model, some networks show an almost linear relation to the activation probability. The linear threshold model shows similar picture. Here the general S-shape is very steep showing some kind of critical threshold or tipping point. Under a certain value the diffusion stays very low. Once this critical value is reached, the diffusion quickly reaches high values and a high network saturation.

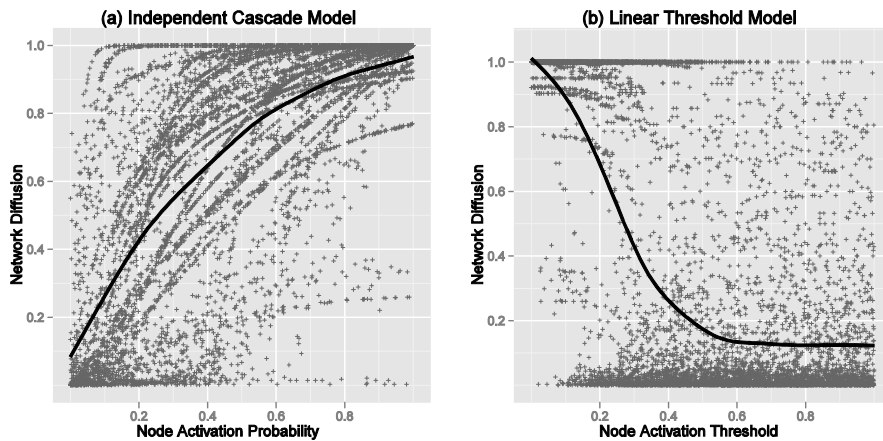


Fig. 1. Simulated diffusion of all networks depending on the activation probability of the network nodes. The figure includes a regression line (solid line).

To draw a more detailed picture, Figure 2 shows the diffusion curves of three exemplary networks, the cond-mat 2005 network, the power grid network and the Erdős collaboration graph network. The plot of the independent cascade model illustrates very clearly the effect of the underlying network. The Erdős graph shows an almost linear relationship whereas the cond-mat 2005 and the power grid network show dramatic differences. At a node activation probability value of 0.45 we can observe a very low mean diffusion of around 15% of the network nodes for the power grid network. On the other hand, the mean diffusion reaches values of 80% for the cond-mat 2005 network at the same value. Interestingly this does not account for the linear threshold model. Here the curve shapes are more similar. An interesting behavior shows the Erdős collaboration graph. As the only network, it reaches diffusion rates above 20% even at very high thresholds.

To evaluate the influence of the network parameters on the diffusions we calculated the correlation between those parameters and the mean diffusion. The results re-

vealed several significant correlations. Obviously the activation probability plays an important role ($r = 0.69$) as well as the node activation threshold ($r = -0.60$). The independent cascade model showed several significant correlations with the mean diffusion: *Average path length* ($r = -0.34$), *Network diameter* ($r = -0.38$), *Average node closeness* ($r = -0.39$), *Average node eigenvector* ($r = -0.38$), *Network density* ($r = -0.31$). Using the linear threshold model, the following significant correlations could be found: *Average node betweenness* ($r = -0.37$), *Network density* ($r = -0.34$). For both models we can state that a dense, highly connected network leads to higher diffusion rates. It is further important, that the network shows a small spatial extension. As mentioned in the introduction, we know that contemporary online social networks like Facebook show very small average path lengths and a small diameter, although those networks contain hundreds of millions of nodes. This directly relates to the dynamics of small world networks of Watts and Strogatz, showing that diseases (in our context information) spread more easily in networks characterized as highly clustered yet having small path lengths [14].

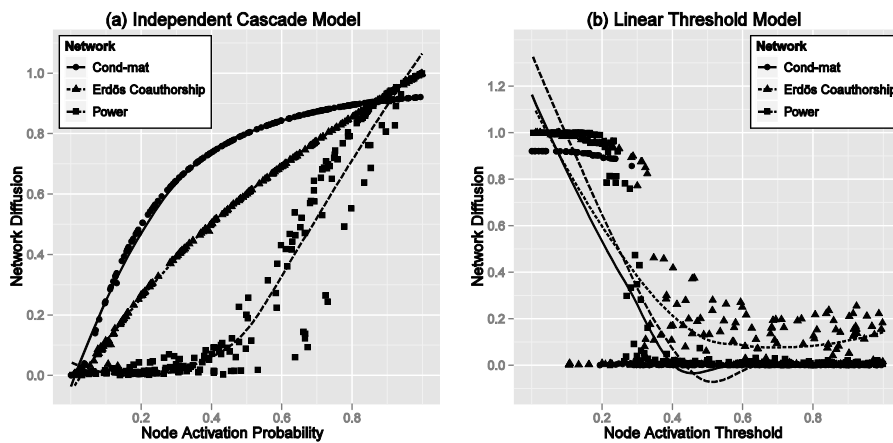


Fig. 2. Diffusion curves of three real world networks given the node activation probability (Independent Cascade Model) and the node activation threshold (Linear Threshold Model). The figure includes regression lines.

In many real world scenarios, marketers will not have a matching dataset to run simulations and create scenarios. Therefore artificially generated networks might be used, as we did in this paper. Since this is a convenient alternative, we were interested if generated networks show a significant effect on the diffusion. An ANOVA on the effect of the mean diffusion regarding an underlying real world or generated network was conducted. The ANOVA revealed a significant mean difference for both diffusion models: $F(1, 7198) = 84.56, p < .001, \omega = .01$ for independent cascade model, $F(1, 7198) = 14.48, p < .001, \omega = .002$ for linear threshold model. Although a significant difference exists, the effects are very low. Effect sizes like Kirk's ω or Cohens' d should always be calculated in order to determine the substantial effect. The gener-

ated networks tend to a slight overestimation of the diffusion. Marketers should bear this in mind when using generated networks.

4.2 Influence of Startnode Criterion - Independent Cascade Model

Figure 3 shows the estimated density from a kernel density estimator. The figure highlights a rather minor effect of this parameter on the diffusion. A conducted ANOVA showed a significant effect of the startnode criterion on the diffusion, $F(6, 7192) = 2.62, p < .01, \omega = .001$. Pagerank showed the highest overall diffusion mean of 0.7 ($\sigma = .34$) and betweenness ($\sigma = .36$) and eigenvector ($\sigma = .36$) the lowest mean with 0.65. However, the means hardly differ and the effects are very small. Moreover, as Figure 1, already showed, the standard deviation is very high. Only the difference between pagerank and betweenness as well as pagerank and eigenvector centrality showed small effects according to Cohen's d . Interestingly, the use of random startnodes vs. any other criterion showed very low effects to almost no effects. We can state that for the independent cascade model, the startnode criterion seems to have a minor importance having the pagerank criterion with the highest values.

4.3 Influence of Startnode Criterion - Linear Threshold Model

As seen in Figure 3, the selection of a startnode criterion has a greater effect on the linear threshold model diffusion. An ANOVA also proves the visual impression with a significant effect and higher effect values, $F(6, 7193) = 27.26, p < .001, \omega = .021$. Still the overall effect is small to medium according to Kirk's ω . *Betweenness* showed the highest diffusion mean with 0.42 ($\sigma = .44$) followed by *closeness*, *degree* and *PageRank* with 0.39 ($\sigma = .44, .43, .43$). *Clustering* and *random* showed the lowest diffusion means with 0.25 and 0.26 ($\sigma = .39, .40$). Looking at pairs of groups we can observe differences between *clustering* and *betweenness* (Cohen's $d = -.41$), *random* and *betweenness* (Cohen's $d = -.38$), *clustering* and *closeness* (Cohen's $d = -.33$), *random* and *closeness* (Cohen's $d = -.30$), *degree* and *clustering* (Cohen's $d = -.33$), *eigenvector* and *clustering* (Cohen's $d = -.26$), *pagerank* and *clustering* (Cohen's $d = -.35$), *random* and *degree* (Cohen's $d = -.30$) and *random* and *pagarank* (Cohen's $d = -.32$). We can thus state that start startnode criterion has a larger effect using the linear threshold model. Here the *betweenness* centrality showed the highest overall diffusion mean and random performed significantly lower. The overall effect is, however, rather small. Furthermore, the standard deviation in every group was very high showing a high degree of dispersion.

To get some deeper insights on the effect of the startnode criterion, we calculated an ANOVA and the effect sizes for every network dataset separately. The results show a very diverse picture. For some datasets the startnode criterion using the independent cascade model had no effect at all (dolphins network, $\omega = 0.003$), whereas other networks showed a high sensitivity to this parameter (hep-th, $\omega = 0.10$). Even more evident is the picture using the linear threshold model. Again some networks showed no effect at all and others showed a very high effect (Watts-Strogatz Network 7, $\omega = 0.46$). We can conclude that the effect of the startnodes strongly differs from

network to network. To answer the question what characteristics may be related to this behavior, we calculated correlations of all network characteristics with the omega squared ω effect size a network showed on the selection of the startnodes. The results revealed that the *Number of components* and the *Average clustering coefficient* are positively correlated with the effect ($r = 0.29$ and $r = 0.37$) and *Degree distribution power law fit* is negatively correlated with the effect ($r = -0.38$) for the independent cascade model. This leads to the conclusion that highly clustered and fragmented networks show a high sensitivity to the selection of the initial start nodes. Concerning the linear threshold model, the results differ. Here we can observe a significant correlation with the *Number of network communities* ($r = 0.50$). Again, the conclusion is similar. If a network is highly fragmented, the selection of initial start nodes is more important.

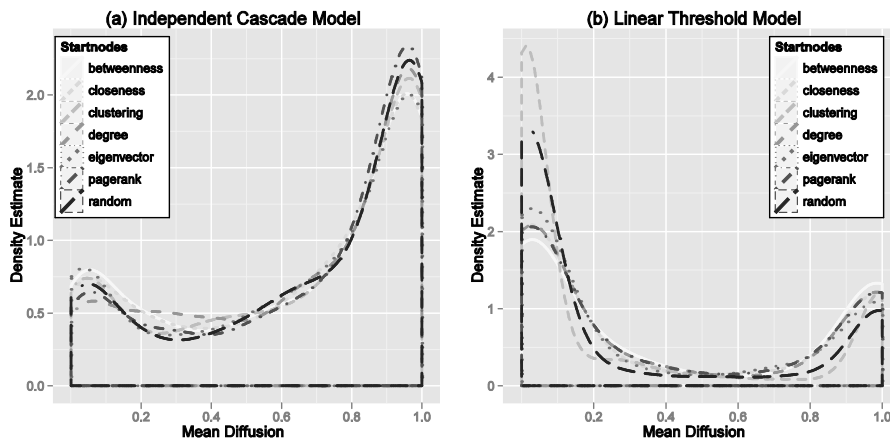


Fig. 3. Estimated densities of the mean diffusion grouped by the startnode criterion.

4.4 Influence of the Number of Initial Seeding Nodes

To prove whether the number of initial nodes has a significant influence on the resulting diffusion we calculated the overall correlation between the mean diffusion of a simulation run and the number of initial nodes used. We further calculated the same correlations for every network separately. Over all networks we can observe small but significant correlations between the mean diffusion and the number of initial start nodes, $r = 0.05$ for independent cascade model, $r = 0.15$ for the linear threshold model. Looking at the correlations for every network dataset separately, we can again state a very diverse picture. Using the independent cascade model, the correlations reach from no correlation at all to high significant correlations of $r = 0.40$ (Barabási-Albert 1). Similarly the linear threshold model shows even higher significant correlations up to $r = 0.59$ (Dolphins Network). To draw further conclusions we correlated the effect (r values) of the initial number of start nodes with the network characteristics of all networks. The results show that for the independent cascade model the *Average clustering coefficient* ($r = 0.48$), *Average node betweenness* ($r = 0.74$) and *Network densi-*

ty ($r = 0.68$) show significant correlations. The findings indicate that especially dense networks show a high effect regarding the number of initial start nodes. Marketers should therefore pay attention to the underlying density. Most real world networks, however, show a very low density. For those applications this parameter is of less importance.

5 Network Diffusion Prediction

To exploit the findings for a marketer's decision making we conducted a classification analysis. Goal of the analysis is to investigate if the diffusion may be predicted if network characteristics (Diameter, Number of Nodes, Density, etc.) and diffusion parameters (Number of start nodes, startnode criterion, Assumed activation probability, etc.) are given. This gives a marketer the possibility to predict a diffusion (the success of a campaign) based on assumptions about the general network characteristics. A marketer can therefore easily create worst-case, best-case and intermediate scenarios to manage possible campaigns. On this account a C4.5 decision tree was created and evaluated using a 10 fold cross-validation. Since the decision tree only classifies categorical variables, we created a 10 bin class variable from the Mean Diffusion variable in the original dataset. This leads to ten diffusion classes, 1 for very low and 10 for very high diffusion. The decision tree is supposed to predict the correct diffusion class based on network characteristics given. As seen in Table 1, the first decision tree achieved notable results with an overall accuracy of 86.012% for the independent cascade model. False predictions are scattered close to the real class, especially for low and high diffusions.

Table 2. Confusion matrix of decision tree classification (independent cascade model).

Confusion Matrix N=7200

Classified as										Real class
1	2	3	4	5	6	7	8	9	10	
832	67	14	2	3	1	2	4	0	0	1 n=925
68	200	42	7	2	0	0	0	0	0	2 n=319
10	44	176	37	12	2	0	2	0	0	3 n=283
3	7	34	163	47	6	2	2	1	0	4 n=265
3	1	3	39	206	25	13	4	2	1	5 n=297
0	0	1	2	39	264	41	6	5	1	6 n=359
0	0	1	2	10	38	341	57	9	0	7 n=458
0	0	1	1	1	2	46	478	57	0	8 n=586
0	0	0	1	0	2	8	45	755	62	9 n=873
0	0	0	0	0	0	0	5	52	2778	10 n=2835

The results of the linear threshold model show similar values and an overall prediction accuracy of 85.847%.

6 Conclusions

6.1 Practical Implications

The results of our analysis highlight the impact of the network structure on the diffusion process. This should be evaluated carefully when viral marketing is considered as a marketing tool. Dense networks and networks showing small average path lengths are very efficient for information diffusion whereas highly clustered networks are rather disadvantageous. If the network structure and the diffusion process are unknown, assumptions about the diffusion model need to be drawn since the diffusion varied considerably from the independent cascade to the linear threshold model. Bampo et al. showed that the network structure and the transmission behavior can be estimated during the first generations of a running campaign [5]. A campaign manager could use the estimated artificial network to forecast the campaign, considering the slight diffusion overestimation of such networks, as we have shown. We can further state that seeding matters, but strongly differs depending on the underlying diffusion model and the network structure. If the underlying network is fragmented and clustered, the seeding method is more important. This has already been indicated by Shakarian and Paulo who found that simulated diffusions in highly clustered networks required more initial seed nodes to reach a certain diffusion level [10]. The results also highlight that a campaign manager should rather concentrate on a strong growth than on massive seeding, since the number of initial seeds showed a minor effect.

6.2 Future Work

Our results also highlight that viral marketing is not a panacea to today's marketing. This directly confirms the findings of Leskovec et al. [44]. If the network structure is very disadvantageous, a campaign needs very high activation probability (or very low thresholds) to spread into the whole population. This might be unrealistic. Nevertheless, we can expect that virtual online networks will show even smaller average path lengths and a higher density in the future. Moreover, if currently still independent networks like Twitter and Facebook become more and more connected through aggregator services and mobile devices, the clustering is reduced, yet leading to a greater potential of electronic WOM, and therefore, viral marketing. Furthermore, the diffusion models used in this work might not be fully suited to describe spreading behavior. The behavioral characteristics of single nodes have not been taken into account. This should be addressed in the future to make the models more realistic. We further propose to conduct similar analyses on complex networks comprising different types of relations and multiple layers, as these networks might provide a more holistic reflection of the online communication.

References

1. Hui, K.L., Lee, S.-Y.T., Png, I.P.L.: Consumer Privacy and Marketing Avoidance: A Static Model. *Management Science* 54 (6), 1094–1103 (2008)
2. Solomon, M.R.: *Consumer Behavior*. Pearson Education (2006)
3. Porter, L., Golan, G.J.: From Subservient Chickens to Brawny Men: A Comparison of Viral Advertising to Television Advertising. *Journal of Interactive Advertising* 6 (2), 26–33 (2006)
4. Sattelberger, F.: *Erfolgsprognose bei Produktneueinführungen: Eine Untersuchung unter besonderer Berücksichtigung von Word-of-Mouth-Effekten*. LIT, Münster (2010)
5. Bampo, M., Ewing, M.T., Mather, D.R., Stewart, D., Wallace, M.: The Effects of the Social Structure of Digital Networks on Viral Marketing Performance. *Information Systems Research* 19 (3), 273–290 (2008)
6. Hinz, O., Skiera, B., Barrot, C., Becker, J.: Seeding Strategies for Viral Marketing: An Empirical Comparison. *Journal of Marketing* 75 (6), 55–71 (2011)
7. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146. ACM Press, New York, USA (2003)
8. Kimura, M., Saito, K., Nakano, R., Motoda, H.: Finding Influential Nodes in a Social Network from Information Diffusion Data. In: Liu, H. et al. (eds.): *Social Computing and Behavioral Modeling*. Springer US (2009)
9. Kimura, M., Saito, K., Nakano, R.: Extracting influential nodes for information diffusion on a social network. In: *AAAI'07 Proceedings of the 22nd national conference on Artificial intelligence - Volume 2*, pp. 1371–1376. AAAI Press (2007)
10. Shakarian, P., Paulo, D.: Large Social Networks can be Targeted for Viral Marketing with Small Seed Sets. In: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1–8. IEEE (2012)
11. Thompson, B.C.: Is the Tipping Point Toast?. *California Management Review* 43, 44–63 (2008)
12. GfK Marktforschung GmbH, Bereich Online Research: Alternative Werbeformen, Zentrale Ergebnisse, Welle 3, http://webguerillas.com/media/press/downloads/pdf/gfkstudie_2007.pdf
13. Leskovec, J., Adamic, L.A.: The Dynamics of Viral Marketing. *TWEB* 1, 1–46 (2008)
14. Watts, D.J., Strogatz, S.H.: Collective dynamics of “small-world” networks. *Nature* 393 (4), 440–442 (1998)
15. Barabási, A., Albert, R.: Emergence of Scaling in Random Networks. *Science* 286 (5439), 509–512 (1999)
16. Sun, T., Chen, W., Liu, Z., Wang, Y., Sun, X., Zhang, M., Lin, C.-Y.: Participation Maximization Based on Social Influence in Online Discussion Forums. In: *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media ICWSM'11* (2011)
17. Tillett, H.E.: Infectious Diseases of Humans: Dynamics and Control. *Epidemiology and Infection* 108 (1), p. 211 (1992)
18. Rogers, E.M.: *Diffusion of Innovations*. Free Press (1995)
19. Valente, T.W.: *Network Models of the Diffusion of Innovations*. Hampton Press (1995)
20. Bass, F.M.: A New Product Growth for Model Consumer Durables. *Management Science* 15 (5), 215–227 (1969)
21. Goldenberg, J., Libai, B., Muller, E.: Using complex systems analysis to advance marketing theory development. *Academy of Marketing Science Review* (9) (2001)
22. Mahajan, V.: *New-Product Diffusion Models (International Series in Quantitative Marketing)*. Springer (2000)

23. Junapudi, V., Udgata, G.K., Udgata, S.K.: Study of diffusion models in an academic social network. In: Janowski, T., Mohanty, H. (eds.): ICDCIT 2010. LNCS, Vol. 5966, pp. 267–278. Springer, Berlin Heidelberg (2010)
24. Kempe, D., Kleinberg, J., Tardos, É.: Influential Nodes in a Diffusion Model for Social Networks. In: ICALP 2005. LNCS, Vol. 3580, pp. 1127 – 1138. Springer (2005)
25. Granovetter, M.: Threshold models of collective behavior. *American journal of sociology* 83 (6), 1420–1443 (1978)
26. Goldenberg, J., Libai, B., Muller, E.: Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters* 12 (3), 211–223 (2001)
27. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*. Cambridge University Press (1994)
28. Newman, M.: *Networks: An Introduction*. Oxford University Press (2010)
29. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30, 107–117 (1998)
30. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. *Physical Review E* 74 (3), 22 (2006)
31. Newman, M.E.J.: Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics* 46, 323–351 (2005)
32. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.: The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology* 56, 396–405 (2003)
33. Knuth, D.E.: *The Stanford GraphBase: A Platform for Combinatorial Computing*. Addison Wesley, Reading (1993)
34. Matei, R., Iamnitchi, A., Foster, P.: Mapping the Gnutella network. *IEEE Internet Computing* 6 (1), 50–57 (2002)
35. Opsahl, T., Panzarasa, P.: Clustering in weighted networks. *Social Networks* 31, 155–163 (2009)
36. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. *Physical Review E* 74 (3), 22 (2006)
37. Newman, M.E.: The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America* 98, 404–9 (2001)
38. Guimer, R., Danon, L., Diaz-Guilera, A., Giralt, F., Arenas, A.: Self-similar community structure in a network of human interactions. *Physical Review E* 68, (2003)
39. Arenas, A., Danon, L., Díaz-Guilera, A., Gleiser, P., Guimerá, R.: Community analysis in social networks. *The European Physical Journal B - Condensed Matter* 38, 8 (2004)
40. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., Arenas, A.: Models of social networks based on social distance attachment. *Physical Review E* 70 (2004)
41. Erdős, P., Rényi, A.: On random graphs. *Publicationes Mathematicae (Debrecen)*. 6, 290–297 (1959)
42. Csardi, G., Nepusz, T.: The igraph software package for complex network research. *InterJournal. Complex Sy*, 1695 (2006)
43. R Development Core Team: *R: A Language and Environment for Statistical Computing*, <http://www.r-project.org> (2010)
44. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. *ACM Transactions on the Web* 1, 5–es (2007)
45. Katz, E., Lazarsfeld, P.: *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Transaction Publishers (2005)

A Comprehensive Analysis of E-Government Adoption in the German Household

Cigdem Akkaya, Petra Wolf, and Helmut Krcmar

Technische Universität München, Garching by Munich, Germany
{cigdem.akkaya,petra.wolf,krcmar}@in.tum.de

Abstract. Much of the prior research on IS adoption recognizes that cultural characteristics of the nations influence their adoption behaviors significantly. In the context of e-government, more empirical research is necessary to understand the adoption behaviors of different nations. Our research focuses on understanding the antecedents of e-government adoption in the German household, which has not been adequately addressed to date. Based on the findings of two representative cross-sectional studies, we derived a comprehensive research model and tested it with 1,000 users in the German household on the specific example of e-filing. While the factors of data protection and security were mentioned as crucial in the descriptive studies, the explanatory analysis with LISREL revealed that compatibility is the main antecedent of e-filing adoption in the German nation, followed by relative advantage and perceived risk. Implications for practice and future research are discussed.

Keywords: IS Adoption, E-Government, Germany, LISREL, Structural Equation Modeling

1 Introduction

Despite the enormous potential of new public management, most e-government initiatives are faced with adoption challenges. Until now, development of e-government services has been primarily guided by supply side factors and technological possibilities rather than user needs [1]. There is an increasing body of research claiming that a user-centered e-government strategy is essential, if e-government is to succeed [2]. Therefore, empirical research becomes critical in understanding the expectations of citizens and their decision making mechanisms towards using online public services.

Prior literature suggests that national culture shapes perceptions of citizens thus facilitates or impedes adoption of new technologies [3]. According to Patel and Jacobson [4], factors influencing e-government adoption can be influenced by cultural backgrounds of the users, however there is a lack of empirical studies examining the impact of national culture on e-government adoption. Some technology related factors may be salient in all cultures. Yet, risk perceptions and privacy concerns, which play key roles in the adoption of other online technologies, vary considerably from one nation to another [5]. Consequently, more empirical research is necessary to under-

stand the nations' expectations to achieve worldwide adoption of e-government services [1], [4].

The Cap Gemini Report on e-government [6] recognizes Germany as one of the top performers in full online availability and sophistication of online public services. In the Global Competitiveness Report [7] Germany is ranked second for the quality of its infrastructure worldwide. Moreover, the Federal Government provides full support and dedication to make Germany one of the leaders of e-government in Europe. The widespread adoption of e-government services has been part of the national strategy over a decade, with the specific target of "enabling Germany to become one of the e-government leaders of Europe" [8]. However, the adoption problems of e-government in Germany seem to persist, especially in the household context [9]. Other than some municipal level adoption research [10], the major antecedents influencing adoption in Germany have not been subject to empirical research with a representative study. Hence, the factors motivating or hindering households' use of online public services as well as their expectations, demands and concerns for a better e-government adoption are yet to be understood.

The presented research aims to contribute to a better understanding of e-government adoption in Germany by addressing this research gap in literature. In particular, the main research question guiding our research is as follows: What are the antecedents of e-government adoption in the German household?

The document is organized as follows. The next section provides a brief review of the literature and discusses the high sensitivity of the German citizens towards data protection and security. Section three summarizes the insights gained by the descriptive cross-sectional studies. The research model and hypotheses derived are elaborated in the following section. Section five summarizes the methodology used and the results are analyzed in section six. After discussing the main findings, implications and limitations, the paper concludes with a high-level summary of the key findings arising from our research.

2 Literature Review

2.1 Technology Adoption in E-government Research

It is common practice for e-government researchers to draw on constructs, theories and models from IS literature to study e-government adoption [11]. Up to now, various factors have been discussed as being the prominent factors of IS adoption. Some constructs consistently explain a large percentage of variance such as perceived usefulness (relative advantage) [12] and ease of use (simple usability) [12]. Yet, previous research suggests that some determinants related to beliefs and values vary by culture [13], necessitating empirical research conducted in each culture individually [14].

We argue that perceived risk and trust are among the essential constructs that need to be analyzed to understand e-government adoption. Broad literature on e-commerce adoption recommend building trust of consumers [15] for decreasing their risk perceptions to foster adoption [5]. In fact, the risk involved in e-government is actually

much higher than e-commerce [16]. In e-commerce, consumers risk the theft of credit card information and access to personal preferences by third parties. However, in e-government, citizens transmit their income tax records, digital identities and even permanent characteristics such as biometrics, which would have severe life-long consequences in case of an unwanted third party access. Hence, the key factors of perceived risk and trust become crucial for the context of e-government adoption.

It is necessary to distinguish between the trust of the transmission medium and trust of the service provider, as in the case of e-commerce [17]. Consumers' decision to adopt e-commerce is affected not only by the perceptions of technology but also by the beliefs about the credibility of the e-vendor. Likewise, the perceptions of users on technological security of the transmission medium are significantly different from the perceptions on public organizations' ability to deliver their commitments or the ethical use of the collected data. The latter requires trust in government, which may be difficult to establish. In particular, citizens tend to suspect that government watches everything and gathers data about citizens through various channels [18]. Worldwide data scandals and identity thefts intensify concerns of users towards online public services even more.

It is widely acknowledged that national culture influences how people react to perceived risk and handle trust [19]. Previous studies have demonstrated how espoused cultural values affect a nation's behavior by altering their belief structures [3], [14], also in the context of e-government [20-21].

2.2 High Sensitivity of the German Nation towards Data Protection and Security

The German nation is highly sensitive towards the issues of data protection and security. The high risk-aversion of the German nation is a widely recognized phenomenon [22] as confirmed by the well-known cultural classifications [19], [23]. The past experience of surveillance state in the German history may be seen as an influence factor on the high risk-aversion at the national level, which deserves further research.

An overview of data protection laws in Germany reflects that Germany has one of the strictest data protection laws in the European Union. In 1983, the German Federal Constitutional Court acknowledged an individual's "right of informational self-determination" and ascertained that the privacy of personal data is a fundamental right in German constitutional law [24]. Since then, the basic criterion for handling personal data – including analysis, processing and further use – by any third party has been a given right to the individual who owns the data. Processing of personal data is only permitted with the consent of the individual, independent of its context and sensitivity. Indeed, various nationwide initiatives involving transfer of sensitive personal data have been protested nationwide by citizens, non-governmental organizations and political parties and even abandoned due to infringement to personal privacy [25].

The term "transparent citizen" (in German "gläserner Bürger") – which was originally used to define an anatomical human model made of plastic – has become the metaphor for violating the privacy of citizens in Germany [26]. It implies a complete fluoroscopy (X-Ray screening) of a citizen in terms of his complete personal data and

tracking of his activities by the government. We argue that the data privacy and security concerns of the citizens need to be analyzed carefully, if e-government in Germany is to succeed. Especially, large-scale projects involving transfer of sensitive personal data are not likely to be successful without a thorough understanding of various risk perceptions of citizens.

3 Insights Gained from Descriptive Research

We have followed a two-step methodology, which combines descriptive and explanatory research. Firstly, we conducted two descriptive studies in 2010 and 2011 – with two different nationwide representative samples – to receive insights into the factors that influence e-government adoption in Germany, which were then analyzed in 2011 as part of an explanatory study. Two cross-sectional studies over these two years were conducted to increase the validity of findings, as one time cross-sectional design would only provide a ‘snapshot’ at a given point of time [27].

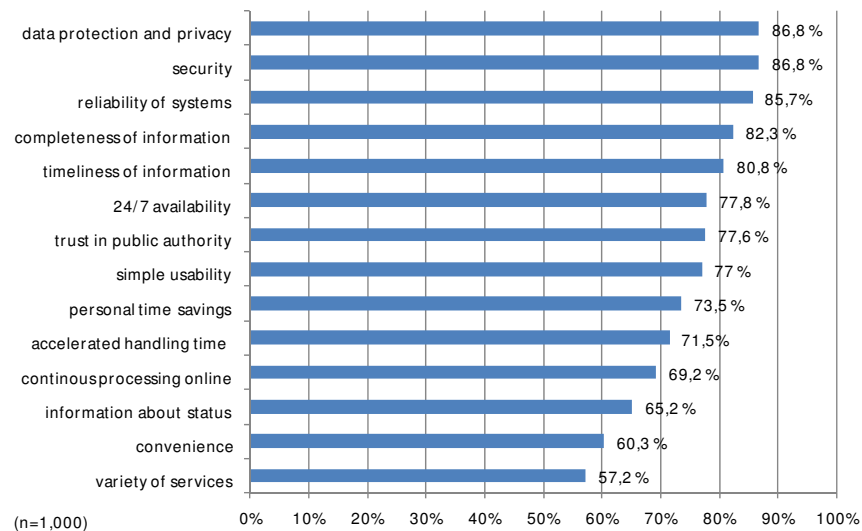


Fig. 1. Factors influencing the adoption of e-government in Germany (2011) (based on [28])

The descriptive study included question on the familiarity and satisfaction of e-government services in Germany. For the purposes of this paper, we only consider the question that aims to provide better understanding on factors influencing e-government adoption. Figure 1 summarizes the results in 2011 and Table 1 analyzes the change of the factors over the study in 2010 regarding this specific question.

As depicted in Figure 1 above, the factors of data protection, privacy and security are the most important considerations for one’s e-government engagement (86,8%) followed by the reliability of systems (85,7%). These factors highlight the importance of citizens’ trust in government and technology, including the supporting infrastructure and the transmitting medium. Whilst the trust in government is mentioned implic-

itly under ‘data protection and privacy’, the importance of trust in government became more apparent under ‘trust in public authority’ (77,6%).

Table 1. The change in factors influencing adoption over the two years (based on [28-29])

		N	Mean	Std. Dev.	Std. Err. Mean	t	df	sig. (p)	diff.
data protection and privacy	2010	1002	4,583	0,667	0,021	1,440	1865	.150	No
	2011	1000	4,533	0,876	0,028				
security	2010	1002	4,534	0,673	0,021	0,824	1886	.410	No
	2011	1000	4,505	0,863	0,027				
reliability of systems	2010	1002	4,402	0,697	0,022	-0,060	1895	.953	No
	2011	1000	4,404	0,884	0,028				
completeness of information	2010	1002	4,003	0,707	0,022	-7,388	1886	.000	Yes
	2011	1000	4,271	0,906	0,029				
timeliness of information	2010	1002	4,202	0,696	0,022	0,094	1868	.925	No
	2011	1000	4,199	0,911	0,029				
24/7 availability	2010	1002	4,242	0,738	0,023	2,236	1868	.025	Yes
	2011	1000	4,156	0,967	0,031				
trust in public authority	2010	1002	4,172	0,783	0,025	-0,481	1914	.631	No
	2011	1000	4,191	0,968	0,031				
simple usability	2010	1002	4,160	0,725	0,023	0,440	1883	.660	No
	2011	1000	4,144	0,933	0,029				
personal time savings	2010	1002	4,084	0,794	0,025	1,544	1941	.123	No
	2011	1000	4,024	0,945	0,030				
accelerated handling time	2010	1002	4,057	0,785	0,025	1,788	1915	.074	No
	2011	1000	3,986	0,970	0,031				
continuous processing online	2010	1002	3,924	0,816	0,026	-0,046	1915	.963	No
	2011	1000	3,926	1,008	0,032				
information about status	2010	1002	3,782	0,838	0,026	-1,698	1949	.090	No
	2011	1000	3,851	0,983	0,031				
convenience	2010	1002	3,684	0,912	0,029	-1,776	1969	.076	No
	2011	1000	3,761	1,033	0,033				
variety of services	2010	1002	3,694	0,798	0,025	-0,195	1911	.846	No
	2011	1000	3,702	0,990	0,031				

The independent-samples-t-test analysis presented in Table 1 suggests that, the importance of most factors has not changed significantly in 2011. In the overall, we conclude that the results of the two years did not differ considerably. Afterwards, we proceeded to defining the appropriate research model and developing the hypotheses for our explanatory study, which is explained next.

4 Research Model and Hypotheses

The descriptive studies reveal that perceived risk, trust in technology and trust in government authorities play an important role in German society. Moreover, they illustrate that citizens value relative advantage – correspond to 24/7 availability, accelerated handling time and personal time savings in these studies – and simple usability of

the services, which were regarded as being important by at least 71,5% of the respondents. By combining these results with the prior IS adoption literature, we derived our research model. Gefen et al. [30] suggest that confirmatory research should be conducted by using research models with strong theoretical basis. Hence, we combined the selected constructs from the Roger's [31] 'Theory of Perceived Attributes' and the 'Trust and Risk Model' of Belanger and Carter [32].

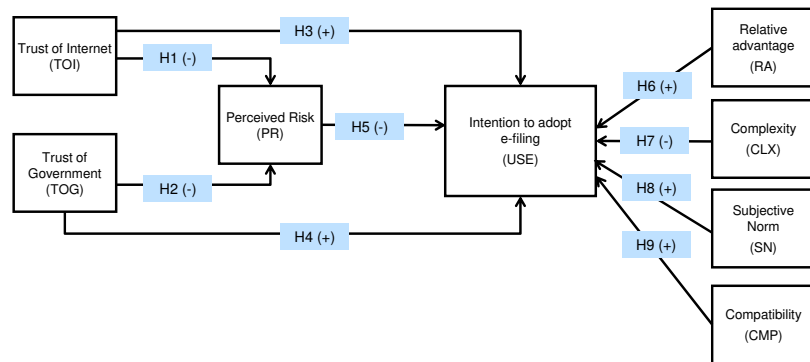


Fig. 2. Research Model

Roger's theory states that potential adopters judge an innovation based on their perceptions of five attributes: relative advantage, complexity, compatibility, observability and trialability [31]. Based on a comprehensive review and meta-analysis, Tornatzky and Klein [33] found that relative advantage, complexity and compatibility are the most significant factors in explaining adoption of innovations, so these were included in our research model. In order to reflect the issues of trust and perceived risk, we included trust of the Internet, trust of the government and perceived risk from the Belanger and Carter's model of e-government adoption. Finally, we added the construct of subjective norm. There is a considerable amount of literature indicating that social influences play an important role in determining the acceptance and usage behavior of new adopters of information technologies [34]. Figure 2 presents our proposed research model and hypotheses.

The intention to use e-government services was questioned by the specific example of e-filing for a number of reasons. First, due to the federal structure of the government in Germany, there may be differences among the services offered by different states and municipalities. Yet, individual income tax filing is a nationwide service of the German Tax Administration, without any specific differences in states or municipalities. Second, tax filing is a civic obligation, which should be familiar to a large amount of the population. Third, the acceptance rate of e-filing in Germany is still relatively low compared to other countries. Thus, questioning particularly based on this example would also deliver the reasons hindering its nationwide acceptance. Finally, some of the existing studies on e-government adoption [35-37] use the specific example of e-filing as well, which would enable cross-cultural comparison of different nations.

5 Methodology

5.1 Data Collection

The sample for the explanatory study was randomly selected to be representative of the German population to ensure high validity. The online survey was conducted between 8th and 31th of August 2011. The final sample included 1,000 Internet drawn from the target population of all German households over 18 years old, who had household Internet access, for an error margin of 3,1% at a 95% confidence level. The data is weighted to be representative for the total online population by central features like gender, age and formal education. The sample's age ranged from 18-75, while 81,1% respondents were in the 18-54 age group. 51,8% of the participants were female, 83,5% of them used the Internet several times a day and 84,6% had already been using the Internet for more than five years.

5.2 Instrument Development

Survey instrument was developed by using existing pre-validated measures of similar constructs in literature [17], [34], [38-40] that were modified to fit to the research context. The instrument consisted of thirty-two indicators on a structured seven-point Likert scale ranging from "strongly agree" to "strongly disagree". The initially developed instrument was initially pilot tested, which helped to reveal problems with question wording, lack of clarity and ambiguity. Such questions were improved before administrating the scale to the intended sample. Face and content validity of the instrument were evaluated by two professors who are experts in the area of trust and IS research, two IS experts with PhD degrees in IS and one marketing research expert.

The internal consistency of the scale, the Cronbach's alpha, has ranged between 0.74 and 0.93. Four items having lower reliability were removed which were the reverse-scored ones on the survey instrument, suggesting that the direction of the wording may have caused the problem. The reliabilities of all measures surpassed the 0.70 cut-off value [41] thus all constructs had acceptable reliability.

Podsakoff et al. [42] suggest that studies using single source, self-reported data should be checked for common method bias. Harman's single-factor test [43] was applied by conducting an exploratory factor analysis to examine the presence of the bias. The principal components factor analysis resulted in seven factors with no factor accounting for the majority of the variance, while the first factor accounted for 34,75% of the variance. This suggests that method bias did not overly influence the responses in this study.

6 Data Analysis and Results

Due to the existence of various latent variables, the second generation multivariate analysis technique Structural Equation Modeling (SEM) [30] was selected for data analysis. We carried out the analysis in accordance with the two-stage methodology

suggested by Anderson and Gerbing [44]. The first step was to establish convergent and discriminant validity of the proposed constructs.

6.1 Measurement Model

Convergent and discriminant validity of the constructs were examined with several tests. The confirmatory factor analysis (CFA) conducted by using LISREL 8.80 [45] demonstrated a good model fit (SRMR = 0.039, CFI = 0.98, RMSEA = 0.061, NFI = 0.97) [41], [45], supporting both convergent and discriminant validity.

Table 2. Convergent and discriminant validity

	CR	AVE	MSV	ASV
TOI	0,774	0,547	0,269	0,074
TOG	0,882	0,714	0,269	0,090
PR	0,897	0,688	0,339	0,165
RA	0,866	0,622	0,615	0,284
CLX	0,934	0,779	0,618	0,262
USE	0,830	0,644	0,601	0,274
SN	0,914	0,729	0,038	0,018
CMP	0,917	0,788	0,618	0,324

In addition to CFA, we also checked the construct validity to ensure convergent and discriminant validity of the variables before including them into the hypotheses testing. As illustrated in Table 2, all variables in the study have provided a sufficient convergent validity. The standardized factor loadings were highly significant. Composite reliabilities of all variables exceeded the minimum limit of 0.70 and were larger than the average variance extracted (AVE). The average variance extracted estimates were all above the recommended 0.50 level [41], [46], which implied that more than one-half of the variances observed. With regard to discriminant validity, we compared the maximum shared squared variances (MSV) between factors and average shared squared variance (ASV) with the average variance extracted. All constructs surpassed this test thus the discriminant validity was established [45].

As another test for discriminant validity, the square root of the average variance extracted for each construct was compared against its correlations with other constructs [46], as shown in Table 3. This discriminant validity assessment has also revealed that all constructs displayed sufficient discriminant validity.

Table 3. Construct correlation matrix

	SN	TOI	TOG	PR	RA	CLX	CMP	USE
SN	0,854							
TOI	0,161	0,740						
TOG	0,178	0,519	0,845					
PR	0,054	-0,187	-0,208	0,829				
RA	0,194	0,252	0,284	-0,483	0,789			
CLX	0,037	0,213	0,272	-0,461	0,689	0,883		
CMP	0,140	0,220	0,272	-0,541	0,784	0,786	0,887	
USE	0,074	0,169	0,234	-0,582	0,695	0,638	0,775	0,803

To sum up, the conducted CFA and convergent validity tests showed that we did not have any issues of convergent and discriminant validity. Next, the structural model and the hypotheses were examined.

6.2 Structural Model

The structural equation model was estimated using the maximum likelihood method of LISREL 8.80 [45] to examine the relationships among the proposed constructs. The model fit measure of ‘standardized root means square residual’ (SRMR) had a value of 0.086, exceeding both the good fit level of 0.050 and the acceptable fit level of 0.080 [47]. In order to improve the model fit, the modification indices of LISREL suggested the addition of a path from compatibility to perceived risk. This modification was made in accordance with the previous empirical studies which found that compatibility is negatively related to perceived risk [48]. With this modification, the model provided a very good fit (SRMR = 0.041, CFI = 0.98, RMSEA = 0.059, NFI = 0.97) [41], [45], which is considered to be ‘excellent’ according to [47] with the fit indices of SRMR < 0.08, CFI > 0.95 and the RMSEA < 0.06.

Figure 3 presents the structural relationships among the constructs and standardized path coefficients. Perceived risk is predicted by trust of the Internet ($\beta = -0.14$), trust of the government ($\beta = 0.11$) and compatibility ($\beta = -0.46$). These variables totally explain a variance of 23% on perceived risk ($R^2 = 0.23$). Intention to use is jointly predicted by perceived risk ($\beta = -0.24$), compatibility ($\beta = 0.41$), relative advantage ($\beta = 0.26$) and subjective norm ($\beta = 0.06$) by explaining a total variance of 63% ($R^2 = 0.63$).

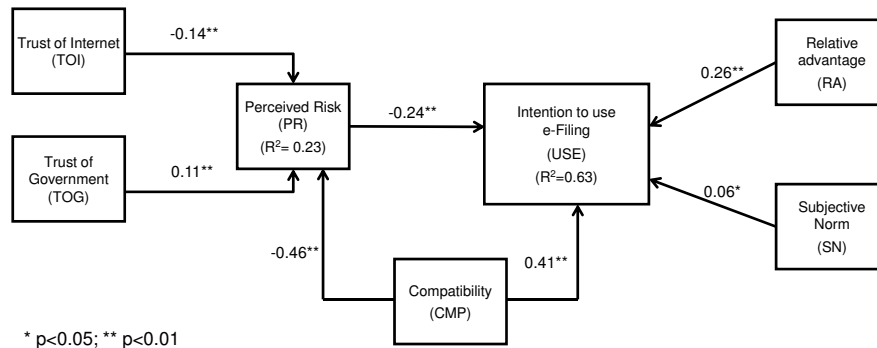


Fig. 3. Standardized path coefficients for significant relationships

Table 4 summarizes the hypotheses tests. Overall, five of the nine proposed hypotheses are accepted. Higher levels of trust of the Internet decrease perceived risk (H1). As the degree of trust on the Internet as the transmitting medium increases, the degree of perceived risk decreases. Similarly, higher levels of perceived risk affect intention to use e-filing negatively (H5). If users perceive high amount of risk to make their tax declarations online due to any reason, they will be less willing to use e-filing.

Table 4. Path coefficients and hypothesis testing

Hypotheses	Relationship	Estimate	Std-Error	t-value	Supported
H1	TOI->PR	-0.14	0.043	-3.25	YES
H2	TOG->PR	0.11	0.040	2.70	NO
H3	TOI->USE	-0.01	0.033	-0.20	NO
H4	TOG->USE	0.03	0.031	1.01	NO
H5	PR->USE	-0.25	0.027	-9.47	YES
H6	RA->USE	0.29	0.039	7.49	YES
H7	CLX->USE	-0.03	0.044	-0.76	NO
H8	SN->USE	0.06	0.025	2.52	YES
H9	CMP->USE	0.39	0.046	8.61	YES

Relative advantage behaves also as expected (H6) by positively influencing intention to use. If an online service provides more advantages than its paper-based version, citizens would prefer to use the former. Subjective norm affects intention to use positively as well (H8), however its effect is barely significant. This finding suggests that, there is only a low amount of peer influence on citizens for determining the behavior of e-filing. Perhaps, this is caused by the sample of experienced Internet users. Yet, the experiences of the peers could have been influential on the people having very limited amount of online experience. We suggest future research to explore this aspect with other samples to understand this relationship.

Compatibility influences intention to use e-filing significantly, therefore H9 is also supported. Overall, we observe that intention to use e-filing is largely driven by compatibility. Besides this direct positive effect, it has also an indirect effect on intention

to use by decreasing perceived risk. Thus it seems that taxpayers are influenced by the compatibility of the e-filing method with their life and work styles.

Notably, H2 is not supported. We expected trust of the government would negatively affect perceived risk however in contrast; it has a slight positive effect. This contradicts with the literature that perceived risk decreases when trust is present [32], [49]. Future research should explore this relationship with new samples. H3 and H4 are also rejected. They suggest that trust of the Internet and trust of the government increase the intention to use e-filing, however such a direct effect cannot be observed. The impact of trust of the Internet is mediated by perceived risk, consistent with findings obtained by previous studies [50]. Future research may focus on understanding the mediating effect of perceived risk on trust rather than analyzing its direct effect.

Surprisingly, H7 is rejected. Complexity does not have a significant effect on intention to use. This contradicts with the highly recognized models of IS adoption research [40], [51]. Indeed, e-filing in Germany is a relatively complex e-government service compared to other services. Thus the complexity of e-filing could still be an important antecedent to adoption for the users having limited amount of experience with the Internet and online technologies. Future studies should explore the impact of complexity using such a sample for verification.

7 Discussion

Overall, the descriptive studies of the presented research reveal important insights about the expectations, concerns and demands of the citizens towards adopting e-government services. The findings of the confirmatory research suggest that compatibility, relative advantage and perceived risk are the main antecedents influencing adoption in the German household.

Our research aims both theoretical and practical contributions. Despite its development in recent years, research on adoption of e-government is still in its infancy. Scarce research on adoption of e-filing has mainly utilized the technology acceptance model (TAM) by Davis [40], theory of planned behavior (TPB) by Ajzen [52] and the Unified Theory of Acceptance and Use of Technology (UTAUT) by Venkatesh et al. [51]. Except some initial attempts [53], Roger's Theory of Perceived Attributes [31] have not received much attention to analyze adoption of e-filing, which was applied in this research. This research also provides a contribution to the cultural context of e-government adoption. While adoption in different nations receives increasing interest, antecedents of e-government adoption in Germany remained relatively under-researched. Our research should be considered as a building block for future research on understanding the adoption behavior of the citizens in Germany.

Moreover, the empirical findings of this paper provide valuable insights to the policy makers to promote the adoption of e-government nationwide. The cross-sectional studies over the two years have drawn attention to the data protection and security related concerns as important inhibitors of e-government adoption. Thus, citizens should be aware of and have control over the personal information to be stored [54]. Privacy policies involving access control should be documented and communicated.

In addition, government authorities should pay more attention to delivering services that are compatible with the life and working styles of the citizens and concentrate on increasing the relative advantages of the online services over the traditional ones.

Some limitations of the study should be acknowledged which point to directions for future research. First, in order to reach out to nationwide representative samples, we used online questionnaires with households having a PC and an Internet access; therefore the sample had a considerable amount of previous Internet experience. This may have biased our results, especially in terms of the constructs complexity, subjective norm and trust of the Internet. We suggest future research analyze the influences of these constructs on samples having no household PCs and only very limited experience with the Internet. Second, due to the emphases of data protection and security in the cross-sectional descriptive studies, the constructs of trust of the Internet and trust of the government were integrated in the research model. Although our selection was based on prior literature, trust constructs may not have fully reflected the data protection and security concerns of the users. Instead, we suggest future research to test the direct impacts of data protection and security concerns on e-government adoption, rather than the indirect impact of trust, which may seem irrelevant for the respondents. Third, it should be noted that the purpose of this research was to understand the adoption behavior of the German citizens rather than investigating the cultural differences among nations. For such a purpose, a cross-cultural study would be necessary using the same research model and instrument to understand adoption in different nations. Although perceived ease of use [55], social influence [56] and trust of the government [57] found to be among the crucial antecedents of the intention to use e-filing in other cultures, this may or may not be caused by the cultural differences among the nations, which needs to be confirmed further in cross-cultural studies analyzing the moderating effect of national culture on adoption. Finally, the descriptive studies were not based on a specific e-government service but the explanatory research was conducted on the specific example of e-filing. We suggest further research to validate the findings of this research by using other examples of e-government services in Germany.

8 Conclusion

In contrast to IS adoption literature in other online contexts, research in e-government adoption is still in its infancy. Despite its development in the last decade, more empirical research is still necessary, to understand and boost e-government adoption worldwide. Research in similar contexts suggests that adoption behaviors of nations are influenced by the espoused cultural values, beliefs and norms. Even though e-government adoption in some cultures has been subject to research, some others remain relatively under-researched.

We analyzed the determinants of e-filing adoption in Germany with a comprehensive empirical research combining two cross-sectional descriptive studies and an explanatory study, being all nationwide representative. The descriptive studies demonstrated the sensitivity of the nation towards the issues of data protection and security

therefore we included the constructs of trust and perceived risk in our research model. The structural equation modeling analysis conducted in the second stage revealed that compatibility, relative advantage and perceived risk are the main determinants of adoption in the German household. In light of these findings, we suggest future researchers to analyze decision making of households with other research models and samples, in order to contribute to the existing knowledge on e-government adoption.

References

1. Verdegem, P., Verleye, G.: User-centered E-Government in practice: A comprehensive model for measuring user satisfaction. *Government Information Quarterly* 26, 487-497 (2009)
2. Schedler, K., Summermatter, L.: Customer orientation in electronic government: Motives and effects. *Government Information Quarterly* 24, 291-311 (2007)
3. Srite, M., Karahanna, E.: The role of espoused national cultural values in technology acceptance. *MIS Quarterly* 30, 679-704 (2006)
4. Patel, H., Jacobson, D.: Factors influencing citizen adoption of E-government: a review and critical assessment. In: *ECIS 2008 Proceedings*, pp. 1058-1069 (2008)
5. Luo, X.: Trust production and privacy concerns on the Internet: A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management* 31, 111-118 (2002)
6. Capgemini: *Digitizing public services in Europe: putting ambition into action* (2011)
7. World Economic Forum: *The Global Competitiveness Report 2010-2011*. (2011)
8. Bundesministerium des Innern: *Umsetzungsplan 2009 - E-Government 2.0* (2009)
9. United Nations: *E-Government Survey* (2012)
10. Heierhoff, L., Hofmann, S.: Adoption of Municipal e-Government Services - A Communication Problem? In: *AMCIS 2012 Proceedings, Paper 7* (2012)
11. Gefen, D., Warkentin, M., Pavlou, P.A., Rose, G.M.: EGovernment Adoption. In: *Americas Conference on Information Systems, Paper 83, Dallas, Texas* (2002)
12. Davis, F.D.: User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int. Journal of Man-Machine Studies* 38, 475-487 (1993)
13. Doney, P.M., Cannon, J.P., Mullen, M.R.: Understanding the influence of national culture on the development of trust. *Academy of Management Review* 23, 601-620 (1998)
14. Gefen, D., Heart, T.: On the Need to Include National Culture as a Central Issue in E-Commerce Trust Beliefs. *Journal of Global Information Management* 14, 1-30 (2006)
15. Slyke, C.V., Belanger, F., Comunale, C.L.: Factors influencing the adoption of web-based shopping: the impact of trust. *ACM SIGMIS Database* 35, 32-49 (2004)
16. Gefen, D.: Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM SIGMIS Database* 33, 38-53 (2002)
17. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-Commerce: an integrative typology. *Information Systems Research* 13, 334-359 (2002)
18. Belanger, F., Hiller, J.: A framework for e-government: privacy implications. *Business Process Management Journal* 12, 48-60 (2006)
19. Hofstede, G.: *Cultures and organizations: software of the mind*. McGraw-Hill, New York (1991)

20. Carter, L., Weerakkody, V.: E-government adoption: A cultural comparison. *Information Systems Frontiers* 10, 473-482 (2008)
21. Warkentin, M., Gefen, D., Pavlou, P.A., Rose, G.M.: Encouraging citizen adoption of e-Government by building trust. *Electronic Markets* 12, 157-162 (2002)
22. The Lauder Institute: First-hand perspectives on the global economy (2009)
23. House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W., Gupta, V. (eds.): *Culture, Leadership and Organizations: The GLOBE Study of 62 Societies*. Sage, Thousand Oaks (2004)
24. Bundesrepublik Deutschland: Volkszählungsurteil (1983)
25. Klopp, T.: Zentralregister für Streikende. *Zeit Online* (2009)
26. Akkaya, C., Wolf, P., Krcmar, H.: Factors Influencing Citizen Adoption of E-Government Services: A Cross-Cultural Comparison (Research in Progress). In: 45th Hawaii International Conference on System Science (HICSS), pp. 2531-2540 (2012)
27. Punch, K.F.: *Introduction to social research*. Sage, London (2005)
28. Krcmar, H., Wolf, P., Wolf, M., Dirtheuer, K.: *eGovernment Monitor In: Initiative D21 e.V., Institute for Public Information Management (ipima) (eds.)* (2011)
29. TNS Infratest: (N)ONLINER Atlas 2010 eGovernment Monitor (2010)
30. Gefen, D., Rigdon, E.E., Straub, D.: An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly* 35, iii-A7 (2011)
31. Rogers, E.M.: *Diffusion of Innovations*. The Free Press, New York (1995)
32. Belanger, F., Carter, L.: Trust and risk in e-government adoption. *Journal of Strategic Information Systems* 17, 165-176 (2008)
33. Tornatzky, L.G., Klein, K.J.: Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. *IEEE Transactions on Engineering Management* 29, 28-45 (1982)
34. Taylor, S., Todd, P.A.: Understanding information technology usage: a test of competing models. *Information Systems Research* 6, 144-176 (1995)
35. Wu, I.-L., Chen, J.-L.: An extension of Trust and TAM model with TPB in the initial adoption of on-line tax: an empirical study. *Int. J. of Human-Computer St.* 62, 784-808 (2005)
36. Schaupp, L.C., Carter, L., Hobbs, J.: E-File Adoption: A Study of U.S. Taxpayers' Intentions. In: *Proceedings of the 42nd Hawaii International Conference on System Sciences*, pp. 1-10 (2009)
37. Fu, J.-R., Farn, C.-K., Chao, W.-P.: Acceptance of electronic tax filing: a study of taxpayer intentions. *Information & Management* 43, 109-126 (2006)
38. Teo, T., Srivastava, S., Jiang, L.: Trust and electronic government success: an empirical study. *Journal of Management Information Systems* 25, 99-132 (2008)
39. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research* 2, 192-222 (1991)
40. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of Information Technology. *MIS Quarterly* 13, 319-340 (1989)
41. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: *Multivariate data analysis: a global perspective*. Pearson, Upper Saddle River et al. (2010)
42. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88, 879-903 (2003)
43. Harman, H.H.: *Modern factor analysis*. University of Chicago Press, Chicago (1976)

44. Anderson, J.C., Gerbing, D.W.: Structural Equation Modeling in practice: a review and recommended two-step approach. *Psychological Bulletin* 103, 411-423 (1988)
45. Jöreskog, K.G., Sörbom, D.: LISREL 8: Structural Equation Modeling with the SIMPLIS command Language. Scientific Software International, Inc., Chicago, IL (1998)
46. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* XVIII, 39-50 (1981)
47. Hu, L.-t., Bentler, P.M.: Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1), 1-55 (1999)
48. Holak, S.L., Lehmann, D.R.: Purchase intentions and the dimensions of innovation: An exploratory model. *Journal of Product Innovation Management* 7, 59-73 (1990)
49. Featherman, M.S., Pavlou, P.A.: Predicting e-services adoption: a perceived risk facets perspective. *International Journal of Human-Computer Studies* 59, 451-474 (2003)
50. Schaupp, L.C., Carter, L.: The impact of trust, risk and optimism bias on E-file adoption. *Information Systems Frontiers* 12, 299-309 (2010)
51. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. *MIS Quarterly* 27, 425-478 (2003)
52. Fishbein, M., Ajzen, I.: *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley (1975)
53. Hui-Chih, W., Her-Sen, D., Feng-Chang, L.: Determinants of E-government service adoption: an innovation diffusion perspective. In: *International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 3458-3461 (2007)
54. Hoffman, D.L., Novak, T.P., Peralta, M.: Building Consumer Trust Online. *Communications of the ACM* 42, 80-85 (1999)
55. Wang, Y.-S.: The adoption of electronic tax filing systems: An empirical study. *Government Information Quarterly* 20, 333-352 (2003)
56. Carter, L., Schaupp, L., Evans, A.: Antecedents to e-File adoption: the U.S. perspective. In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, pp. 1-10 (2008)
57. Hussein, R., Mohamed, N., Ahlan, A.R., Mahmud, M.: E-government application: an integrated model on G2C adoption of online tax. *Transforming Government People, Process and Policy* 5, 225-248 (2011)

The Coevolution of Network Structure and Perceived Ease of Use

Johannes Putzke¹, Kai Fischbach², Detlef Schoder¹, and Peter Gloor³

¹ University of Cologne, Cologne, Germany
{putzke, schoder}@wim.uni-koeln.de

² University of Bamberg, Bamberg, Germany
kai.fischbach@uni-bamberg.de

³ MIT Center for Collective Intelligence, Cambridge, USA
pgloor@mit.edu

Abstract. Perceived Ease of Use (PEoU) is one of the most central constructs in IS research. However, it has been examined only from an individual perspective. This article conceptualizes PEoU as a network construct. Results indicate three things. Firstly, the higher a person's PEoU, the more likely she or he is sought for advice. Secondly, there is a greater likelihood that a person seeks out another person for advice if the other person also seeks out the first person for advice. Thirdly, a person's PEoU will become similar to that of other persons' she or he seeks out for advice.

Keywords: SIENA, Social Network Analysis, Technology Acceptance Model, TAM

1 Introduction

While there are only a few constructs as central in IS research as Perceived Ease of Use (PEoU), research on PEoU has focused the individual perspective only. While an actor's embeddedness in a social network may exert a significant influence on the actor's attitudes, few works acknowledge that an actor's attitudes and behavioral intentions are influenced by that actor's embeddedness in a social network (compare [1]). For example, it is likely that an actor's PEoU influences and is influenced by the PEoU of her or his surrounding peers, which would mean, for example, that if an actor perceives an information system to be easy to use, that perception should spread to her or his neighbors in a social network (and vice versa). Nevertheless, this mutual influence of PEoU has not yet been researched; the reason may be attributable to the fact that, until recently, no adequate statistical methods to test such kind of hypothesis had been developed. Now, however, some new statistical methods allow for the creation of such models.

Research of this kind is of great importance. For example, managers who have to choose which employees should be given further vocational training may wish to consider the employee's embeddedness in a social network.

Hence, the current work has three main objectives:

1. Conceptualize PEOU as a network construct. In doing so, PEOU should be conceptualized both as an antecedent to network structure and as an outcome of network structure.
2. Propose a model that incorporates PEOU as a network construct.
3. Validate the model empirically.

This article is organized as follows. Section 2, Theory, reviews the related literature and develops four hypotheses. Section 3, Method, describes the study's context, participants, measurement, and modeling approach. Section 4, Results, highlights our findings. Section 5, Discussion, discusses the theoretical and managerial implications of our findings. Section 6, Conclusions, summarizes the results, notes their limitations, and provides some suggestions for further research.

2 Theory

The first part of the literature review highlights the conceptual underpinnings of Perceived Ease of Use. The second part highlights models that examine the evolution of social networks.

2.1 Technology Acceptance and Perceived Ease of Use

Perceived Ease of Use (PEoU) belongs to the most central construct in IS research. It is a core construct of the technology acceptance model (TAM) [2], which is one of the most firmly established models in IS research (compare [3-4]). Perceived Ease of Use is defined as "the degree to which a person believes that using a particular system would be free of effort" [2]. Related constructs can also be found in myriad other theories such as the theory of planned behavior, innovation diffusion theory, and social cognitive theory [5]. In these theories, the related constructs are referred to as, for example, "perceived behavioral control", "self-efficacy", and "complexity." Perceived behavioral control is defined as "the perceived ease or difficulty of performing the behavior and is assumed to reflect experience as well as anticipated future impediments and obstacles" [6]. Complexity is defined as the "opposite of ease of use" [7], that is, "the degree to which an innovation is perceived as relatively difficult to understand and use." Finally, (Computer) Self-Efficacy is defined as "a judgement of one's capability to use a computer. It is not concerned with what one has done in the past, but rather with judgements of what could be done in the future" [8].

Despite the high prominence of these constructs in IS research, the fact that PEOU has been conceptualized only from an individual's perspective (i.e., residing within each individual and isolated from other individuals) leaves a notable gap in the research. The individuals' embeddedness in a social network has been neglected nearly completely (for an exception, see, [1]). Rather, the influence of the other actors on a focal individual is captured by constructs such as "subjective norm," defined as "the

person's perception that most people who are important to him think he should or should not perform the behavior in question" [9].

However, the models do not hypothesize how PEOU is influenced by network structure, and how PEOU influences network structure, which leads to the second focus of our literature review.

2.2 Social Network Perspective, Embeddedness and the Evolution of Social Networks

In this paper, we assume that each focal individual (ego) is embedded in a social network of alters (see Figure 1).

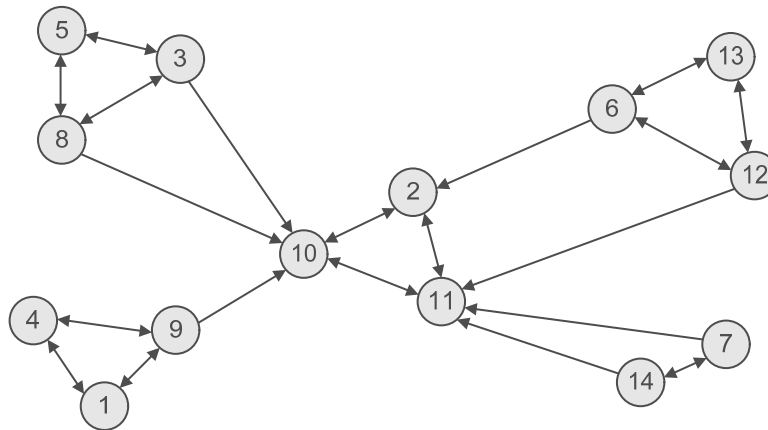


Fig. 1. Social Network

Figure 1 illustrates a networked perspective on PEOU. In the figure, each circle represents an individual. A pair of individuals is connected by an arrow if one actor seeks advice from the other actor.

As we show in the hypotheses development, we assume that PEOU spreads through this social network through a process of contagion. This process of contagion can be modeled with other models that examine the co-evolution of social networks and actor characteristics. Here we review these types of models. Early models examining the evolution of social networks (for an introduction, see [10]) primarily explore how structural characteristics of networks (such as transitivity, reciprocity, and degree-centrality) influence the process of network evolution (see, e.g. [11]).

Recent developments in these types of models (see, e.g. [12]) now also allow for the integration of several actor characteristics, which are permitted to co-evolve with the social network over time (e.g. [13]) so that the social network (and the actor characteristics) can be dependent and independent variables concurrently. Hence, these models allow for statistical tests of causal relationships between network structure and actor characteristics (such as Perceived Ease of Use) that were not previously possible.

2.3 Hypotheses Development

We structure the development of our research hypotheses in two parts. In the first part, we propose two hypotheses regarding the co-evolution of social network and Perceived Ease of Use. The first of these addresses Perceived Ease of Use as an antecedent to network evolution. The second explains Perceived Ease of Use as an outcome of network structure. In the second part of our hypotheses development, we propose two hypotheses that examine the effects of endogenous¹ network variables on the process of network evolution. These two variables serve as control variables in our study.

Perceived Ease of Use. We begin with a hypothesis that considers Perceived Ease of Use as an antecedent to network structure. Generally, a person who perceives a computer system to be easy to use has a high expertise in using the system (compare [14]). Furthermore, the greater a person's expertise, the more likely he or she is sought out for advice by other people [15]. In summary,

Hypothesis 1 (PEoU alter): The higher a person's Perceived Ease of Use, the more likely she or he is sought out for advice.

The second hypothesis concerns Perceived Ease of Use as an outcome of network structure, and is based on social influence theories, one of the most prominent being social comparison theory (e.g. [16-17]). Social comparison theory assumes that individuals compare themselves with their surrounding peers. During these comparisons, they adapt their individual attributes so that they become more like their peers. In a similar vein, we assume in this paper that people compare their Perceived Ease of Use to each other. We expect that people who observe others with a high PEoU of a system will perceive the system easier to use for themselves. Conversely, we expect that people who observe others that perceive a system difficult to use will perceive the system difficult to use for themselves. Hence,

Hypothesis 2 (PEoU similarity): Over time, the likelihood that a person's Perceived Ease of Use will become similar to that of other persons' she or he seeks out for advice is greater than a random change in Perceived Ease of Use.

Endogenous Network Effects. We next propose two hypotheses that examine the effects of endogenous network variables on the process of network evolution. These two variables serve as control variables. As our first endogenous network hypothesis, we hypothesize that people do not seek advice for free. Rather, building and maintaining relationships for seeking advice is associated with some cost (c.f. [18]). Consequently, people who are embedded in an advice network with many partners are less likely to seek out new people for advice than are people who have only a few con-

¹ In this paper, we use the term "endogenous" in the sense of Contractor et al. (2006), that is, endogenous variables do not incorporate factors other than the focal relationship itself. In particular, they may not include attributes of the actors in the network.

tacts. Numerous studies that examine the scale-free property of social networks (e.g. [19]) and models of dynamic network evolution (e.g. [13]) support this finding. Hence,

Hypothesis 3 (Density): The higher the number of a person's partners for seeking advice, the lower the likelihood that he or she will seek out new partners for advice.

As a second endogenous network hypothesis we hypothesize mutual/ reciprocated ties, that is, if person i seeks out person j for advice, person j should also seek out person i for advice. Several theories explain reciprocal ties, including social exchange theory [20], resource dependency theory [21], and network exchange theory [22], compare [23]. For example, social exchange theory assumes that relationships are built through an individual cost-benefit analysis: the benefits for an individual are the positive elements of a relationship such as advice, support, or friendship; the costs are the effort required to maintain a relationship. In general, scholars agree that relationships evolve over time into trusting, loyal, and mutual commitments as actors obey certain rules of exchange. Reciprocity is probably the best-known exchange rule [24]; it means, in theory, giving advice to someone is considered to be a previous investment in a relationship that must be reciprocated. However, some researchers state that relationships in advice networks tend to be nonreciprocating/ asymmetric, that is, a less well-informed actor is more likely to seek advice from a more well-informed actor than vice versa (e.g. [15], [25-26]). One possible explanation for non-reciprocal dyads is actors striving for social status, compare [25]. Actors that have acquired a certain social status by being sought out for advice will attempt to preserve this status advantage by seeking advice from third parties rather than from the actors that have sought them out for advice. Another possible explanation for non-reciprocated dyads is that actors sought out for advice by a certain individual i will not seek advice from this particular individual, as they doubt the individual's capabilities.

Nevertheless, we assume that the last two explanations do not hold for this study. First, asking someone for advice in the learning situation of the study was not associated with a loss of status. Furthermore, people who sought out others for advice were not perceived to be unknowledgeable. Hence,

Hypothesis 4 (Reciprocity): There is a greater likelihood that a person seeks out another person for advice if the other person also seeks out the first person for advice.

3 Method

To test the proposed hypotheses, we used data collected during a June 2010 PhD course on social network analysis taught in the IS department of a leading research university.

3.1 Context

The aim of the five-day course was to familiarize students with a software system for longitudinal social network analysis that had been co-developed by one of the authors. Morning sessions of the course were devoted to training; in the afternoons, when students applied the software, they were free to seek advice from other course participants. Most students were first-year PhD students. Rather than grades, a certificate was awarded for attendance, and participation in the course was voluntary. Three participants had extensive experience with the social network analysis software taught in the course.

3.2 Participants

The unit of analysis in our study is the individual student. The course had 15 participants, one of whom dropped out during the course. Before our final analysis, the student who dropped out was excluded due to an extensive amount of missing data. Participation in the course formed an appropriate boundary for our study because the class members interacted in the context of the system that bound them with interdependent processes and a shared symbol system [1]. There was one woman among the 14 participants, which is typical for a course taught in an IS department. The average age of the respondents was 32.46 years, with a standard deviation of 7.95 years. The youngest participant was 24 years old; the oldest was 48. Although the sample size is quite small ($n = 14$), it is sufficient for the proposed methodology since the model has only 3 predictors.

3.3 Measurement

Like Sykes et al. [1], we collected data with a survey administered before students used the new system immediately after the first training session. Furthermore, we collected data on the third and on the fifth days at the conclusion of the afternoon sessions.

We measured Perceived Ease of Use with 5 items (compare [2], [5], [27]): (1) Learning to operate <the system> is easy for me; (2) I find it easy to get <the system> to do what I want it to do; (3) It is easy for me to become skillful at using <the system>; (4) I find <the system> easy to use; (5) My interaction with <the system> would be clear and understandable. All items were measured using seven-point se-

semantic differentials anchored with “strongly disagree” and “strongly agree” as well as numbers from -3 to +3. Later, we aggregated the 5 items to a single value for PEOU.

Furthermore, we asked the following question: During the last <n> days of the seminar, how often and for how long did you seek advice from the following persons? We used a seven-point semantic differential anchored with <never> to <very long and often> for this question. We later dichotomized the answers as described in Sykes et al. [1]. In the following section, $X(t) = X_{ij}(t)$ denotes an $n \times n$ adjacency matrix, where $X_{ij} = 1(0)$ represents a tie (no tie) from actor i to actor j ($i, j = 1, \dots, n$) in period t , that is, player i responded at least “0” on the semantic differential.

3.4 Model

To examine the dynamic co-evolution of network structure and Perceived Ease of Use, we employed a stochastic actor-driven modeling approach proposed by Snijders, e.g. [13], [28-29]. The first application of this methodology in IS research is a recent article by Putzke et al. [30].

The advantage of Snijders’ methodology is that the same variable can be interpreted concurrently as both an independent and a dependent variable, as we show in the following paragraphs. This makes it possible to establish a causal relationship between structural network variables and PEOU.

Snijders models the co-evolution of network structure and actor characteristics as a continuous-time Markov process $Y(t) = (X(t), Z_{h1}(t), \dots, Z_h(t))$ on the space of adjacency matrices $X(t)$ as well as actors’ characteristics Z_{hi} ($h = 1, \dots, H$) (in this case, PEOU). To derive a transition matrix between two states $y(t_m)$ and $y(t_{m+1})$, Snijders decomposes each change between the states into so-called “micro steps.” Micro steps are randomly determined moments in time whose queue time follows an exponential distribution with rate parameters $\lambda_i^{[X]}$ and $\lambda_i^{[Z_h]}$ that we assume to be constant and independent between actors. At these randomly determined moments in time, one of the actors has the opportunity either to: change a tie variable X_{ij} (i.e. $\hat{y} = (x(i \Rightarrow j), z)$); change his or her own characteristics Z_{hi} by δ (i.e. $\hat{y} = (x, z(i \uparrow_h \delta))$); or change nothing (i.e. $\hat{y} = y$). These changes occur with probabilities $p = (x(i \Rightarrow j) | x(t), z(t))$ and $p = (z(i \uparrow_h \delta) | x, z)$ respectively. Whereas $p = (x(i \Rightarrow j) | x(t), z(t))$ denotes the probability that actor i changes its tie to actor j (conditioned on all other ties being constant, and given actor characteristics), $p = (z(i \uparrow_h \delta) | x, z)$ denotes the probability that actor i ’s characteristic h will decrease or increase by δ . To obtain transition intensities, Snijders multiplies the rate functions by the probabilities of an actual change taking place, which leads to the transition matrix

$$q_{ij} = \begin{cases} \lambda_i^{[X]}(y) p(x(i \Rightarrow j) | x, z) & \text{if } \hat{y} = (x(i \Rightarrow j), z), \\ \lambda_i^{[Z_h]}(y) p(z(i \uparrow_h \delta) | x, z) & \text{if } \hat{y} = (x, z(i \uparrow_h \delta)), \\ -\sum_i \left\{ \sum_{j \neq i} q(y; (x(i \Rightarrow j), z)) + \sum_{\delta \in \{-1, 1\}} q(y; (x, z(i \uparrow_h \delta))) \right\} & \text{if } \hat{y} = y, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

To estimate the full model, the change probabilities $p = (x(i \Rightarrow j) | x(t), z(t))$ and $p = (z(i \Rightarrow \delta) | x, z)$ have to be specified, which we do as discrete choice models in multinomial logit form (cf. McFadden, 1974). That is

$$p = (x(i \Rightarrow j) | x(t), z(t)) = \frac{e^{u_i^{[X]}(\beta, x(i \Rightarrow j)(t), z(t))}}{\sum_k e^{u_k^{[X]}(\beta, x(i \Rightarrow j)(t), z(t))}} \quad (2)$$

where $u_i^{[X]}$ denotes the deterministic part of a utility function that actor i attributes to the network configuration. For example, a utility function that allows only a test of H3 (density) and H4 (reciprocity) might be defined as

$$u_i^{[X]}(\beta^{[X]}, y) = \beta^{density} \sum_j x_{ij} + \beta^{reciprocity} \sum_j x_{ij} x_{ji} \quad (3)$$

Analogously, the formulas for the behavioral evolution of PEOU can be derived (for a more detailed discussion, see [13]).

The following effects were included in the utility functions:² The *PEoU-alter* effect was measured as $\sum_j x_{ij} PEOU_j$ (cf. hypotheses 1), that is, actor i 's utility function $u_i^{[X]}$ increases by actor j 's PEOU if actor i seeks advice from actor j (i.e. $x_{ji} = 1$). Hence, a positive parameter $\beta^{PEoU\ alter}$ indicates that students with a higher PEOU are more likely to be sought out for advice than are students with a lower PEOU.

PEoU average similarity (c.f. hypothesis 2) was measured as

$$\frac{1}{x_{i+}} \sum_j x_{ij} \left(\frac{\overline{\max_{ij} |PEoU_i - PEoU_j| - |PEoU_i - PEoU_j|}}{\max_{ij} |PEoU_i - PEoU_j|} - \frac{\overline{\max_{ij} |PEoU_i - PEoU_j| - |PEoU_i - PEoU_j|}}{\max_{ij} |PEoU_i - PEoU_j|} \right) \quad (4)$$

where $\frac{\overline{\max_{ij} |PEoU_i - PEoU_j| - |PEoU_i - PEoU_j|}}{\max_{ij} |PEoU_i - PEoU_j|}$ is the mean of all similarity scores $\frac{\max_{ij} |PEoU_i - PEoU_j| - |PEoU_i - PEoU_j|}{\max_{ij} |PEoU_i - PEoU_j|}$.

The similarity score between actor i and actor j calculates the difference between the PEOU of the two actors $|PEoU_i - PEoU_j|$ (in absolute values) and standardizes this difference by the range of all actors' PEOU $\max_{ij} |PEoU_i - PEoU_j|$. Hence a positive parameter $\beta^{PEoU\ similarity}$ indicates that an actor's PEOU tends to become similar to the PEOU of those actors she or he seeks out for advice. However, the total influence of the actor she or he seeks out for advice is the same regardless of their number.

A general *drive toward high PEOU* (linear shape effect) was measured as $PEoU_i$. This effect was added as an additional control variable, since course participants became more familiar with the software during the course and hence would be expected to perceive the software to be easier to use over time.

General tendency to seek advice from alters (density/outdegree effect) is measured as

² See Snijders et al. (2007) for more information about the measures.

$$\sum_j x_{ij} \tag{5}$$

(cf. hypothesis 3), that is, actor i 's utility function $u^{[x]}$ increases by value 1 if actor i seeks out actor j for advice, because the corresponding value in the adjacency matrix x_{ij} equals 1 if actor i seeks advice from actor j (and is 0 otherwise). Consequently, a negative parameter $\beta^{outdegree}$ indicates that actor i does not seek advice randomly, but that each occasion of seeking advice is associated with some "cost" for actor i .

Number of mutual ties (reciprocity) is measured as

$$\sum_j x_{ij}x_{ji} \tag{6}$$

(cf. hypothesis 4), that is, actor i 's utility function $u^{[x]}$ increases by value 1 only if actor i seeks advice from actor j ($x_{ij} = 1$) and actor j seeks advice from actor i ($x_{ji} = 1$). If one of these ties is missing (i.e. $x_{ij} = 0$ or $x_{ji} = 0$), the product will equal 0. Consequently, a positive parameter $\beta^{reciprocity}$ indicates a greater likelihood that actor i seeks advice from actor j if actor j also seeks advice from student i .

4 Results

We conducted a nested model comparison to test the proposed hypothesis, see [29]. All models were estimated using RSiena and RSienaTest 1.0.12.186. In a series of Neyman-Rao tests, we compared a model that allows both PEOU effects to vary freely against a baseline model that restricts one (or both) PEOU parameters to zero, but includes all control variables. We report no measure of explained variation because there are, as yet, no satisfactory measures for this stochastic actor-driven modeling approach.

Table 1. Model Results

	Beta	s.d.	t-value	p-value
Network Dynamics				
Rate Parameter (t=1)	2.310	.837	2.760	.003
Rate Parameter (t=2)	1.334	.468	2.850	.002
Outdegree	-1.697	.381	-4.454	<.001
Reciprocity	3.279	.739	4.437	<.001
PEoU alter	0.120	.053	2.264	.011
Behavior Dynamics				
Rate Parameter (t=1)	5.772	2.3642	2,441	.007
Rate Parameter (t=2)	2.780	.984	2.825	.002
PEoU linear shape	2.498	5.999	0.416	.339
PEoU average similarity	295.900	.444	666.441	<.001

The series of Neyman-Rao tests indicate that the inclusion of both PEOU effects in the model (see Table 1) at the same time increases model fit, and that the increased model fit can be attributed to both effects, that is, to PEOU alter ($\chi^2 = 6.642$; $d.f. = 1$; $p < .01$) as well as to PEOU similarity ($\chi^2 = 7.223$; $d.f. = 1$; $p < .01$).

Furthermore, both effects are found to be statistically significant and in the expected direction. Hence, Hypothesis 1 and Hypothesis 2 are supported, and we can conclude that 1) students are more likely to seek advice from a partner that perceives the system as easy to use, and 2) that the students' PEOU over time tends to become similar to those of their surrounding peers.

Concerning the control variables (Hypothesis 3 and Hypothesis 4), the results are in line with our expectations. The negative outdegree effects indicate that seeking ties is associated with some costs and the positive reciprocity effect indicates that students are more likely to seek advice from students who sought advice from them. However, the linear shape effect turns out to be positive but non-significant. Hence, there is limited support for the proposition that course participants perceived the software to be easier to use over time.

Further results can be found in the Appendix, where Table 2 provides the correlations of estimates and Table 3 highlights the tie changes between subsequent observations. Indeed, the Jaccard coefficients (see Table 3) fall within acceptable levels.

We further tested for time heterogeneity in parameters between both time periods, compare [31]. In the original model, the objective function was defined as:

$$f_i^{net}(x) = \sum_k \beta_k^{net} s_{ik}^{net}(x), \quad (7)$$

where $s_{ik}^{net}(x)$ were the effects as defined above. This means that all parameters β were assumed to be stable over time. To test for time heterogeneity in the parameters, the objective function was defined as:

$$f_{ij}^{(a)}(x) = \sum_k (\beta_k + \delta_k^{(a)} h_k^{(a)}) s_{ik}(x(i \rightsquigarrow j)) \quad (8)$$

where $h_k^{(a)}$ is a vector of time dummies and $\delta_k^{(a)}$ is the time dummy interacted effect parameter for effect k in period a . The dummy variables $\delta_k^{(1)}$ were assumed to be 0 for all k , so that the first period is considered to be the base period. A joint score test for parameter heterogeneity is then given as:

$$H_0: \delta_k^{(a)} = 0 \forall k, a \quad (9)$$

$$H_1: \delta_k^{(a)} \neq 0 \text{ for some } k, a \quad (10)$$

For the further estimation of the test, see [31].

The results of a joint test of dummy parameters for all effects (outdegree, reciprocity, PEOU alter, PEOU linear shape, PEOU similarity) revealed that it is not necessary to introduce time dummies for a particular effect ($p = .746$).

Further, we assessed goodness of fit using the indegree distribution as auxiliary statistic (see Figure 2) of a Monte Carlo Mahalanobis Distance Test (compare [32]). Results indicate that goodness of fit is good (Mahalanobis Distance = 16.384; $p = .173$).

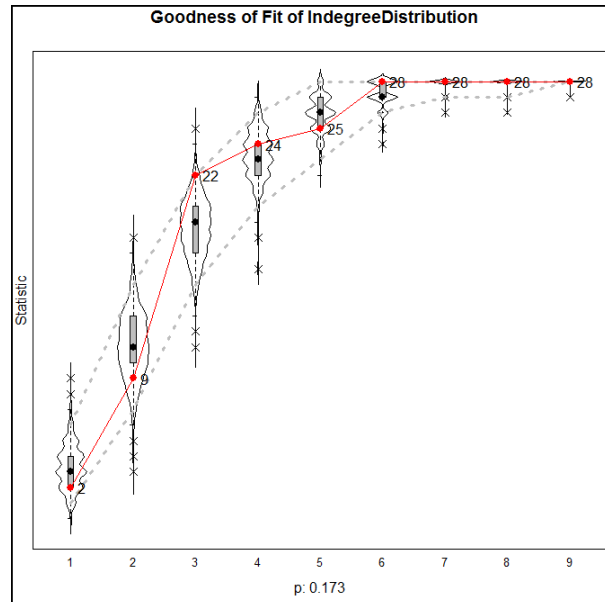


Fig. 2. Goodness of Fit of Indegree Distribution

5 Discussion

In this study, we theorized PEoU as a network construct. In particular, we hypothesized that people are more likely to seek advice from a partner who perceives a system as easy to use. Furthermore, we hypothesized that people's PEoU over time tends to become similar to that of their surrounding peers. The empirical study lent support to both hypotheses.

5.1 Theoretical Contributions and Implications

The paper contributes to IS research in several ways. First, the model is a fundamental shift in our understanding of PEoU. Whereas PEoU used to be examined from an individual perspective in structural equation models, the results show that PEoU is an antecedent as well as an outcome of network structure. Future research should, therefore, take a network perspective on PEoU.

Second, the paper introduced a new methodology from sociology into IS literature. Apart from [30], it is the first paper in IS research that uses a methodology that can examine the co-evolution of actor characteristics and social network. However, this paper exceeds the paper [30] as it tests for time heterogeneity and reports goodness of fit using the indegree distribution as auxiliary statistic of a Monte Carlo Mahalanobis Distance Test.

5.2 Practical Implications

In addition to the theoretical insights, this study also has some implications for practitioners. Results showed that people's PEOU tends over time to become similar to that of their surrounding peers, which means that managers should not regard PEOU as a construct only on the individual level. Rather, managers should examine each individual's embeddedness in a social network before selecting who will take part in a system training course. With the careful selection of individuals, PEOU will spread through the network and, as a consequence, less people will need formal training courses. This will result in cost savings and enhanced organizational performance. This study offered a practical insight that the selection of individuals for training courses should be driven not only by their personal characteristics, but also to some extent by their position in the social network.

Another managerial implication is that managers should pay special attention to the structure of the social networks within and between organizational units when seeking to improve performance outcomes through information systems. Only active management of such networks will optimize network flow and assure that PEOU will spread through the network. Enhancing the level of PEOU in such a way may also prove useful for reducing IT resistance in enterprise systems implementation projects.

Finally, this paper introduced a new methodology from sociology into the IS literature. The proposed methodology can be applied by IS practitioners in a variety of different contexts. For example, practitioners can use the proposed methodology for link prediction (e.g., on social networking platforms).

6 Conclusions

In this article, we conceptualized PEOU as a network construct. Our results indicate that people are more likely to be sought out for advice the higher their PEOU. Furthermore, there is a greater likelihood that a person seeks out another person for advice if the other person also seeks out the first person for advice. Finally, we found that a person's PEOU will become similar to that of other persons' she or he seeks out for advice.

Of course, as with any empirical study, ours is subject to some limitations that could be seen as affecting the rigor and relevance.

First, we examined a single construct from TAM only. We neglected Perceived Usefulness as well as the user's Behavioral Intention to use a system. Future research should examine these constructs in more detail. The proposed methodology offers an interesting way to examine the co-evolution of social networks and these two constructs.

Second, we examined only one type of network (i.e., the advice network). However, there may be other networks that influence an actor's PEOU, such as a friendship network or acquaintance network. Future research should examine related questions with networks other than the advice networks.

There are several fruitful areas where the methodology can be employed in IS research. For example, future research can examine the adoption of information tech-

nologies taking a network perspective. Basically, the method is appropriate for various types of analyses that examine the co-evolution of social network and actor characteristics. Our hope is that our research will assist others in conducting these types of studies and form the basis for substantial future research into the co-evolution of social networks, attitudes, and behavioral intentions.

7 Appendix

Table 2. Correlations of Estimates

Rate Parameter (t=1)	1								
Rate Parameter (t=2)	-0.068	1							
Outdegree	-0.292	-0.138	1						
Reciprocity	0.046	0.017	-0.597	1					
PEoU alter	0.004	0.035	-0.467	0.417	1				
Rate Parameter (t=1)	-0.202	0.085	0.040	-0.046	0.019	1			
Rate Parameter (t=2)	-0.155	0.101	-0.019	0.031	0.065	0.560	1		
PEoU linear shape	0.193	-0.129	0.003	0.041	-0.073	-0.826	-0.699	1	
PEoU average similarity	-0.195	-0.016	0.029	0.019	-0.115	-0.205	0.012	0.070	1

Table 3. Tie changes between subsequent observations

	0 => 0	0 => 1	1 => 0	1 => 1	Jaccard Index
from period 1 to period 2	155	13	2	12	.444
from period 2 to period 3	147	10	2	23	.657

References

1. Sykes, T.A., Venkatesh, V., Gosain, S.: Model of Acceptance with Peer Support: A Social Network Perspective to Understand Employees' System Use. *MIS Quarterly* 33, 371-393 (2009)
2. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 319-340 (1989)
3. Hu, P.J., Chau, P.Y.K., Sheng, O.R.L., Tam, K.Y.: Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems* 16 (2), 91-112 (1999)
4. Venkatesh, V.: Where To Go From Here? Thoughts on Future Directions for Research on Individual-Level Technology Adoption with a Focus on Decision Making*. *Decision Sciences* 37 (4), 497-518 (2006)
5. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. *MIS Quarterly* 27, 425-478 (2003)
6. Ajzen, I.: The Theory of Planned Behavior. *Organ Behav Hum Dec* 50, 179-211 (1991)

7. Thompson, R.L., Higgins, C.A., Howell, J.M.: Personal Computing - toward a Conceptual-Model of Utilization. *MIS Quarterly* 15, 125-143 (1991)
8. Compeau, D.R., Higgins, C.A.: Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Quarterly* 19, 189-211 (1995)
9. Fishbein, M., Ajzen, I.: Belief, attitude, intention and behaviour: An introduction to theory and research. Addison-Wesley, Reading, MA (1975)
10. Doreian, P., Stokman, F.N.: Evolution of social networks, Vol. 1. Routledge (1997)
11. Wasserman, S., Pattison, P.: Logit models and logistic regressions for social networks .1. An introduction to Markov graphs and p. *Psychometrika* 61, 401-425 (1996)
12. Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P.: Recent developments in exponential random graph (p*) models for social networks. *Soc Networks* 29, 192-215 (2007)
13. Snijders, T.A.B., Steglich, C.E.G., Schweinberger, M.: Modeling the co-evolution of networks and behavior. In: Montford, K.v., Oud, H., Satorra, A. (eds.): Longitudinal models in the behavioral and related sciences. Lawrence Earlbaum, Newark, NJ (2007)
14. Li, X., Troutt, M.D., Brandyberry, A., Wang, T.: Decision factors for the adoption and continued use of online direct sales channels among SMEs. *Journal of the Association for Information Systems* 12 (4) (2011)
15. Nebus, J.: Building collegial information networks: A theory of advice network generation. *Academy of Management Review* 31 (3), 615-637 (2006)
16. Festinger, L.: A theory of social comparison processes. *Human Relations* 7, 117-140 (1954)
17. Felps, W., Mitchell, T.R., Hekman, D.R., Lee, T.W., Holtom, B.C., Harman, W.S.: Turnover Contagion: How Coworkers' Job Embeddedness and Job Search Behaviors Influence Quitting. *Academy of Management Journal* 52 (3), 545-561 (2009)
18. Bolton, P., Dewatripont, M.: The Firm as a Communication-Network. *Quarterly Journal of Economics* 109 (4), 809-839 (1994)
19. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. *Science* 286, 509-512 (1999)
20. Blau, P.M.: Exchange and power in social life. Transaction Publishers, New Brunswick, New Jersey (1992)
21. Pfeffer, J., Salancik, G.R.: The external control of organizations: A resource dependence perspective. Stanford University Press, Stanford, CA (2003)
22. Willer, D.: Network exchange theory. Praeger Publishers, Westport, CT (1999)
23. Contractor, N.S., Wasserman, S., Faust, K.: Testing Multitheoretical, Multilevel Hypotheses about Organizational Networks: An Analytic Framework and Empirical Example. *The Academy of Management Review* 31, 681-703 (2006)
24. Cropanzano, R., Mitchell, M.S.: Social exchange theory: An interdisciplinary review. *J Manage* 31, 874-900 (2005)
25. Agneessens, F., Wittek, R.: Where do intra-organizational advice relations come from? The role of informal status and social capital in social exchange. *Soc Networks* (in press)
26. Carley, K.M., Krackhardt, D.: Cognitive inconsistencies and non-symmetric friendship. *Soc Networks* 18, 1-27 (1996)
27. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer-Technology - a Comparison of 2 Theoretical-Models. *Manage Sci* 35, 982-1003 (1989)
28. Snijders, T.A.B.: Stochastic Actor-oriented Models for Network Change. *Journal of Mathematical Sociology* 21, 149-172 (1996)
29. Snijders, T.A.B., van de Bunt, G.G., Steglich, C.E.G.: Introduction to Stochastic Actor-Based Models for Network Dynamics. *Soc Networks* 32, 44-60 (2010)

30. Putzke, J., Schoder, D., Fischbach, K., Gloor, P.A.: The Evolution of Interaction Networks in Massively Multiplayer Online Games. *Journal of the Association for Information Systems* 11 (2010)
31. Lospinoso, J.A., Schweinberger, M., Snijders, T.A.B., Ripley, R.M.: Assessing and accounting for time heterogeneity in stochastic actor oriented models. *Advances in Data Analysis and Classification* 147-176 (2011)
32. Lospinoso, J.A., Snijders, T.A.B.: Goodness of fit for Stochastic Actor Oriented Models. *Sunbelt XXXI, St. Pete's Beach, Florida* (2011)

A Structural Perspective on Belief Formation

Daniel Grgecic

Goethe-University Frankfurt, Frankfurt am Main, Germany
grgecic@wiwi.uni-frankfurt.de

Abstract. Research on IT adoption has shown that object-based beliefs about IT systems have a profound impact on subsequent IT usage. However, we still need to identify antecedents of object-based beliefs in order to understand how the belief formation process can be influenced. This research builds upon and extends Adaptive Structuration Theory to examine how IT-related factors influence the formation of object-based beliefs. To test our research model, we surveyed 183 users of a student information system. The proposed model was supported, providing evidence that values, meaning and functionalities provided by an IT system positively affect information and system quality.

Keywords: Structuration Theory, IS Adoption, Object-based Beliefs

1 Introduction

Information Technology (IT) pervades important aspects of human life at different levels, such as individuals, teams or organizations. The interplay between IT and individuals has been investigated, especially within two major research streams – the technology adoption and the user satisfaction literature – and both research streams have converged on a shared understanding of the salient predictors of individuals' acceptance and intentions to use IT [1-2]. An important and long-standing research question in the field of Information Systems (IS) research deals with individual beliefs about IT systems that have shown to have a large effect on subsequent IT behaviors [3-4]. While research on IT adoption and IS success has made important steps to understand which kind of beliefs contribute to a successful adoption process [5] there is still a lack of knowledge about the formation and antecedents of object-based beliefs [6]. This knowledge is crucial in order to inform researchers as well as practitioners how IT design influences users and how IT systems can be improved to foster the IT adoption process [7]. Practice can only benefit from adoption theories if researchers understand how to influence any kind of beliefs through IT design. This study builds upon and extends Adaptive Structuration Theory (AST) in order to examine IT-related factors that influence individual object-based beliefs about IT systems [8]. AST proposes that users are related to IT systems through two communication channels, namely functional affordance and symbolic expression [9]. Both communication channels can be regarded as sources for structure that determine to some degree how people use and interact with IT. Thus, AST offers a good starting point to

investigate how these IT-related sources for structure affect the formation of object-based beliefs and how these beliefs affect the usage of IT systems. This study seeks to apply AST in order to identify and empirically test antecedents of object-based beliefs which leads to the following research question: To what extent do the structures provided by IT systems affect the formation of object-based beliefs?

2 Theoretical Grounding

2.1 Structuration Theory and IT Adoption

AST serves as a foundation to develop a theoretical framework for the structural features of an IT system [8]. It is one of the most influential structural theories in IS research [10] that describes “the production and reproduction of social systems through members’ use of rules and resources in interaction” [11]. Structuration can be described as the process of social structures that shape peoples’ actions and beliefs and that are shaped by peoples’ actions. Thus, at its core AST is a holistic attempt to examine the interplay of advanced technologies, social structures, and human action [8], [12]. According to a recent reconceptualization of AST, technical objects are related to human agents through two concepts: “functional affordance” and “symbolic expression” [9]. These structures are not directly attributed to the technical object itself but to the relation between technical objects and users, thus this conceptualization emphasizes the importance of technology-human interactions. In other words, every user or user group perceives, understands, and grasps the structures that are provided by technical objects (functional affordance, symbolic expression) differently, thus the technology in use and the user (the relation between them) are inextricably connected and cannot be studied separately. The structure provided by an IT system determines to some degree the outcomes of human-technology interactions [8]. However, we have to differentiate between behaviors that are determined by IT systems and how IT systems are perceived by users before any action takes place. This differentiation is important because structures provided by IT systems are indirectly influencing IT behavior through the formation of object-based beliefs [5]. According to Fishbein and Ajzen [13], beliefs can be influenced by observations, by information that are received from any other source, or by different inference processes. One among many other sources that affect the formation of beliefs are the structures that are provided by an IT system [8]. Thus, IT structures can be regarded as antecedents of object-based beliefs. Object-based beliefs, such as information or satisfaction quality, in turn affect the use of IT systems. The assumed relationships that guide this study are summarized in Figure 1.

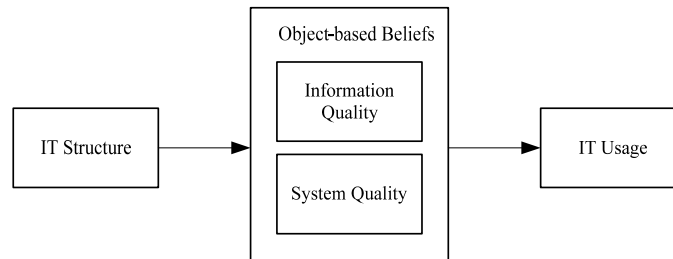


Fig. 1. Conceptual Research Framework

2.2 Object-Based Beliefs and IT Usage

IS literature is largely influenced by the Theory of Reasoned Action (TRA) [13]. According to TRA object-based beliefs can be understood as the perception a user has about an IT artifact. A user links information about an IT artifact to an attribute and is forming his/her attitude towards this special object in focus. The user satisfaction literature in general dealt to a large degree with the role of object-based beliefs as key constructs that determine the success of IT systems. One of the most important object-based beliefs is IS quality, that can be separated into information quality and system quality, both of which determine the user satisfaction and the use of an IT system [14-15]. In accordance with the IS Success Model information quality as well as system quality both will positively affect the usage of IT systems [14], [16].

H1: Information Quality will positively affect IT use.

H2: System Quality will positively affect IT use.

2.3 IT Structure and Object-Based Beliefs

The term “affordance” [17] refers to actionable properties between any real-world object and an actor. Affordances are relations between objects and actors in special situations and can be described as cues and instructions that are offered by an object to an individual in order to provide opportunities for particular types of individual behavior [18]. Functional affordances comprise “the possibility for goal-oriented action afforded by technical objects from designers to a specified user group (potential use of an IT object)” [9]. They are purposefully designed to assist and help users to accomplish tasks [19]. Therefore, the functional affordance of an IT system refers to the potential uses one can make of a technical object in order to achieve a certain goal. The concept of functional affordance provides a perspective that recognizes how features of certain technical objects favor, shape, invite, or at the same time constrain a set of specific uses [9]. A central assumption of AST is that IT systems are an important part of the structuration process and that the structural potential of IT systems elicits cognitive and behavioral reactions. The possibilities that technical objects afford for action may or may not be perceived by individuals in differing ways and therefore elicit different kinds of beliefs. This means that human-technology interactions are perceived and processed by users the same way as interpersonal interac-

tions and therefore lead to the formation of different beliefs based on users' direct experiences with IT systems. The interaction with IT systems will therefore positively influence the formation of object-based beliefs. Users who understand, grasp and ultimately know the functionalities provided by IT systems and know how the underlying functionalities can be used will therefore perceive a higher system quality and information quality [7].

H3: Functional Affordance will have a positive effect on Information Quality.

H4: Functional Affordance will have a positive effect on System Quality.

Similar to the concept of functional affordance, a symbolic expression is not a property of a technical object but a relational concept that connects technical objects and users. Symbolic expressions can be understood as "the communicative possibilities of technical objects for a specified user group" [9] that enable the interpretation of technical objects. For example, symbolic expressions include "messages" that help users interact with technical objects or functionalities, or messages pertaining to designers' or users' goals and values. Symbolic expressions are not to be confused with designer's intentions or user's perceptions. It is true that IT systems express "messages" and provide information that are intended by designers. However, they may also provide information that is not intended by designers and users may or may not perceive certain signs, symbols, or messages differently due to the fact that every user has a different background, expertise, or knowledge base. The conceptualization of symbolic expression is closely related to the conveyance of values [9], [20-21], even though the concept is not inherently limited to the domain of values. An expression can be understood as the manner or form in which a thing is expressed in words, or in the special case of a symbolic expression, in which a thing is expressed in any kind of symbol. The understanding of a symbol from the user perspective is as important as values that are conveyed by an IT system. While meaning or understanding of a symbol does also promote some kind of values, because the concept is inherently connected to values of a symbol, meaning is mostly considered as the user interpretation of an underlying real-world phenomenon (or abstract concept) that a symbol refers to [22]. Symbols serve as a means of communication and successful communication requires the know-how to produce the relevant signs/symbols with the intended meaning [23]. In general, IT systems can promote values such as control or reliability on an aggregate level; however, the understanding of perceptual cues needs to be considered in more detail as well. If a symbol is to convey meaning it must be identified by a user group and the symbol must communicate a similar meaning to all users within a group. For instance, concerning the example of Wikipedia, do the users understand what the meaning of the "edit button" is and how it has to be used? What this discussion amounts to is that we propose to subdivide the concept of symbolic expression into two distinct sub-dimensions: communication of values and communication of meaning. Communication of values deal with values that should be conveyed by an IT system to support certain functionalities or tasks. Communication of meaning comprises the understanding of functionalities that are provided by an IT system. Defining the concept this way has the advantage of supporting potential analyses of the relationships between functional affordances, symbolic expressions and object-

based beliefs in more detail. Thus, the understanding of an IT system (its functionalities) and the values conveyed by an IT system are in focus of the investigation. This conception also allows directly answering the question whether users understand the functionalities of IT systems. If users understand the meaning of symbols that are conveyed by the IT system they will perceive and understand the functionalities afforded by the IT system. The less effort is required for users to understand the meaning of an underlying functional affordance, the easier it will be to know what this functionality will do and ultimately how this functionality can be used. Thus:

H5: Communication of Meaning will positively affect Functional Affordance.

The same reasoning is applied to the relation between communication of values and functional affordance. If the values conveyed by an IT system support the general intent of the same IT system, users will be more prone to perceive and ultimately use the functionalities as intended.

H6: Communication of Values will positively affect Functional Affordance.

As has already been outlined, the structure provided by an IT system partially influences the formation of object-based beliefs. Users will attribute a higher quality to the information provided by an IT system if information are easily understandable. Thus, we assume a positive relationship between communication of meaning and information quality. As for the link between communication of meaning and system quality, we expect that users will perceive a higher system quality, if they generally understand the IT system and its functionalities.

H7: Communication of Meaning will positively affect Information Quality.

H8: Communication of Meaning will positively affect System Quality.

E-commerce literature has already investigated the effects of web assurance seals on trust and information disclosure [24-25]. These examples show how values are communicated by certain seals that are presented on websites and how these values influence personal beliefs or behaviors. If there is a perceived congruence between the values of the IT system and the values and goals of users [26], users will have a positive image of the underlying IT system and attribute positive values to the IT system. Therefore, we expect the following:

H9: Communication of Values will positively affect Information Quality.

H10: Communication of Values will positively affect System Quality.

The research model and underlying hypotheses are summarized in Figure 2.

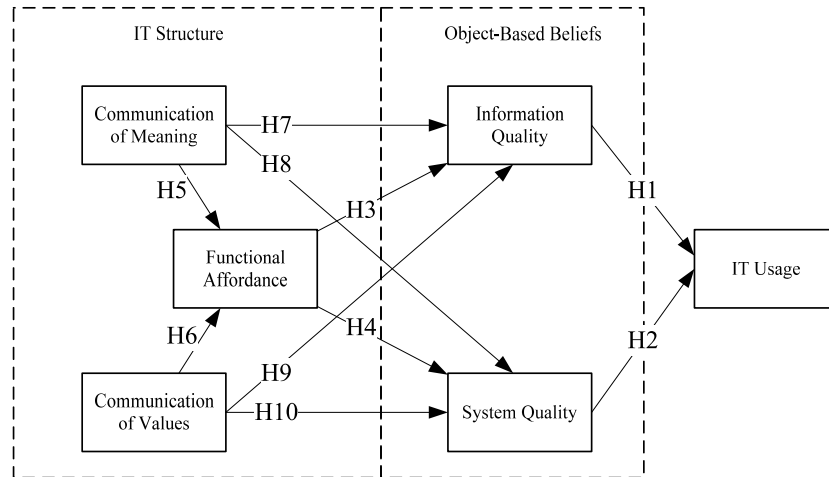


Fig. 2. Research Model

3 Methodology

3.1 Research Site and Data Collection

Data for this study was collected by surveying users of a computerized student information system (SIS) in place at a Western university. The IT system provides students with information about lectures, seminars, and courses. Moreover, it offers the possibility to plan and manage the entire semester. The use of the IT system is mandatory. Above this, the SIS is ideally suited to test the research model, since its functionalities as well as its scope are limited and the purpose of the system is precisely determined. The students attended a basic Information Systems course during the winter term 2011. Out of approximately 380 students a total of 200 students participated in the online questionnaire. After removing all questionnaires that were incomplete or not reliable a total of 183 usable questionnaires were received. Respondents ranged from 18 to 42 years of age, with a mean age of 21.8 (sd = 2.8). On average, students (106 male and 77 female students) were in their second semester (mean 2.2, sd = 0.88).

3.2 Measurement Model

We had to develop new items for the structural concepts. To ensure content validity, we followed the two-staged approach proposed by Burton-Jones and Straub [27]. First, four students were interviewed in order to find out about the most common functionalities and in order to understand the IT system from the point of view of a student. This procedure helped to make sure that all important functionalities and values were considered in the survey.

Table 1. Measurement Model (**p < 0.001, *p < 0.01, *p < 0.05)

Construct	Items	Description	Loadings
Communication of Meaning	COM1	I know how to download the offered course materials.	0.73***
	COM2	I know how to use the forums.	0.77***
	COM3	I know how to apply for a course.	0.7***
	COM4	Generally, I understand the basic functionality of the system.	0.86***
	COM5	In general, I understand how the system works.	0.88***
	COM6	I know how to use the system and its functionalities.	0.84***
Functional Affordance	FA1	The system offers the possibility to learn about the offered courses.	0.57***
	FA2	The system offers the possibility to download course materials.	0.61***
	FA3	The system offers the possibility to exchange opinions with other students.	0.77***
	FA4	The system provides information about seminars and lectures.	0.61***
	FA5	The system offers the possibility to use a forum.	0.76***
Communication of Values	COV1	Reliability	0.7***
	COV2	Effectiveness	0.87***
	COV3	Efficiency	0.92***
	COV4	Productivity	0.86***
	COV5	Control	0.75***
Information Quality [5], [28]	IQ1	In general, the system provides me with high-quality information.	0.8***
	IQ2	I am satisfied with the quality of the information.	0.84***
	IQ3	Overall, I would give the information from the system a high rating in terms of quality.	0.82***
System Quality [5], [28]	SQ1	Overall, I would give the quality of the system a high rating.	0.88***
	SQ2	Overall, the system is of high quality.	0.92***
	SQ3	In terms of system quality, I would rate the system highly.	0.93***
IT Use	USE1	I use the system on a regular basis.	0.84***
	USE2	I use the system to access my course materials.	0.81***
	USE3	I use the system regularly.	0.87***

Based on a previous study, the theoretical deliberations, and the interviews, measures were created that tie together the constructs in the research model and that seemed suitable to reflect the underlying causal relationships. All constructs, except functional affordance, were operationalized with reflective indicators [29]. All variables were measured using multiple items on 7-point Likert-type scales, ranging from “strongly agree” to “strongly disagree” (Table 1). Communication of meaning (COM) was measured using six reflective items. Students were asked if they knew and understood how different functionalities provided by the SIS work. Communication of values (COV) was operationalized using five reflective items. The set of relevant values was determined with the help of the interviews. Since the SIS serves as a tool to support students during their studies, the focus was on values that are associated with productivity. Students were asked to rate to what degree the SIS conveyed the proposed values. IT use (USE), information quality (IQ) and system quality (SQ) were also modeled as a reflective construct using 3 items per construct [5], [28]. The construct of functional affordance (FA) was operationalized as a formative construct [30]. The decision to model a formative construct was based on different criteria [29-30]. We were especially interested to find out what kind of different functionalities were offered by the SIS and to what extent these functions were important for the users. While one could easily define general items that ask to what extent the IT system as a whole is used, we wanted to focus on the most important features that build and therefore define the construct functional affordance. Therefore, changes in the variables influence the meaning of the formative construct. In other words, if different functionalities were added to the system the functional affordance of the SIS (thus what the IT system offers to do) would change significantly. In addition, the formative measures may not be interchangeable since every measure accounts for a unique dimension of the formative construct. We added control variables as IT use may vary across users with different demographic characteristics such as age, gender, degree and semester.

4 Data Analysis and Results

4.1 Factorial Analysis of Symbolic Expression

Before the research model was tested we conducted an exploratory factor analysis to check the two-dimensionality of symbolic expression. A principal components analysis was conducted on the 6 items for COM and the 5 items for COV with orthogonal rotation (varimax). The results of the analysis verified the sampling adequacy for the analysis (Kaiser-Meyer-Olkin measure (KMO) = 0.87). All KMO values for the individual items were higher than 0.79, which is above the acceptable limit of 0.5. Bartlett’s test of sphericity was significant ($\chi^2(91) = 1711.9, p < 0.001$). The analysis resulted in two components with eigenvalues over Kaiser’s criterion of 1 that explained altogether 70.62% of the variance. Given the large sample size, the analysis of the scree plot and the Kaiser’s criterion two components were retained in the analysis. The analysis of the rotated component matrix showed that all items that were assumed to be correlated loaded higher on one single component than on any other component

with factor loadings above 0.66. The items that cluster on the same components suggest that component 1 represents COM (Cronbach's $\alpha = 0.88$) and component 2 represents COV (Cronbach's $\alpha = 0.88$).

4.2 Scale Validation

The measures and the research model were tested by using SmartPLS 2.0 [31-32]. Internal consistency and convergent validity were examined by assessing item loadings, composite reliability, and average variance extracted (AVE). All factor loadings are significant (Table 1) and lie above the recommended threshold of 0.7 [31]. Composite reliabilities (CR) are above 0.8 and each AVE is above 0.50 (Table 2), indicating that the measurements are reliable and the latent construct can account for at least 50 percent of the variance in the items [33]. Discriminant validity was also achieved since the correlations between each pair of latent variables are less than the square root of AVE [34]. The traditional evaluation criteria such as factor loadings and AVE are not applicable for the evaluation of formative measurement models. Because these measures assume high internal consistency (high intercorrelating indicators) they are inappropriate for formative indicators, where no theoretical assumption is made about inter-item correlation [30], [35].

Table 2. Reliabilities and Correlation Matrix

Construct	Composite Reliabilities	AVE	COM	COV	FA	IQ	SQ	USE
COM	0.913	0.64	0,8					
COV	0.913	0.68	0.476	0,82				
FA	n/a	n/a	0.5221	0.408	n/a			
IQ	0.863	0.68	0.4738	0.483	0.446	0.82		
SQ	0.936	0.83	0.551	0.559	0.416	0.673	0.91	
USE	0.88	0.71	0.466	0.371	0.54	0.42	0.268	0.84

Diagonal elements represent the square root of the AVE. Off diagonal elements are the correlations.

For functional affordance, construct validity was assessed by using principal components analysis to examine the item weights for the measurement model [30]. The results show that three weights are significant while two weights (FA2 and FA4) are insignificant (Table 3). However, small absolute and insignificant weights should not inevitably be misinterpreted as a poor measurement model [31]. Instead one should further examine each indicator's weight (relative importance) and loading (absolute importance) [36]. For FA2 and FA4 the loadings are positive and significant, thus these items can be considered to be an important aspect of FA (Table 1).

Table 3. Factor Weights and Variance Inflation Factor (***p < 0.001, **p < 0.01, *p < 0.05)

Construct	Items	Outer Weights	VIF
Functional Affordance	FA1	0.32**	1.224
	FA2	0.258	1.325
	FA3	0.449**	1.354
	FA4	0.069	1.712
	FA5	0.363**	1.811

The insignificant weight of FA2 and FA4 should therefore be interpreted as their relative contribution to FA after controlling for the other functions. In sum, these results are consistent with the expectation that these five different functions might not be equally important to the functional affordance construct. To ensure that multicollinearity does not pose a problem, the variance inflation factor (VIF) statistic was computed. Table 3 shows that VIF values for all items are below the threshold of 3.3 [37].

4.3 Common Method Bias

As with all self-reported data, there is a potential for common method bias (CMB) resulting from multiple sources such as consistency motif [38]. To check for CMB a common method factor was included in the PLS model whose indicators included all the principal constructs' indicators. Each indicator's variances that was substantively explained by the principal construct and by the method was calculated (see [39]). The CMB analysis only included the reflective constructs, because to the best of our knowledge there still is no agreed upon method for testing CMB for formative constructs. The results demonstrate that the average substantively explained variance of the indicators is 0.7, while the average method-based variance is 0.007. The ratio of substantive variance to method variance is about 100:1. In addition, none of the method factor loadings are significant. Given the small magnitude and insignificance of method variance, it is unlikely that CMB poses a serious concern for this study. Since all items were randomly shuffled within the survey it can be assumed that CMB does also not pose a problem for the formative FA items.

4.4 Structural Model

Figure 3 provides the R2 and path coefficients along with their respective significance levels from PLS analysis. The link between IQ and IT use is significant thus offering evidence for H1. The path between SQ and IT use is not significant. Thus, there is no support for H2. As for H3 and H4, Figure 3 shows a significant link from FA to IQ but a non-significant link to SQ. We found support for H3, but H4 had to be refused. We could find a significant correlation between COM and FA and between COV and FA which supports H5 and H6. The effect of COM on IQ as well as on SQ is positive

thus supporting H7 and H8. H9 and H10 propose that COV impacts IQ and SQ. Figure 3 shows significant paths from COV to IQ and SQ. We thus found support for H9 and H10. With regard to the four control variables included in the model, none of them were significantly related to IT use. Thus, we conclude that the variance in IT use does not depend on the control variables. The results indicate that the model explained 20% of the variance in IT use. COM, COV and FA account for 34% of the variance in IQ. COM and COV explained 42% of SQ. 30% of the variance of FA is explained by COM and COV. The model's capability to predict was tested by computing the cross-validated redundancy measures for each construct (Stone-Geisser's Q2). The blindfolding procedure is only applied to endogenous latent variables that have a reflective measurement model operationalization. All measures were larger than zero indicating that the latent construct exhibit predictive relevance [31].

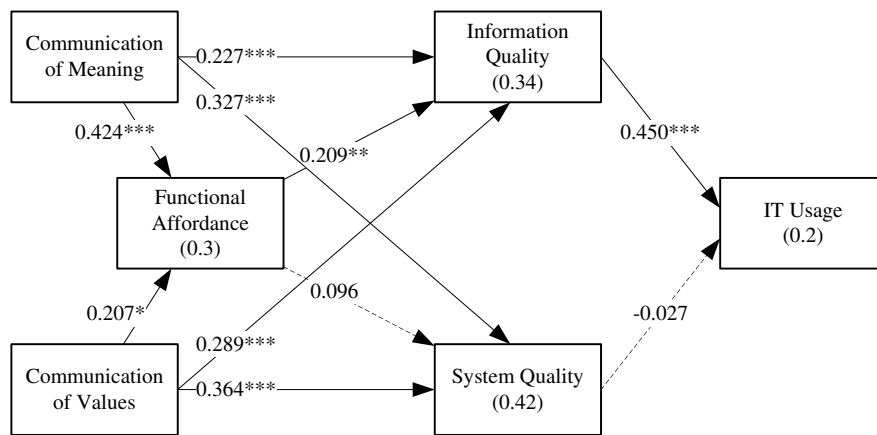


Fig. 3. Summary of Model Results (R^2 are reported in parentheses, Path Significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$)

5 Discussion and Conclusion

In order to understand IT-related factors that contribute to the formation of object-based beliefs, this study explored the effects of structural concepts (as proposed by AST) on information and system quality. The findings of this study offer statistical support for the relation between structural concepts and object-based beliefs. First, we divided the construct of symbolic expression into communication of values and communication of meaning and provided empirical evidence for the two-dimensional structure of symbolic expression. Symbolic expression consists of different types of values and meanings that are conveyed by an IT system. The factorial analysis of communication of values and communication of meaning indicates that both sub-constructs are distinct dimensions of symbolic expression. Communication of values have a positive impact on functional affordance which shows that values conveyed by an IT system positively affect how functionalities are perceived. In the

case of the present study, the SIS supports values such as efficiency, productivity and control which ultimately impact the awareness of the functionalities offered by the SIS. Communication of meaning positively affects functional affordance and provides some evidence that the understanding of the underlying functionalities contributes to the awareness of the functionalities afforded by the SIS. Users perceive a high system quality if they basically understand the meaning of the functionalities that and if the values conveyed by the system support its general intent. The same reasoning applies to the link between communication of values/communication of meaning and information quality. The better users understand the underlying functionalities and the better the IT system communicates values that support its general intent, the higher is the information quality of the IT system regarded by the users. Interestingly, our study provides some evidence that functional affordance directly influences the perceived information quality but not the overall system quality of the SIS. It seems that functionalities provided by the SIS are primarily important when it comes to the assessment of information that an IT system offers. In our case students used the possibility to exchange opinions with other students and used the forum to discuss relevant problems and questions. This interaction leads to new information that are perceived to be of high quality. Overall system quality depends only on a clear understanding of the system and positive values that are provided by the system. There is no significant relationship between system quality and IT use, as concluded in previous research [15]. From a theoretical perspective, this link was expected to be a strong one, that is, a high perceived service quality should lead to a higher level of IT use. In order to further investigate this surprising result we tested if the effect of system quality on IT use is mediated by information quality. To test the mediation, information quality was removed from the model and an additional confirmatory analysis was computed. The link between system quality and IT use became significant ($\beta = 0.276$, $p < 0.01$) which indicates that the effect of system quality on IT use is fully mediated by information quality. The reason for the inconsistency of the direct link between system quality and IT use could be the target IT system in this study. Because the system is mandatory and students do not have any alternative but to use the SIS the overall quality might not be important as long as the IT system provides adequate information. The proposed research model explained 34% of the variance in information quality and 42% of the variance in system quality which can be regarded as a moderate or weak R2 [31]. However, it depends on the specific research discipline and topic if values are considered to be high or low. Given the fact that object-based beliefs are only partially affected by IT-related structural concepts, the R2 values appear to be quite acceptable. Other sources for structure that were not part of this study, such as social norms, organizational resources, tasks or user characteristics, also play an important part in the process of belief formation and could explain additional variance in system quality and information quality [7-8].

The research model offers several advantages to researchers and practitioners interested in the social dynamics of human-IT interaction. The conceptualization based on AST encourages researchers as well as practitioners to investigate the relation between human agents and the IT system in more detail since the functional affordance construct has to be adapted to every new IT system and the user groups in

focus. Researchers have to realize which kinds of functionalities are provided to a certain user group to support a specific task in order to develop items that grasp the functional affordance of an IT system. The same applies to communication of values and meaning. Researchers have to be aware of what kind of values an IT system is supposed to provide; e.g. an ERP system should convey different values than collaborative tools. Thus, this study comes together nicely with Jones and Karsten's [10] call for more attention on the interaction between technology and human action although this quantitative research approach only provides a static snapshot of the human-IT interaction. Practitioners can benefit as well from our research. We demonstrated that not all functionalities of the SIS are equally important. The model therefore helps practitioners to evaluate what functionalities are essential and contribute the most to object-based beliefs. In addition, our research model underlined the importance of values and meaning that are communicated by an IT system. Thus, the design and operation of an IT system should directly target and support its general goals; e.g. an online shop should provide values such as control, convenience or trust in order to attract and retain customers [21], [24]. An obvious limitation of this study pertains to the sample and the IT system that was investigated. Focusing only on the students as the target users of the SIS allowed to control for extraneous factors such as different use intentions and objectives, different user types and so forth. Future research should examine the model across different populations and different IT systems, especially where IT use is completely voluntary. The proposed model explains some variance in information and system quality and therefore seems to be applicable to situations where users are more or less forced to use a certain system. In a mandatory setting, the provision of "right" structures (functionalities, values and meaning) to support a task might even be more important than in a voluntary setting since users do not have a choice to switch to another IT system. Despite the aforementioned limitations, our research adds to the existing knowledge on technology adoption. The structural potential of IT systems cannot fully determine the formation of object-based beliefs since IT systems are embedded in an organizational environment that provides different structures such as norms and values. However, Giddens acknowledges the value of decomposing structuration by taking institutions as a backdrop and by focusing on the structural potential of technical objects that shape and generate social structures [40]. This bracketing artificially segments structuration, but it is still admissible and justified for methodological purposes [12].

Acknowledgments

This research is supported by grants from the German Research Foundation (HO 2196/4-1).

References

1. DeLone, W.H., McLean, E.R.: Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research* 3 (1), 60-95 (1992)
2. Venkatesh, V., et al.: User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27 (3), 425-478 (2003)
3. Sabherwal, R., Jeyaraj, A., Chowa, C.: Information System Success: Individual and Organizational Determinants. *Management Science* 52 (12), 1849-1864 (2006)
4. Wu, J., Lederer, A.: A Meta-Analysis of the Role of Environment-Based Voluntariness in IT Acceptance. *MIS Quarterly* 33(2), 419-432 (2009)
5. Wixom, B.H., Todd, P.A.: A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research* 16 (1), 85-102 (2005)
6. Benbasat, I., Barki, H.: Quo vadis TAM? *Journal of the Association for Information Systems* 8 (4), 211-218 (2007)
7. Al-Natour, S., Benbasat, I.: The Adoption and Use of IT Artifacts: A New Interaction-Centric Model for the Study of User-Artifact Relationships. *Journal of the Association for Information Systems* 10 (9), 661-685 (2009)
8. DeSanctis, G., Poole, M.S.: Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory. *Organization Science* 5 (2), 121-147 (1994)
9. Markus, M.L., Silver, M.S.: A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit. *Journal of the Association for Information Systems* 9 (10/11), p. 609-632 (2008)
10. Jones, M.R., Karsten, H.: Giddens's Structuration Theory and Information Systems Research. *MIS Quarterly* 32 (1), 127-157 (2008)
11. Griffin, E.: *A First Look at Communication Theory*. McGraw-Hill (2003)
12. Poole, M.S., DeSanctis, G.: Structuration Theory in Information Systems Research: Methods. In: Whitman, M.E., Woszczynski, A.B. (eds) *The Handbook of Information Systems Research*. IGI Global: Hershey, PA, USA (2003)
13. Fishbein, M., Ajzen, I.: *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley (1975)
14. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems* 19 (4), 9-30 (2003)
15. Petter, S., McLean, E.R.: A meta-analytical assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level. *Information and Management* 46 (3), 159-166 (2009)
16. DeLone, W.H., McLean, E.R.: Information Systems Success Revisited. In *Proceedings of the 35th Hawaii International Conference on System Sciences* (2002)
17. Gibson, J.J.: The theory of affordances. In: Shaw, R.E. (ed.): *Perceiving, Acting, and Knowing*, J.B. Lawrence Erlbaum Associates, Hillsdale, NJ (1977)
18. Chemero, A.: An Outline of a Theory of Affordances. *Ecological Psychology* 15 (2), 181-195 (2003)
19. Hartson, H.R.: Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology* 22 (5), 315-338 (2003)
20. De Souza, C.S., Preece, J.: A Framework for Analyzing and Understanding Online Communities. *Interacting with Computers: The Interdisciplinary Journal of Human-Computer Interaction* 16 (3), 579-610 (2004)

21. Grange, C., Benbasat, I.: Online Shopping: The Functions and Symbols of Design Artifacts. In: Proceedings of the 43rd Hawaii International Conference on System Science (2010)
22. Margolis, E., Laurence, S.: Concepts. Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/concepts/> (2006)
23. Bühler, K.: Theory of Language: The Representational Function of Language. J. Benjamins Pub. Co, Amsterdam et al. (1990)
24. Hu, X., et al.: The effects of Web assurance seals on consumers' initial trust in an online vendor: A functional perspective. *Decision Support Systems* 48 (2), 407-418 (2010)
25. Hui, K.-L., Teo, H.H., Lee, S.-Y.T.: The Value of Privacy Assurance: An Exploratory Field Experiment. *MIS Quarterly* 31 (1), 19-33 (2007)
26. Soliman, M.A., Beaudry, A.: Understanding Individual Adoption and Use of Social Computing: A User-System Fit Model and Empirical Study. In: International Conference on Information Systems (2010)
27. Burton-Jones, A., Straub, D.W.: Reconceptualizing System Usage: An Approach and Empirical Test. *Information Systems Research* 17 (3), 228-246 (2006)
28. Rai, A., Lang, S., Welker, R.: Assessing the validity of IS success models: an empirical test and theoretical analysis. *Information Systems Research* 13 (1), 5-69 (2006)
29. Jarvis, C.B., Mackenzie, S.B., Podsakoff, P.M.: A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research* 30 (2), 199-218 (2003)
30. Petter, S., Straub, D., Rai, A.: Specifying Formative Constructs in Information Systems Research. *MIS Quarterly* 31 (4), 623-656 (2007)
31. Hair, J.F., Ringle, C.M., Sarstedt, M.: PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice* 19 (2), 139-151 (2011)
32. Ringle, C.M., Wende, S., Will, A.: SmartPLS 2.0 (Beta). Hamburg (2005)
33. Jöreskog, K.G., et al.: Lisrel 8: New Statistical Features. Scientific Software International, Lincolnwood, Illinois (2001)
34. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18 (1), 39-50 (1981)
35. Straub, D., Boudreau, M.-C., Gefen, D.: Validation Guidelines for IS Positivist Research. *Communications of the AIS* 13, 380-427 (2004)
36. Cenfetelli, R.T., Bassellier, G.: Interpretation of Formative Measurement in Information Systems Research. *MIS Quarterly* 33 (4), 689-707 (2009)
37. Diamantopoulos, A., Siguaw, J.A.: Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *British Journal of Management* 17 (4), 263-282 (2006)
38. Podsakoff, P., et al.: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology* 88 (5), 879-903 (2003)
39. Liang, H., et al.: Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. *MIS Quarterly* 31 (1), 59-87 (2007)
40. Giddens, A.: The Constitution of Society: Outline of Theory of Structuration. University of California Press, Berkeley, CA, USA (1984)

Teaching the Chief Information Officers: An Assessment of the Interrelations within their Skill Set

Matthias Boehm, Carl Stolze, and Oliver Thomas

University of Osnabrück, Information Management and Information Systems,
Osnabrück, Germany
{matthias.boehm, carl.stolze, oliver.thomas}@uni-osnabrueck.de

Abstract. Due to the high volatility in the field of information technology (IT) and the rapid technological advancements, all IT professionals constantly have to be able to evaluate trends and put them into context. This is especially true for those who fulfill the role of a company's chief information officer (CIO). But if there is a gap between the required set of skills and those needed, training becomes necessary. In order to plan training programs, one has to know the skill set of current CIOs. We investigate this by conducting workshops with 21 CIOs from a diverse set of companies. The purpose of this paper is to better understand the skill items of CIOs and how they interrelate.

Keywords: CIO, training, skill gap, interrelation analysis

1 Motivation

In times of economic uncertainty, constant change has become the norm: volatility, multiplicity, versatility, and mobility are the 2012 mega trends for information technology (IT) professionals [1]. In order to cope with these trends, IT staff has to keep track of technological developments and decide if and how to adopt them [2]. The rapid technological advancements multiply the complexity and increase the need for personnel with appropriate knowledge and skills [3]. But a one-size-fits-all approach for training IT personnel is bound to fail as different skills are required during an individual's career [3]. The relative importance of technical, personal, administrative and conceptual skills changes over time [4]. However, most IT training programs are still targeted at starters and pure technical skills whilst neglecting skills required for leadership positions with their lifelong learning requirement [5].

As IT is now part of the very fabric of modern life and business, the management of IT also gains importance [6-7]. Management of IT means the alignment of IT and business requirements and the linkage of the organizational strategy to the IT strategy [8-9]. IT management is usually the task of the information systems (IS) function within a company, headed by the Chief Information Officer (CIO) [10]. While in the U.S. the name CIO is widely accepted for this strategic position, in Europe often the title IT director is used to describe a more operational role. In this paper we follow the

understanding of Joia and define the CIO as the IT professional that is leading the organizational unit responsible for IT within an organization [10].

It is no wonder that CIOs need to have a broad set of skills to excel in their job [2]. In literature, broad skill categories for CIOs, like knowledge and managerial skills [11], technical and non-technical skills [12], project management skills [13], IT capabilities and well-honed communication skills [14], are named. In most cases, they are based on theoretical models [9], [15]. However, these models describe required skills only on a very high abstraction level.

To plan an actual training curriculum, these broad categories do not provide the required level of detail nor do they necessarily reflect the skill levels of current CIOs. Therefore, a concretization and derivation of interrelations between current skills is necessary for planning comprehensive training programs [3]. Hence, we derived the following research question:

RQ: What are skill items current CIOs possess and how are they interrelated?

In order to answer our research question, we proceed as follows: Based on a comprehensive literature review [16], we conduct two workshops. Due to the exploratory nature of our research this procedure has been chosen instead of conducting a Delphi study or large-scale survey. After presenting the key results, implications are discussed and a conclusion and outlook are given.

2 Background

Usually, the term skill is used interchangeably with the concept of competence [7]. However, there is a subtle difference: While skills are defined as the ability to apply knowledge and use know-how to complete tasks and solve problems, competencies are the proven ability to use knowledge, skills and personal, social, and/or methodological abilities, in work or study situations and in professional and personal development [17]. In the following we use the term skill to stress that the described abilities have not yet been proven.

In literature, skills are usually divided into technical (hard) and personal (soft) skills [18]. Hard skills, like programming, become less and less important, and companies increasingly emphasize soft skills, like leadership skills, when hiring new IT personnel [19-20].

A mixture of different kinds of skills is necessary for CIOs [14], [18]. Back in 1983, Zmud described general skill categories IT professionals should fulfill: Next to organizational (unit) knowledge and skills, which generally are seen as soft, "hard" IS knowledge and technical skills are required [21]. Throughout the years, those skill lists have been extended and adapted to the specific role of the CIO. Later research also stressed the economic importance of the CIO [11]. Joia went further by integrating six models of CIO skills, like Earl [9] or Broadbent and Kitziis [15], into one. Additionally, a questionnaire had been sent by Joia to (Brazilian) CIOs for verification. The resulting skill categories are knowledge of the business, understanding the organ-

izational context, ability to influence the organization, technical expertise, external networking, management of the IT operation, and capacity to innovate using new IT [10]. Other authors like Ekimci and Ozkan [22], Luftman and Ben-Zvi [23], or Peppard [7] use similar skill categories for their research. What most approaches have in common is that they use general skill listings without analyzing skills in concrete technologies or methods.

Since 1991, CIGREF has been developing a nomenclature of IS roles which provides a description of the roles existing in the IT Departments. It consists of 36 competences which are associated to 33 roles of IT professionals [24]. One of these roles is the Chief Information Officer (CIO) (cf. Table 1). This to-be profile consists of 10 general skill categories (and associated levels of proficiency and responsibility) a CIO should fulfill. Nevertheless, one has to keep in mind that these items are not the skills which CIOs actually possess.

Table 1. To-be profile of the CIO role [24]

Areas of concern	General skill category	Level
Plan	IS and Business Strategy Alignment	5
	Service Level Management	4
	Business Plan Development	5
	Sustainable Development	4
Enable	Personnel Development	4
Manage	Risk Management	3
	Relationship Management	4
	Business Change Management	5
	Information Security Management	3
	IT Governance	5

One of the key challenges in the IT environment is the transformation of technically skilled entry-level hires into midlevel IT managers with strong business management skills [25]. In order to achieve this, comprehensive training programs [26] are necessary that implement the requirements of CIGREF. Boehm et al. developed such an approach for teaching CIOs. Their framework (cf. Figure 1) is based on an iterative assessment of 116 existing training programs together with expert interviews of researchers and practitioners. The authors identified core modules which can be grouped into methodology courses and personal skills modules [27].

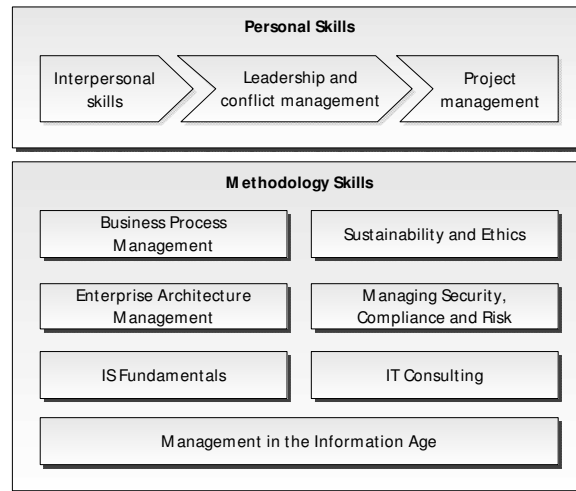


Fig. 1. Framework for Teaching CIOs [27]

The first methodology module, Management in the Information Age (MIA), integrates aspects like internet economics, management basics as well as consulting methods and approaches. In the field of IS Fundamentals (ISF), the basics of IS/IT, enterprise software, business intelligence, and operations research are covered. Enterprise Architecture Management (EAM) discusses all aspects of Business-IT alignment and IT service management. Process-oriented approaches and methods belong to the field of Business Process Management (BPM). The IT Consulting (ITC) module focuses on knowledge and methods related to business and IT advice. Within the field of Managing Security, Compliance and Risk (MSCR) subjects like IT security, legislation, contracts, and risk management are discussed. Sustainability and Ethics (SE) deal with green IT, management ethics, and social aspects of information management as well as intercultural studies. Finally, the personal skill modules cover aspects on Interpersonal skills, Leadership and conflict management as well as Project management. The modules in the framework can be seen as concrete skills for CIOs and, therefore, these skill items will be used in further analysis of the CIOs' skills [3].

3 Methodology

For answering our initially laid out research question, we conduct a multi-perspective analysis. For doing so, several research methods could have been employed. With the help of a large-scale survey among CIOs, one could identify nearly all skills of CIOs. A Delphi study could be used to identify research and practice agendas for training certain skills. We decided to employ neither of these two methods. There are several reasons for this decision: Next to time and resource related reasons, the main reason has been the fact that we wanted to closely get in touch with the CIOs. We are convinced that thorough discussions are more helpful in answering our research questions

than large-scale studies [28]. Hence, we considered the workshop method as ideal for our investigation. Although the analyzed phenomenon is hardly nascent, we chose this method because it allowed us to discuss the answers of the CIOs, to give context to answers, and to iterate on the given answers several times.

The data collection took place in two workshops – one in June 2011 and one at the end of 2011. Next to comprehensive group discussions, qualitative and quantitative instruments were employed.

The self-reporting approach of Vygotsky is used for methodological foundation. This is a process of looking at the self in order to consider aspects that one is asked to evaluate. Participants are asked to respond to a set of survey questions which rate their own knowledge, expertise, or skills [29]. This instrument is seen as a valuable depiction of the reality, although there are the limitations like overestimating or underestimating the level of expertise [30]. Nevertheless, it is claimed that the self-reporting approach, for example in form of think-aloud protocols of experts, can be seen as a good instrument for the study of skills [31].

Within the first workshop, the skill lists found in literature have been worked on with eight experts from practice. Besides three CIOs, we invited two executive level managers, two department managers and one executive coach to gain a 360 degree view on the topic. Within the workshop advantages and disadvantages of various approaches have been discussed by the two women and six men. Furthermore, the participants have been asked to further elaborate on required skills.

The second workshop was attended by twenty-one CIOs from different industries. Except for one, all have been male. The aim of this meeting was the self-assessment of skills. For doing so, we developed a questionnaire which is based on our literature analysis and the first workshop. The participants have been asked to fill out the first questionnaire. In this questionnaire, the CIOs assessed their skills on a 5-point Likert-type scale. Within the analysis, the aim is to investigate relationships between skills. This can help to design more integrated programs addressing interconnected skills better. As the variables are ordinal ranked, a specific measure for correlation to identify interconnections has to be selected. We chose Kendall's τ_B because it is as powerful as the traditional Pearson coefficient but is explicitly tailored to ordinal variables and more useful for smaller sample sizes [32]. Kendall's τ correlation coefficients do not reflect to what extent two variables move together. Rather, they assess the difference between the probabilities that the observed data are in the same order versus the probability that the observed data are not in the same order. In other words, it assesses the association between the rankings applied to the different variables (the similarity of data orderings).

3.1 Psychometric Properties

In order to show how well our instrument measures the construct of interest, we calculated its psychometric properties. In general, there are two broad types of psychometric properties: reliability and validity. In order to ensure the appropriateness of our research instrument, we tested it for reliability and content validity.

Reliability refers to the internal consistency of a measurement instrument. It means that a measure (in our case the questionnaire) consistently reflect the construct that it is measuring. Therefore, it is the test's ability to measure the construct of interest consistently. Mostly, the Cronbach's α is used here, whereas it should be higher than .8 in order to be accepted [33]. In our case, Cronbach's α equals .841, which can be seen as high reliability.

Validity in general refers to whether an instrument measures what it was designed to measure. It is the question of how well the test accurately measures the construct of interest. In order to accept the validity of the observed measurement, there are major estimation methods: construct, content, concurrent and predictive validity. Construct validity involves understanding the theoretical rationale underlying the obtained measurements and hence relates the construct of interest to other constructs in order to develop an entire theoretical framework for the phenomenon being measured [34]. Concurrent validity involves correlating two different measurements of the same phenomenon. Predictive validity refers to the ability of a measured phenomenon at one point in time to predict another phenomenon at a future point. Content validity is the evidence that the content of a test corresponds to the content of the construct it was designed to cover. As this research concentrates on one concept and uses one concept at one specific point in time, only content validity can be examined in the following.

For this purpose, experts can be consulted. In order to ensure content validity, a comprehensive literature review on skills of CIOs was conducted. Within a pre-test, the questionnaire was filled out by experts (professors and IS professionals). They reviewed it and gave helpful comments on how to change the questionnaire. Therefore, both content and clarity could be improved. Finally, a sample of respondents (separate from those included in the pre-test) was asked to check the questionnaire. These and all pre-test respondents were excluded from the sample used for data analysis.

4 Results

The participants of the first workshop confirmed that the chosen framework by Boehm et al. [27] is valid. They concluded that the modules of the framework can be seen as general skill categories. Next to these ten skills, the following nine important skill categories have been added to the list: IT Governance (ITG), Cloud Computing (CC), Outsourcing (Out), Virtualization (Vir), (Process) Modeling (Mod), Customer and Employee Satisfaction (Sat), Coaching (Coa), Enterprise Resource Planning (ERP), and IT Training (ITT). These skills are also seen as highly relevant for CIOs in literature (cf. for example [20], [22], [35–38]).

One can remark that the skills of the framework by Boehm et al. [27] and those of the participants exist at different levels of analysis. For example, ERP (one category identified by the participants) could fit within the IS Fundamentals category identified by Boehm et al. [27]. Therefore, participants of the workshop elaborated together with the authors an integrated list of skills (cf. Table 2). In this list the level of analysis has been defined for each skill. For example, while the skill Enterprise Resource Planning

is seen on a specific and practical level, IS Fundamentals is seen on a more general and theoretical level. Altogether, skills can be defined on a general, theoretical, specific and/or practical level. With this list we ensure the comparability of skills.

Table 2. Integrated List of Skills

Skill	Source	Level of Analysis
Management in the Information Age	Boehm et al. [27]	General and theoretical
IS Fundamentals	Boehm et al. [27]	General and theoretical
IT Consulting	Boehm et al. [27]	Theoretical and practical
Enterprise Architecture Management	Boehm et al. [27]	General and theoretical
Business Process Management	Boehm et al. [27]	Theoretical
Managing Security, Compliance and Risk	Boehm et al. [27]	Theoretical and practical
Sustainability and Ethics	Boehm et al. [27]	General and theoretical
IT Governance	1 st Workshop	Specific and practical
Cloud Computing	1 st Workshop	Practical
Outsourcing	1 st Workshop	Practical
Virtualization	1 st Workshop	Specific and practical
Modeling	1 st Workshop	Specific and practical
Customer and Employee Satisfaction	1 st Workshop	Practical
Enterprise Resource Planning	1 st Workshop	Specific and practical
Coaching	1 st Workshop	Theoretical and practical
IT Training	1 st Workshop	Theoretical and practical
Interpersonal Skills	Boehm et al. [27]	Theoretical and practical
Leadership Management	Boehm et al. [27]	Theoretical and practical
Project Management	Boehm et al. [27]	Practical

From the 21 CIOs of the second workshop, 17 questionnaires have been completely filled out and therefore could be used for the analysis. All respondents are male and their average age is 39.44, with the youngest participant being 30 years old and the oldest 56. Although most participants have different job titles, their position can be basically subsumed by the term CIO. Nearly 53% of the participants' enterprises have less than 250 employees. These firms can be seen as small or medium-sized enterprises (SMEs). However, also a relatively large amount of participants (35%) comes from companies with more than 1,000 employees. The business sectors have been compiled using the North American Industry Classification System (NAICS). According to this system, half of the CIOs come from companies offering professional, scientific, and technical services (50%). In 32% of the cases, CIOs work at manufacturing companies, followed by enterprises working in the field of agriculture, forestry, fishing, and hunting (12%). One CIO works at a real estate company (6%). Altogether, one can say that the participants of the workshops reflect to a good extent the global market in terms of size of enterprises and business sectors.

The results of the CIO self-assessment of the 19 final skills which took place in the second workshop are depicted in Figure 2. The skills have been coded from -2 (= lowest rank) to 2 (= highest rank). For each skill its abbreviation (Abb.), average value (Avg.), standard deviation (SD) and rank is shown in the left part of Figure 2. The right part shows a graphical representation of the average values.

Skill	Abb.	Avg.	SD	Rank
Management in the Information Age	MIA	,65	,996	5
IS Fundamentals	ISF	,63	1,204	6
IT Consulting	ITC	,59	1,121	8
Enterprise Architecture Management	EAM	,41	,939	10
Business Process Management	BPM	,59	1,278	9
Managing Security, Compliance and Risk	MSCR	-,06	1,197	15
Sustainability and Ethics	SE	-,41	1,543	17
IT Governance	ITG	-,18	1,185	16
Cloud Computing	CC	,12	1,269	11
Outsourcing	Out	,06	1,345	14
Virtualization	Vir	,76	1,251	2
Modeling	Mod	,12	1,269	12
Customer and Employee Satisfaction	Sat	,06	1,569	13
Enterprise Resource Planning	ERP	,63	1,360	7
Coaching	Coa	-,56	1,459	19
IT Training	ITT	-,44	1,365	18
Interpersonal Skills	Int	,88	1,054	1
Leadership Management	Lea	,76	1,251	3
Project Management	Proj	,76	,903	4

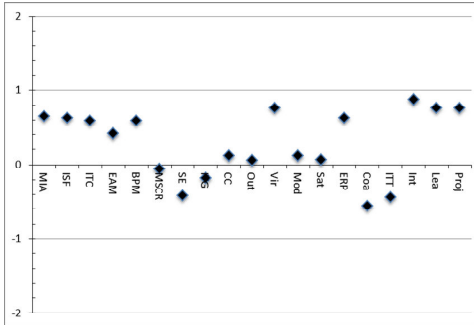


Fig. 2. Self-assessed Skills of the CIOs

The highest ranked skills are Interpersonal Skills (Int), Leadership Management (Lea) and Project Management (Proj) as well as the methodological and technical skills Virtualization (Vir) and Management in the Information Age (MIA). Interestingly, all personal skill modules (Int, Lea, and Proj) are among the top 4 of highest ranked skills. CIOs seem to be very confident when it comes to the assessment of their personal skills. The skills in IS Fundamentals (ISF) and Enterprise Resource Planning (ERP) are also ranked relatively high. One has to note that the standard deviation for these skills is quite low and therefore the responses have been quite similar in these cases. Within the top 10 of the highest ranked skills all modules from the framework for teaching IT management and IT consulting can be found, except Managing Security, Compliance and Risk (MSCR) and Sustainability and Ethics (SE). Interestingly, those two modules are in the top 5 list of the lowest ranked skills. Next to MSCR and SE, Coaching (Coa), IT Training (ITT), and IT Governance (ITG) are also ranked quite low. Here, CIOs need further training. Also Outsourcing (Out), Customer and Employee Satisfaction (Sat), Modeling (Mod), and Cloud Computing (CC) are ranked low. In opposite to the highest ranked skills, the standard deviations are high. Therefore, the responses have been quite different in these cases. Nevertheless, the necessity of training can be identified in these cases.

As stated before, interrelations between the skill items shall be investigated. In order to do so, correlations have been calculated (cf. Figure 3). The figure shows the correlation coefficients. The higher the value the stronger is the relationship. As rules of thumb one can say that a coefficient of .30 and above is a strong relationship. If the value is less than .10, the relationship is weak. If the algebraic sign is negative, the interpretation is that the higher a value of one variable the lower is the value of another variable. A two-tailed significance of each correlation is also calculated meaning that the hypothesis is tested for non-correlation between two variables in the population. In addition to the correlations of the 19 skill variables, also the correlation be-

tween the age of the participants and the skill items has been calculated. Using the gender or industry sector was not successful because of the amount of data. As one can see from Figure 3, the age of the participants is positively correlated between ERP, BPM, ITG, Mod, and Sat. The correlations are in each case significant at .05 levels (depicted by * in Figure 3). Therefore, one can generally say that the older the participants are, the better they evaluate their skills in these five areas. Interestingly, no personal skill item is correlated with age. Only the Proj skill item has a relatively strong relationship with the Age variable. Nevertheless, this is not significant.

		Age	MA	ISF	ITC	EAM	BPM	MSCR	SE	ITG	CC	Out	Vir	Mod	Sat	Coa	ERP	ITT	Int	Lea	Proj	
Age	Correlation Coefficient	1,000																				
	Sig. (2-tailed)	.																				
MA	Correlation Coefficient	,059	1,000																			
	Sig. (2-tailed)	,781																				
ISF	Correlation Coefficient	-,299	,067	1,000																		
	Sig. (2-tailed)	,166	,777	.																		
ITC	Correlation Coefficient	,396	,172	-,039	1,000																	
	Sig. (2-tailed)	,059	,452	,866																		
EAM	Correlation Coefficient	,025	-,345	-,396	-,132	1,000																
	Sig. (2-tailed)	,910	,152	,105	,577	.																
BPM	Correlation Coefficient	,505*	-,131	-,330	,221	,158	1,000															
	Sig. (2-tailed)	,014	,563	,148	,320	,498	.															
MSCR	Correlation Coefficient	-,067	,393	-,128	-,174	,394	,167	1,000														
	Sig. (2-tailed)	,756	,090	,584	,446	,101	,457	.														
SE	Correlation Coefficient	,240	,114	,126	,302	-,114	,300	,139	1,000													
	Sig. (2-tailed)	,238	,608	,576	,168	,621	,163	,530	.													
ITG	Correlation Coefficient	,424*	,376	,038	,504*	,000	,242	,259	,508*	1,000												
	Sig. (2-tailed)	,045	,105	,873	,027	1,000	,281	,263	,022	.												
CC	Correlation Coefficient	-,122	,301	,133	,059	,268	-,148	,283	,203	,291	1,000											
	Sig. (2-tailed)	,557	,185	,565	,792	,256	,502	,212	,350	,200	.											
Out	Correlation Coefficient	-,030	,481*	,140	-,150	-,117	-,278	,121	,248	,272	,419	1,000										
	Sig. (2-tailed)	,885	,032	,537	,499	,614	,201	,589	,249	,224	,057	.										
Vir	Correlation Coefficient	,166	,417	-,178	,164	,331	,230	,739**	,412	,420	,500*	,466*	1,000									
	Sig. (2-tailed)	,427	,065	,437	,462	,157	,293	,001	,056	,062	,024	,033	.									
Mod	Correlation Coefficient	,457*	,012	-,229	,378	,161	,635**	,136	,356	,537*	-,183	-,073	,202	1,000								
	Sig. (2-tailed)	,028	,958	,320	,092	,496	,004	,549	,102	,018	,411	,739	,360	.								
Sat	Correlation Coefficient	,437*	-,100	-,360	,317	,148	,539*	,108	,620**	,281	,056	,033	,310	,473*	1,000							
	Sig. (2-tailed)	,038	,662	,107	,158	,531	,015	,634	,004	,216	,800	,881	,161	,034	.							
Coa	Correlation Coefficient	,277	,407	-,158	,062	,028	,124	,301	,319	,263	,345	,371	,441*	,000	,436	1,000						
	Sig. (2-tailed)	,190	,076	,499	,783	,906	,578	,189	,148	,251	,124	,095	,048	1,000	,053	.						
ERP	Correlation Coefficient	,468*	,090	-,417	,491*	-,172	,250	-,243	,144	,290	-,244	,057	,000	,337	,185	,022	1,000					
	Sig. (2-tailed)	,028	,697	,078	,031	,475	,264	,291	,515	,208	,280	,799	1,000	,136	,415	,920	.					
ITT	Correlation Coefficient	,248	,450	-,269	,201	,272	,091	,413	,300	,471*	,593**	,625**	,686**	,081	,371	,589**	,157	1,000				
	Sig. (2-tailed)	,245	,052	,255	,377	,258	,685	,073	,176	,041	,009	,005	,002	,719	,103	,008	,481	.				
Int	Correlation Coefficient	,189	,000	,269	,254	-,237	,174	-,372	,425	,142	,000	,212	-,108	,231	,335	,114	,173	,130	1,000			
	Sig. (2-tailed)	,369	1,000	,249	,264	,321	,436	,106	,054	,538	1,000	,342	,631	,306	,138	,616	,451	,571	.			
Lea	Correlation Coefficient	,175	,286	,110	,433	,040	,042	,112	,412	,509*	,138	,301	,211	,404	,422	,414	,140	,349	,512*	1,000		
	Sig. (2-tailed)	,400	,205	,634	,052	,865	,848	,619	,056	,024	,531	,169	,337	,067	,058	,064	,535	,121	,022	.		
Proj	Correlation Coefficient	,367	,172	,098	,701**	-,302	,075	-,349	,193	,451	-,319	,211	-,088	,447	,066	-,094	,543*	,041	,453	,442	1,000	
	Sig. (2-tailed)	,084	,462	,677	,002	,212	,739	,133	,386	,052	,162	,348	,696	,050	,774	,683	,019	,861	,051	,051	.	

Fig. 3. Skill Correlation Matrix

For getting a better overview of the relationships between the remaining 19 skill items, a skill correlation network diagram has been generated (cf. Figure 4). This diagram focuses on the significant relationships and shows correlations that are either significant at the .05 level (depicted by *) or at the .01 level (depicted by **). The colors depict the methodology skill items (orange), the personal skills (green) and the additional named skills (blue). First of all, one recognizes that skills in the seven methodological variables from the framework by Boehm et al. [27] have no relation-

ship with each other. Therefore, these courses seem to cover a mutually exclusive and collectively exhaustive field.

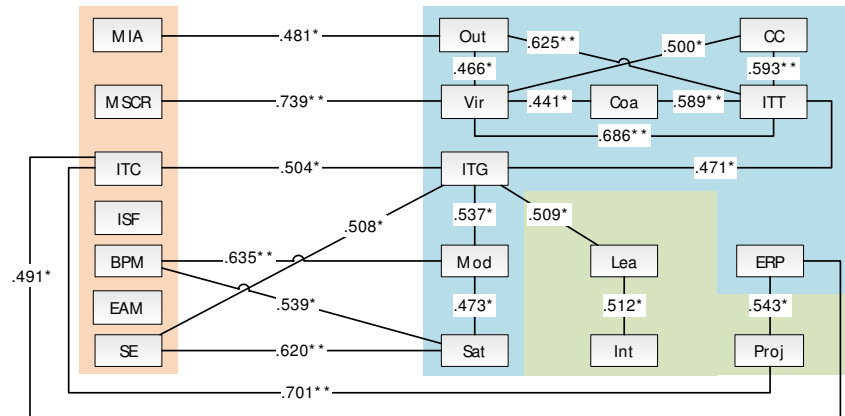


Fig. 4. Skill Correlation Network Diagram

Within the personal skills, the variables Lea and Int are positively correlated. The relationship leads to the supposition that CIOs with high interpersonal skills also have high leadership management skills. This can be seen as a causal relationship because one cannot set one of the skills apart from the other. The Proj variable is not correlated to the other personal skills. However, it has strong relationships with ERP and ITC. Also ERP and ITC are correlated. One might conclude that this is the classical causal relationship within an ERP implementation project and consulting activities within the company. By investigating the other methodological modules one recognizes that two modules, ISF and EAM, have no correlation with others. The reason for this should be investigated in future research. Although the relationship for example between EAM and ITG has been widely discussed in literature [39], the CIOs in our study have not yet acknowledged this link. This research-practitioner gap should be overcome in future. The MIA, BPM, MSCR, and SE module have strong relationships with the other variables. However, MIA is for example connected to outsourcing. MIA is currently only indirectly connected to skills like Vir and ITT. The MSCR module also has a strong relationship to only one other skill, namely Vir. Therefore, CIOs may not discern the links between these skills. A strong relation can be shown between BPM and Mod. As modeling skills are very often required in business process reengineering, this highly significant correlation is comprehensible. Interestingly, BPM is also correlated with Sat. CIOs may need skills especially for employee satisfaction in order to successfully implement BPM. Another strong relationship has been discovered between SE and ITG as well as Sat. Due to that fact that SE is a relatively new field this correlation has not been discovered so far.

The important skills identified in the first workshop are all highly related to each other and the modules of the framework by Boehm et al. [27]. A small sub-network is formed by Out, Vir, Coa, ITT and CC. Here, all skills have strong relationships with each other. Only Out and CC are not linked. This is very interesting because usually it

is said that cloud computing is a form of outsourcing [40]. Nevertheless, CIOs have not yet built up skills in these two areas at the same time. CC is currently only related to Vir and ITT. Though, the topic of CC is much more complex [40]. Vir and ITT are the most complex skills because they have strong relationships with several other skills: Vir is connected to MSCR, CC, Out, Coa, and ITT; ITT is linked to ITG, CC, Out, Vir, and Coa. In planning training programs for CIOs, planners have to keep this fact in mind and design programs which integrate these skills. Additionally, Out with its relation to MIA, Vir and ITT has to be covered as well. Another sub-network is formed by ITG, Mod, Sat and correlations to the respective other skills. ITG is also a very complex skill because it has strong relationships to ITC, SE, Mod, ITT, and Lea. From this list one sees the interdisciplinary approach of IT governance. Next to its relation to BPM, the skill Mod is related to Sat. This can also be explained by their correlations to the CPM skills.

5 Discussion

5.1 Answering the Research Question

The results offer some interesting insights into the working field of the CIO. CIOs have the highest skills in Interpersonal Skills, Leadership Management and Project Management as well as the methodological and technical skills Virtualization and Management in the Information Age. A backlog demand can be identified in the fields of Managing Security, Compliance and Risk, Sustainability and Ethics, Coaching, IT Training, and IT Governance. The analysis of interrelations between the skills also gives several insights. Research-practitioner gaps can be identified by the fact that CIOs have not yet recognized the connection between certain topics, for instance cloud computing and outsourcing. This gap should be overcome in future by establishing a better knowledge transfer between research and practice. Additionally, new skills in Sustainability and Ethics for example are required. As these skills are closely related to others, one has to keep this in mind when designing training programs for CIOs. It is important that CIOs build up skills in two or more areas like cloud computing and outsourcing at the same time. In general, one can learn from the correlation analysis that the CIOs' tasks are interdisciplinary and highly interrelated.

By comparing these results with the to-be profile of the CIO role (cf. Table 1), one recognizes that the skill categories of the role are much broader than the skills discussed in the workshops. Only IT governance and personnel development can be found in both. As our results show, the CIOs' skills in these areas are at least expandable. The average values of .12 (for ITG) and .06 (for Sat) should be higher in order to fulfill the required levels of the to-be profile. As the remaining skills of this study have been investigated on another level of analysis, it is difficult to compare them with those in the to-be profile.

5.2 Implications

From this study, several implications for theory and practice can be drawn. First, researchers should admit that practitioners work on another set of topics out of which some are even solved in scientific literature. It can be said that closer collaboration and discussion of researchers and practitioners are necessary. As Rosemann and Vessey noted, this research is a first step for their applicability checks [41]. Hence, the research-practice gap can be reduced using results of this paper.

Second, this research enhances the knowledge about interrelations of skills. With the help of the results, an in-depth investigation of this field can be started.

Third, the analysis is a know-how basis for teaching CIOs in future: Knowledge about which skills CIOs need further training and which skills are related to one another is highly valuable for doing so. The results can also be used for supervision and controlling of projects as well as training programs in practice. If for example an IT professional wants to attend a seminar on ERP systems, he might search for programs offering also project management and IT consulting skills.

Forth, this paper shows that practitioners increasingly need to be aware of technological developments and take part in training programs in order to stay up-to-date. This analysis can be a starting point for doing so.

5.3 Limitations

Skills of CIOs depend on company's IT strategy, size, industry, culture and more. Therefore, different requirement profiles for CIOs should be specified. For example, a CIO of a small mechanical company needs to know more about production planning, while a CIO of a large public company needs to be expert for e-government. However, the to-be profile of the CIO role is a good basis for further investigations.

Of course, this empirical analysis offers some limitations. The sample size is enough to show valuable results but the research could still benefit from more data. A comparison of CIOs in different countries around the world might be one extension. Additionally, the self-assessment could be augmented by other methods for measuring skill levels. The results of this contribution can be used to conduct surveys among a greater number of CIOs. It should also be tested if the list of 19 skill items can be extended with skills derived from the trend analysis or other contributions in literature [23], [42-43]. However, as discussions within our workshops show, we captured all important skill items.

6 Conclusion and Future Research

In this paper, we investigated the skill set of the CIO. With the help of the findings, the working field of the CIO can be better understood and teaching programs can be designed. After conducting a literature review, two workshops have been carried out. Within the first workshop, important skills for CIOs have been identified. Together with the framework by Boehm et al. [27], these skills have been collected in a list which also depicts the respective level of analysis. These levels ensure the compara-

bility of the skills. In a second workshop, the skill list was shown to CIOs. A questionnaire was used to enable an assessment of the skills of today's CIOs. Several interrelations have been identified.

In future, it is important that researchers and practitioners develop a common frame of reference when talking about skills and trends of IT professionals. More research is necessary for addressing knowledge transfer between CIOs and researchers as well as among CIOs. Here, for example joint workshops of CIOs and researchers could be conducted. Nevertheless, in this paper lies the foundation for doing so.

Acknowledgements

This paper was written in the context of the research project *IMUCON* which is funded by the European Regional Development Fund (ERDF). The authors acknowledge the support by ERDF and all involved project partners as well as the anonymous reviewers.

References

1. Howard, C., Nelson, E., Rollings, M., Santos, J.: 2012 IT Professionals Planning Guide: Volatility, Multiplicity, Versatility, and Mobility. Gartner Research (2011)
2. Boehm, M., Stolze, C., Thomas, O.: Zwischen Information und Innovation: CIO-Weiterbildungskonzepte im Wandel. *IM Information Management und Consulting* 27, 48–56 (2012)
3. Doke, E.R., Williams, S.R.: Knowledge and Skill Requirements for Information Systems Professionals: An Exploratory Study. *Journal of Information Systems Education* 10 (1), 10–18 (1999)
4. Luftman, J.N.: *Managing the Information Technology Resource. Leadership in the information age.* Prentice Hall, Upper Saddle River, N.J. (2004)
5. Smid, G.: Consultants' Learning within Academia. Five devices for the design of university-based learning opportunities for management consultants. *Studies in Continuing Education* 23, 55–70 (2001)
6. Saldanha, T., Krishnan, M.: Leveraging IT for Business Innovation: Does the Role of the CIO Matter? In: *ICIS 2011 Proceedings*, Paper 17 (2011)
7. Peppard, J.: Unlocking the Performance of the Chief Information Officer (CIO). *California Management Review* 52 (4), 73–99 (2010)
8. Henderson, J.C., Venkatraman, N.: Strategic alignment: leveraging information technology for transforming organizations. *IBM Systems Journal* 38 (2.3), 472-484 (1999)
9. Earl, M.J.: The Chief Information Officer: Past, Present and Future. In: Earl, M.J. (ed.): *Information management. The organizational dimension.* Oxford University Press, Oxford (1996)
10. Joia, L.A.: Critical Competencies for the Brazilian CIO. In: *AMCIS 2010 Proceedings*, Paper 57 (2010)
11. Tagliavini, M., Moro, J., Ravarini, A., Guimaraes, T.: Shaping CIO's competencies and activities to improve company performance: an empirical study. In: *ECIS 2003 Proceedings*, pp. 1–15 (2003)

12. Gallagher, K.P., Kaiser, K.M., Simon, J.C., Beath, C.M., Goles, T.: The Requisite Variety of Skills for IT Professionals. *Communications of the ACM* 53 (6), 144–148 (2010)
13. Swartz, N.: Employees Not Receiving Critical Training, Study Says. *Information Management Journal* 39 (2), 7 (2005)
14. Bullen, C., Abraham, T., Galup, S.D.: IT Workforce Trends: Implications for Curriculum and Hiring. *Communications of the Association for Information Systems* 20, Article 34 (2007)
15. Broadbent, M., Kitzis, E.: *The new CIO leader. Setting the agenda and delivering results.* Harvard Business School Press, Boston (2005)
16. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly* 26, xiii–xxiii (2002)
17. European Parliament and European Council: The European Qualifications Framework (EQF) for Lifelong Learning, http://ec.europa.eu/education/lifelong-learning-policy/doc44_en.htm
18. Gorman, M.F.: A Case Study in Effectively Bridging the Business Skills Gap for the Information Technology Professional. *Journal of Education for Business* 86, 17–24 (2010)
19. Joseph, D., Ang, S., Chang, R.H.L., Slaughter, S.A.: Practical intelligence in IT. Assessing Soft Skills of IT Professionals. *Commun. ACM* 53, 149 (2010)
20. Groysberg, B., Kelly, L.K., MacDonald, B.: The New Path To the C-Suite. *Harvard business review* 89, 60–69 (2011)
21. Zmud, R.W.: *Information systems in organizations.* Scott, Foresman & Co., Glenview, Ill (1983)
22. Ekimci, N.A., Ozkan, S.: An Investigation of the Activities and Skill Sets Needed By Senior Information Technology (IT) Managers. In: *Proceedings of the European Conference on Information Management & Evaluation*, pp. 486–497 (2009)
23. Luftman, J.N., Ben-Zvi, T.: Key Issues for IT Executives 2011: Cautious Optimism in Uncertain Economic Times. *MIS Quarterly Executive* 10, 203–212 (2011)
24. CIGREF: Information Systems roles in large companies. HR nomenclature, http://www.cigref.fr/cigref_publications/RapportsContainer/Parus2011/2011_IS_roles_in_large_companies_HR_nomenclature_CIGREF_EN.pdf
25. Zwiag, P., Kaiser, K., Beath, C.M., Bullen, C., Gallagher, K.P., Goles, T., Wion, R.: The information technology workforce: Trends and implications 2005-2008. *MIS Quarterly Executive* 5, 101–108 (2006)
26. Stolze, C., Boehm, M., Zarvić, N., Thomas, O.: Towards Sustainable IT by Teaching Governance Practices for Inter-Organizational Dependencies. In: Nüttgens, M., Gadatsch, A., Kautz, K., Schirmer, I., Blinn, N. (eds.): *Proceedings of IFIP 8.6 Hamburg Conference*, pp. 70–88 (2011)
27. Boehm, M., Stolze, C., Breitschwerdt, R., Zarvić, N., Thomas, O.: An Integrated Approach for Teaching Professionals IT Management and IT Consulting. In: *AMCIS 2011 Proceedings*, Paper 72 (2011)
28. Bryman, A., Bell, E.: *Business research methods.* Oxford Univ. Press, Oxford (2007)
29. Vygotsky, L.S.: *Thought and language.* MIT Press, Cambridge, MA (1962)
30. Dey, S., Sedera, D.: Information Systems Experts: Definition, Identification and Application. In: *ECIS 2011 Proceedings*, Paper 195 (2011)
31. Ericsson, K.A., Charness, N.: Expert Performance. Its Structure and Acquisition. *American Psychologist* 49, 725–747 (1994)
32. Kruskal, W.H.: Ordinal Measures of Association. *Journal of the American Statistical Association* 53, 814–861 (1958)
33. Nunnally, J.C., Bernstein, I.H.: *Psychometric Theory.* McGraw-Hill, New York (1994)

34. Chin, W.W., Gopal, A., Salisbury, W.D.: Advancing the Theory of Adaptive Structuration: The Development of a Scale to Measure Faithfulness of Appropriation. *Information Systems Research* 8, 342–367 (1997)
35. Chan, C.-S.: The changing face of the CIO. *Enterprise Innovation* 7, 16–19 (2011)
36. Corbett, M.F.: Outsourcing and the new IT executive. A Trends Report. *Information Systems Management* 11, 19–22 (1994)
37. von Urff Kaufeld, N., Chari, V., Freeme, D.: Critical Success Factors for Effective IT Leadership. *Electronic Journal of Information Systems Evaluation* 12 (1), 119–128 (2009)
38. Werr, A.: Consultant supported ERP implementation - a learning opportunity? Stockholm School of Economics. SSE/EFI Working Paper Series in Business Administration No. 2005:3, Stockholm (2005)
39. Stolze, C., Zarvić, N., Thomas, O.: Working in an inter-organizational context: the relevance of IT Governance and Business-IT Alignment. *International Journal of Computer Science and Information Security (IJCSIS)* 9, 1–4 (2011)
40. Martens, B., Pöppelbuß, J., Teuteberg, F.: Understanding the Cloud Computing Ecosystem: Results from a Quantitative Content Analysis. In: Bernstein, A., Schwabe, G. (eds.): *Proceedings of the 10th International Conference on Wirtschaftsinformatik WI 2011, Zurich, Switzerland*. Volume 1, pp. 466–476 (2011)
41. Rosemann, M., Vessey, I.: Toward improving the relevance of information systems research to practice: the role of applicability checks. *MIS Quarterly* 32, 1-22 (2008)
42. Hopkins, B.: *The Top 10 Technology Trends EA Should Watch: 2012 To 2014*. Forrester Research, Inc., Cambridge, MA (2011)
43. Gartner Inc.: *Gartner Identifies the Top 10 Strategic Technologies for 2012*, <http://www.gartner.com/it/page.jsp?id=1826214>

Towards Understanding Social Software and Its Impact on Corporate E-Learning Motivation

Thomas Wirtky¹, Sven Laumer¹, Andreas Eckhardt², and Tim Weitzel¹

¹ Otto-Friedrich-University, Bamberg, Germany
{thomas.wirtky,sven.laumer,tim.weitzel}@uni-bamberg.de

² Johann-Wolfgang-Goethe-University, Frankfurt am Main, Germany
eckhardt@wiwi.uni-frankfurt.de

Abstract. This research combines recent discussions in the cross-disciplinary areas of e-learning, social software, and training motivation. On the one hand, e-learning with its role in a fast paced knowledge society, its potential to revolutionize education and the gap between research and practitioners is already in discussion for years. Discussions are becoming more intense with the rise of social software. On the other hand, human resources, life-long learning, and motivating employees to participate in trainings appear critical to organizations. This research focuses on e-learning in the corporate context and examines the impact of social software features on user motivation based on a review of training motivation literature and on 39 interviews conducted in an international IT services company. Findings suggest that the impact of social software features is still unknown, and that differentiating them by the learner's needs leads to further insights.

Keywords: E-learning, Social software, Training motivation, Human resource IS

1 Introduction

Trained employees are the most critical resource in a knowledge society [1] but knowledge becomes quickly outdated due to the increasing pace of changes. Life-long learning becomes more than a popular phrase. The need for flexible ways of training is eminent and no longer met by constantly updating traditional class-room trainings. It is therefore not surprising that e-learning was thought to be one of the fastest growing sectors [2] and to revolutionize education [3]. While successful adoption of e-learning implementations was considerable in academia [4], it was only limited in the corporate context [5]. The reason for this lack of success is unlikely to be the e-learning technology, content or design as these factors hardly differ between academic and corporate contexts. The contextual difference is more likely to stem from the ultimate beneficiary. In the corporate context, it is not the learner ultimately benefiting from e-learning, but the investing corporation. Thus, additional motivation of the learner is one crucial factor for corporate e-learning to be successful [6].

Social software is expected to increase motivation [7-9]. The term social software is used synonymously with web 2.0, which basically refers to web-based software allowing users to not only consume information, but also to generate information and to socially interact with each other [10-11]. The combination of e-learning and social software is also termed e-learning 2.0, which promises a new era in technology enabled learning, potentially providing the means for life-long learning even in dynamic environments [12-14]. Unfortunately, it is not often implemented in the corporate e-learning context yet [15], which is probably why there is no research explaining which and how individual social software features impact motivation in this context. As a result, there is no guidance for practitioners designing, delivering, and implementing e-learning [16], which calls for the research question.

How do social software features impact the motivation for corporate e-learning participation?

The answer could lead to corporate e-learning becoming successful and to life-long learning becoming more than a popular phrase.

This research uses a narrow definition of e-learning to only include asynchronous technologies, i.e., technologies that provide electronic, time and place-independent access to knowledge, which is preprocessed for educational purposes. By contrast, synchronous technologies such as online conferencing and collaboration tools, which facilitate real-time distance trainings, are not included. This is due to the fact that social human interaction is already part of the definition of synchronous e-learning limiting the potential impact of additional social features on motivation.

The subsequent structure of this research is twofold. First, we approach the research question by providing a literature review of the training and e-learning motivation literature. Due to the limited findings, we approach the research question in a second step. By the help of descriptive statistics and an exploratory factor analysis we identify distinct social software factors that potentially impact motivation to participate. This paper finishes with a discussion on the resulting patterns and implications.

2 Reviewing Training Motivation Literature

Since motivation research has a long history and cuts “across all subareas within psychology” [17], a diversity of corporate training motivation models are published in reputable journals of the last 20 years (see Table 1 for examples). Despite the models’ differences, they also share common findings and constructs, which can be associated with more abstract motivational facets [18]. Thus, this review is first concerned with comprehensively identifying these abstract motivational facets. Secondly, it evaluates the comprehensiveness of training motivation models using the identified facets and discusses the models’ potential to explain the impact of social software features on motivation.

2.1 Overview of Abstract Motivational Facets

Figure 1 illustrates a comprehensive overview of motivational facets, which is based on the conceptual overview of Heckhausen & Heckhausen [19].

With Lewin's Force Field Theory in mind, the overview distinguishes between "Personal and environmental factors" as their combination is critical to explain behavior [20-21]. This part also represents a long history of Personal motivation research starting with Darwin, who basically claimed that humans as biological organisms are driven to survive long enough to reproduce [22], and continuing with subsequent researchers, who refined Darwin's drives using different terminology such as needs, instincts, motives or traits and ergs [23-28].

"Attitudes/expectancies" as well as "Behavior" are originally based on Vroom's Expectancy theory and his Valence-Instrumentality-Expectancy model [29], which was later extended by Porter & Lawler III [30] to one of the most complex motivation models. The resulting overview illustrated in Figure 1 accommodates most commonly used constructs such as Self-Efficacy. Defined as the "judgments as to how well one can execute the requisite behavior" [31], the overview would classify it as a performance expectancy.

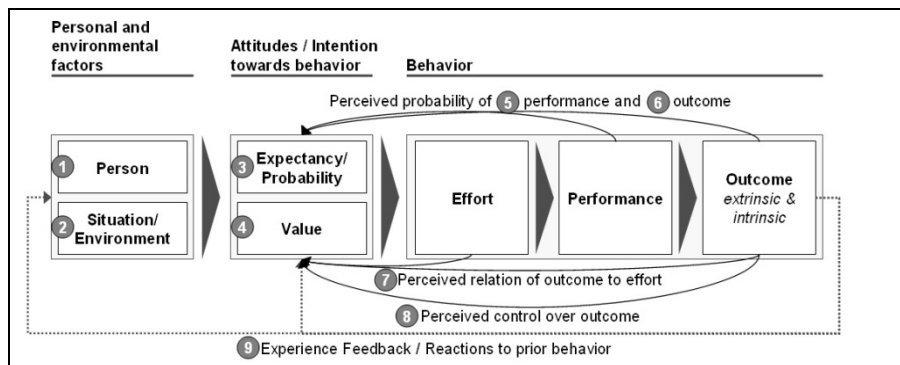


Fig. 1. Comprehensive overview of abstract motivational facets

The motivational facets 7, 8 and 9 in Figure 1 are further extensions. Number 7 and 9 replace Porter & Lawler's III concepts of 'perceived equitable rewards' and 'satisfaction'. Instead, the more comprehensive concept of 'perceived inequity' based on the relation of outcome to effort [32] and a 'feedback loop' suggested by the Social Learning Theory [33] are included. Finally, motivational facet number 8 was added illustrating a different kind of expectancy over the outcome based on the concept of locus of control from the theory of planned behavior (TPB), which postulates high levels of perceived control to strengthen the intention towards a behavior [34].

2.2 Discussion of Review Results

Table 1 contains the results of the review of training motivation models the along nine motivational facets previously illustrated in Figure 1. In the following, these models

are briefly discussed regarding their general theoretical comprehensiveness and regarding their potential to explain the impact of social software features.

Both Birdi et al. [35] as well as Maurer & Tarulli [36] use a remarkably broad range of motivational factors. Unfortunately, the former do not make much use of the underlying basic motivation theory, which leads to the choice of motivational factors appearing rather arbitrary. The latter do not consider the impact of prior experiences critical to the training context [33]. Both models contain the construct “Coworker support”, which provides some ground to suggest social software has an impact on motivation, but neither model is actually suitable to substantiate this suggestion and to explain the impact of different social software features.

The Technology Acceptance Model (TAM) [37] is used by Cheng [8], Venkatesh [38], and Wakefield [7]. All authors extended TAM by adding intrinsic motivational factors, but the resulting models still miss out on other aspects of motivation theory as illustrated in Table 1. Helpful regarding the impact of social software are the models of Cheng and Wakefield as they do claim that system factors such as social software impact usefulness, ease of use, and enjoyment positively [7-8]. Regrettably neither model provides further guidance on individual software features. Furthermore, Wakefield’s et al. findings regarding socialness may not be generalizable to an asynchronous e-learning context, because the authors defined and tested socialness with video-based user interaction [7]. In an asynchronous environment, videos are non-interactive and known not to improve e-learning effectiveness [39].

Noe’s model [40] serves as a basis for both Garavan et al. [41] and Tharenou [42]. The former authors refer to other research publications such as the theory of planned behavior [43-44] or the below discussed comprehensive model of Hurtz & Williams [45], but they do not make use of them in their own model. The latter author found that motivation to learn, outcome expectancies, and supervisor support play an important role in training motivation. Unfortunately, antecedents such as social software are not considered.

The most comprehensive model on training motivation is that of Hurtz & Williams [45] with only minor limitations. It contains only one personal motive, it lacks the motivational concepts ‘perceived inequity’, and it sometimes contradicts previous research findings such as those regarding “perceived control”, which does not correlate with intention as it should according to TPB [43-44]. Unfortunately, social factors only comprise social pressure and social support, which does not include social interactivity as provided by social software features. As a result, even this most comprehensive model does not allow deriving hypotheses regarding the potential impact of social software features on motivation to participate.

For more obvious reasons, the three remaining research models are also not helpful in explaining the potential impact of social software. Chiu et al. [46] exclusively focus on the concepts fairness and quality, which are not only rarely used to explain e-learning participation, but also too broad to explain the impact of social software features. Clark et al. [47] explicitly use a non-comprehensive model excluding system factors, and Noe and Wilk [48] disprove their research model.

The work of Soliman and Beaudry is not included in Table 1 as it is still research-in-progress, but worth mentioning in this context [49]. They explicitly opted against

using an established, empirically supported research model in favor of a more general Person-Environment Fit based approach trying to explain the impact of social software on system use in general. Once finalized, this research could yield findings relevant to the context of corporate e-learning.

Table 1. Comparison of training motivation models by nine antecedents of behavior (see Figure 1)

	(1) Person	(2) Environment	(3) Expectancy	(4) Valence	(5) Probability of performance	(6) Probability of outcome	(7) Inequity / Relation of outcome to effort	(8) Control over outcome	(9) Reactions to prior behavior
Birdi et al. [35]	•	•	•	-	•	•	-	-	-
Cheng [8]	•	•	•	-	•	•	-	-	-
Chiu et al. [46]	-	-	•	-	-	-	•	-	•
Clark et al. [47]	-	•	•	•	-	•	-	-	-
Garavan et al. [41]	-	•	•	-	•	-	-	-	-
Hurtz & Williams [45]	•	•	•	•	•	•	-	-	•
Maurer & Tarulli [36]	•	•	•	•	•	•	-	-	-
Noe & Wilk [48]	•	•	•	-	•	•	-	-	-
Tharenou [42]	•	•	•	•	-	•	-	-	-
Venkatesh. [38]	•	-	•	-	•	•	-	-	-
Wakefield et al. [7]	•	-	•	-	•	•	-	-	-

Summing it up, none of the models discussed, neither the most comprehensive model with regards to motivation theory of Hurtz & Williams [45] nor the most advanced e-learning adoption model of Cheng [8], are detailed enough to explain the impact of social software features on training motivation sufficiently.

3 Identifying Relevant Social Software Artifacts

Although there is no model comprehensively explaining how social software features impact motivation yet (see previous chapter), existing research does support the general assumption that social software features positively impact motivation [8-9], in some cases even on more than one motivational facet [7]. A possible explanation for this influence of social features could be the additional information and support from others, which they provide and which are found to have a positive impact on motivation [50], especially in cases of organizational encouragement [51], of high social presence [52-53], and of a structured and closed societies like the workforce of a corporation [54]. Thus, the proposition, which this research evaluates, is as follows.

Social software features that corporate e-learning users consider favorable i.e., useful in providing relevant information and support from others, have a generally positive impact on motivation, which may vary in the extent of their impact.

The following sections outline the research method, present the findings, and discuss implications for researchers and practitioners.

3.1 Research Method

The aim of this research is to approach the research question by evaluating the proposition stated above. Findings evolve bottom-up from empirical data gathered in structured interviews with corporate e-learning users on their general opinion on social software features in case they were implemented in the corporate e-learning context. This section outlines details on the interviews, the sample, and the analyses performed.

We opted for personal interviews in case participants have questions regarding individual features. The interviews are structured because both social software and motivation theory have a long history and are well-defined. Social software features have been around for nearly ten years [10], and research on motivation theory date back to Aristotle [55]. Besides gender and age, the survey included general items such as “Linking the e-learning environment to a social network has a positive influence on my motivation to participate in e-learning (assumption: I decide freely who can view my information and which information about my e-learning participation is made public)” and specific social software features listed in Table 2. For each item, the survey captured both the level of feature implementation in the e-learning environments used so far and the potential favorability of a feature in this context on 7-point Likert-scales ranging from 1 (not implemented/very unfavorable) to 7 (fully implemented/very favorable) with 4 being neutral on the favorability scale.

The sample needs to fulfill three major requirements for the interview results to be meaningful. Participants need to have a genuine interest in learning using IT, they need to be familiar with e-learning in the corporate context, and they need to be very knowledgeable about social software features. The interviewed sample contains 39 young German IS employees of a well-known international IT services company. The participants are working students who work in different departments of the organization with each department having the freedom to use different e-learning platforms. They work in different locations across Germany, cover a wide range of functions including Marketing, Sales, Consulting, Customer Support, and Development, and work on various IT services including hardware and software. The average age of the sample is 22 years, and it contains 15% females. This sample not only has a high need for flexible ways of learning given the high pace of the IT industry, but also comprises the IS enthusiasm required to be most familiar with social software features.

SPSS version 19 supported both a descriptive analysis and an exploratory factor analysis. The former analysis helps identifying features that users consider as favorable in the e-learning environment (both median and mean greater 4). The latter analysis uncovers patterns of differentiated use of social software features providing further

insights on the potentially different extent of their impact on motivation. It is combined with a parallel analysis of Horn [56], which determines the final number of main factors. O'Conner provides the necessary SPSS script [57].

3.2 Findings

Overall, the *descriptive analysis* reveals that users have a favorable view on social software features in the context of corporate e-learning (M 4.6 SD 1.63), but these features are hardly implemented in the different e-learning platforms that the participants had exposure to (M 2.05; SD 1.44). Table 2 contains the participants' opinion on each individual features and indicates three general findings, which are as follows. First, not favorable features in an e-learning context are personal forums of peers (M 3.9; SD 1.73) and supervisors (M 3.8; SD 1.69) as well as newsfeeds (M 3.8; SD 1.75) and status messages of others (M 4.0; SD 2.00) since both mean and median are equal or below 4. The second general findings is that e-learning users wish for a closer social interaction with a trainer using asynchronous communication (M 5.1; SD 1.37), comments (M 5.2; SD 1.49), and personal forums (M 4.6; SD 1.65). Finally, the interview participants would value the opinion of their peers expressed in ratings (M 5.4; SD 1.57) and comments (M 5.1; SD 1.69) on e-learning modules. They further consider collaborative learning using forums (M 5.1; SD 1.61) as valuable.

Applying an *exploratory factor analysis* in combination with parallel analysis reveals three main components from the favorable social software features of Table 2. The relatively small sample size of 39 still led to a satisfactory Kaiser-Meyer-Olkin value of 0.673 and a significant Bartlett test of sphericity, which both confirm that the data is usable for such an analysis. Table 2 also includes the resulting factor loadings after Varimax rotation if they are greater than 0.5.

3.3 Discussion

This section discusses both the *descriptive analysis* and *exploratory factor analysis* findings and thereby outlines implications for practice and future research. The first of three findings revealed by the *descriptive analysis* implies that neither personal forums of peers and supervisors, nor newsfeeds and status messages of others will be beneficial if implemented in a corporate e-learning environment. With exception of newsfeeds, these features have in common that they provide communication channels allowing others to state their opinion on e-learning content. The reason why these channels are unfavorable is probably because they are indirect. They can be neglected in future corporate e-learning implementations because direct features such as comments and ratings, which allow others to directly state their opinions on respective e-learning module or content items, are more favorable. The reason why newsfeeds are unfavorable is unclear. Newsfeeds is a feature used to combine information from different sources on one website [10]. They potentially allow e-learning users to receive individualized information on newly available e-learning modules without having to visit the e-learning platform. As the interview participants might not have

thought of this advantage, it is advisable not to neglect newsfeeds completely during future implementations, but to assign them to a lower priority.

The second finding, a closer social interaction of the learner with the trainer, addresses a long-criticized downside of traditional asynchronous e-learning, which is the absence of a human trainer consultable in case of questions during the learning process. This desperate need for contact with a trainer also explains why the interview participants consider even an indirect communication channel such as a personal forum of a trainer as favorable. Thus, the general implication is that any social software feature available that allows communication with a trainer is worth implementing.

Finally, as suggested before, corporate e-learning users consider direct channels as favorable, which allow them to contribute, to communicate, and to learn from each other. This potential of social software in the e-learning context is often mentioned [12-14], and its validity supported by the interview participants. Further research and experience is required to identify the best way of how collaborative learning could take place in an asynchronous environment.

The three revealed main components of the *exploratory factor analysis* extends the previous discussion as they appear to be representations of distinguishable user needs in the context of corporate e-learning. Factor A seems to combine features that meet the need of e-learning users to receive additional opinions and meta-information about e-learning modules e.g., ratings or comments of peers and trainers. Furthermore, since e-learning users tend to distinguish between comments of users and comments of trainers, future e-learning environments presenting meta-information should make the same distinction. Possible ways could be to mark comments from the trainer in a different color or to allow participants to hide comments from their peers.

Factors B and C appear to represent the need for two distinct learning methods, which are learning from experts and learning from peers. This implies for practice that building new asynchronously collaborative e-learning modules might be more effective than merely enriching traditional “learning-from-expert” modules with social features. The most obvious and easiest way to implement asynchronous collaboration would be a discussion blackboard or forum, but there is certainly potential for more innovative ways. Similar to a class-room discussion, there could be ways to provide an overall guiding structure to the discussion or to enable users to filter helpful discussion contributions from disturbing individual opinions. It is interesting to note that collaborative forums such as wikis are rather associated with learning from experts (factor B) than with learning from peers (factor C). A possible explanation could be that young IS employees regard collaborative forums such as Wikipedia rather as a mean to receive expert information than as a way to contribute and to collaborate. It could be an indication that corporate e-learning users also do not know how to effectively implement asynchronous collaboration in an e-learning environment, although they generally support the idea of it as mentioned above. Implications for future research on motivation to participate in e-learning include the need to clearly address the distinction between “learning-from-experts” and “learning-from-peers” as the learners appear to have distinct learning preferences, which could impact the motivation to participate.

In summary, the findings on the hypothetical use of social software in a corporate e-learning environment provide the necessary guidance and insights for practitioners and future researchers [16], but they do not finally answer the research question on how social software features impact motivation. Joining other researchers in calling for more design research in the e-learning space [58-59], this research also believes that the iterative trial-and-error approach of design science is the ideal approach extending boundaries to unknown research areas [60-63], which could potentially lead to a conclusive answer of the research question.

Table 2. Descriptive statistics and factor loadings of social software features

	Usefulness in e-learning					Factor loadings ^c		
	N ^a	Mean	SD	Min/Max	Median	A	B	C
<i>Personal forums - text based (blog)</i>								
of peers ^b	38	3.87	(1.73)	1/7	4	-	-	-
of trainers	38	4.61	(1.65)	1/7	5		0.72	
of supervisors ^b	38	3.82	(1.67)	1/7	4	-	-	-
<i>Comments on modules</i>								
from peers	39	5.05	(1.69)	1/7	5	0.60		
from trainers	39	5.18	(1.48)	2/7	5	0.54	0.58	
from supervisors	39	4.62	(1.68)	1/7	5		0.67	
<i>Ratings of modules</i>								
from peers	39	5.41	(1.57)	1/7	6	0.80		
from trainers	39	4.85	(1.57)	2/7	5	0.82		
from supervisors	39	4.38	(1.70)	1/7	5	0.64		
<i>Asynchronous communication (chat)</i>								
with peers	39	4.67	(1.54)	1/7	5			0.78
with trainers	39	5.10	(1.37)	2/7	5			0.80
with supervisors	39	4.10	(1.59)	1/7	4		0.80	
Comments on items	39	4.64	(1.74)	1/7	5	0.74		
Sharing of items with others	39	4.44	(1.80)	1/7	5	0.78		
Collaborative forum (wiki)	39	5.13	(1.61)	1/7	5		0.66	
Earmarking of items (tagging)	39	4.79	(1.47)	2/7	5	0.57		0.58
Search content by person	39	4.13	(1.66)	1/7	4			0.62
Status messages (micro blogs) ^b	38	4.00	(2.00)	1/7	4	-	-	-
Newsfeed on trainings (RSS) ^b	38	3.82	(1.75)	1/7	4	-	-	-

a) N varies due to missing data

b) excluded from factor analysis since mean and median ≤ 4

c) Factor loadings < 0.5 are not displayed

4 Limitations

This research found that the research question on how social software features impact motivation to participate is still open, but it does not provide an empirically supported research model conclusively answering it. Instead, providing the necessary link between expertise in social software and corporate e-learning, it paves the way for future research on this question. Although sufficient for this research purposes, the sample size is relatively small and taken from only one large organization limiting generalizability of the findings.

5 Conclusion

Current research does provide evidence for a positive impact of social software on training motivation, but it does not explain the impact of individual features leaving practitioners without guidance and leading to a gap between e-learning researchers and practitioners [16]. The literature part of this research corroborates the eminence of this gap, while the interview part contributes to closing this gap by providing insights and guidance for future implementations and research. It does not conclusively answer the research question and supports the opinion of other e-learning researcher calling for further design research [58-59], which aims for both conclusively answering the research question and closing the gap to practitioners.

References

1. Drucker, P.F.: The New Society of Organizations. In: Howard, R. (ed.): The Learning Imperative: Managing people for continuous innovation. Harvard Business Press (1993)
2. Martin, G., Massy, J., Clarke, T.: When absorptive capacity meets institutions and (e)learners: adopting, diffusing and exploiting e-learning in organizations. *International Journal of Training and Development* 7, 228–244 (2003)
3. Hiltz, S.R., Turoff, M.: Education goes digital - The Evolution of Online Learning and the Revolution in Higher Education. *Communications of the ACM* 48, 59–64 (2005)
4. Wirt, S., Choy, S., Rooney, P., Provasnik, S., Sen, A., Tobin, R.: The condition of Education 2004 (NECS 2004-077), Washington, D.C. (2004)
5. Fee, K.: *Delivering E-Learning: A complete strategy for design, application and assessment*. Kogan Page Limited (2009)
6. Dublin, L.: If You Only Look Under the Street Lamps...or Nine e-Learning myths. *The e-Learning Developer's Journal*, 1–7 (2003)
7. Wakefield, R. L., Wakefield, K. L., Baker, J., Wang, L. C.: How website socialness leads to website use. *European Journal of Information Systems* 20, 118–132 (2010)
8. Cheng, Y.-M.: Antecedents and consequences of e-learning acceptance. *Information Systems Journal* 21, 269–299 (2011)
9. Wang, M., Vogel, D., Ran, W.: Creating a performance-oriented e-learning environment: A design science approach. *Information & Management* 48, 260–269 (2011)
10. Ebersbach, A., Glaser, M., Heigl, R.: *Social Web*. UVK, Konstanz (2011)
11. O'Reilly, T.: What Is Web 2.0, <http://oreilly.com/web2/archive/what-is-web-20.html>

12. Brown, J.S., Adler, R.P.: Minds on Fire: Open Education, the Long Tail, and Learning 2.0. *Educause Review* 16–32 (2008)
13. Yang, H.H., Yuen, S.C.-Y.: Collective intelligence and e-learning 2.0. *Information Science Reference* (2009)
14. Joint-Research-Center: Learning 2.0: The Impact of Web 2.0 Innovations on Education and Training in Europe. *Dictus Publishing* (2011)
15. Andriole, S.J.: Business impact of Web 2.0 Technologies. *Communications of the ACM* 53, 67–79 (2010)
16. DeRouin, R.E., Fritzsche, B.A., Sales, E.: E-Learning in Organizations. *Journal of Management* 31, 920–940 (2005)
17. Petri, H.L., Govern, J.M.: *Motivation - Theory, Research, and Applications*. Wadsworth Publishing (2003)
18. Colquitt, J.A., LePine, J.A., Noe, R.A.: Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology* 85, 678–707 (2000)
19. Heckhausen, J., Heckhausen, H.: *Motivation und Handeln*. Springer, Heidelberg (2006)
20. Lewin, K.: *Principles of Topological Psychology*. McGraw-Hill (1936)
21. Lewin, K.: The conceptual representation and measurement of psychological forces. *Duke University Press* (1938)
22. Darwin, C.: *The Expression of the Emotions in Man and Animals*. John Murray (1872)
23. Freud, S.: *Triebe und Triebchicksale*. Fischer, Frankfurt, Germany (1952)
24. McDougall, W.U.: *An introduction to social psychology*. John W Luce & Company, Boston, MA, US (1918)
25. Murray, H.A.: *Explorations in personality*. Oxford University Press, New York (1938)
26. Maslow, A.H.: A Theory of Human Motivation. *Psychological Review* 50, 370–396 (1943)
27. Cattell, R.B.: *Personality and motivation structure and measurement*. World Book (1957)
28. Reiss, S.: Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires. *Review of General Psychology* 8, 179–193 (2004)
29. Vroom, V.H.: *Work and Motivation*. New York (1964)
30. Porter, L., Lawler III, E.E.: *Managerial Attitudes and Performance*. Homewood (1968)
31. Bandura, A.: Reflections on Self-Efficacy. In: Rachman, S. (ed.): *Advances in Behavioral Research and Therapy*. Pergamon Press, Oxford, England (1978)
32. Adams, J.S.: Towards an understanding of inequity. *Journal of Abnormal and Social Psychology* 67, 422–436 (1963)
33. Bandura, A.: *Social Learning Theory*. Prentice-Hall, Englewood Cliffs, N.J (1977)
34. Ajzen, I.: Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology* 32, 665–683 (2002)
35. Birdi, K., Allan, C., Warr, P.: Correlates and perceived outcomes of four types of employee development activity. *Journal of Applied Psychology* 82, 845–857 (1997)
36. Maurer, T.J., Tarulli, B.A.: Investigation of perceived environment, perceived outcome, and person variables in relationship to voluntary development activity by employees. *Journal of Applied Psychology* 79, 3–14 (1994)
37. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13 (3), 319–340 (1989)
38. Venkatesh, V.: Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation. *MIS Quarterly* 23 (2), 239–260 (1999)

39. Zhang, D., Zhou, L., Briggs, R.O., Nunamaker, J.F.: Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management* 43, 15–27 (2006)
40. Noe, R.A.: Trainees' Attributes and Attitudes: Neglected Influences on Training Effectiveness. *The Academy of Management Review* 11, 736–749 (1986)
41. Garavan, T.N., Carbery, R., Malley, G.O., Donnell, D.O.: Understanding participation in e-learning in organizations: a large-scale empirical study of employees. *International Journal of Training and Development* 14, 155–168 (2010)
42. Tharenou, P.: The relationship of training motivation to participation in training and development. *Journal of Occupational & Organizational Psychology* 74, 599–621 (2001)
43. Ajzen, I.: From intentions to actions: A theory of planned behavior. In: Kuhl, J., Beckmann, J. (eds.): *Action control: From cognition to behavior*. Springer, Heidelberg (1985)
44. Ajzen, I.: The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 179–211 (1991)
45. Hurtz, G.M., Williams, K.J.: Attitudinal and motivational antecedents of participation in voluntary employee development activities. *Journal of Applied Psychology* 94, 635–653 (2009)
46. Chiu, C.-M., Chiu, C.-S., Chang, H.-C.: Examining the integrated influence of fairness and quality on learners' satisfaction and Web-based learning continuance intention. *Information Systems Journal* 17, 271–287 (2007)
47. Clark, C.S., Dobbins, G.H., Ladd, R.T.: Exploratory Field Study of Training Motivation: Influence of Involvement, Credibility, and Transfer Climate. *Group & Organization Management* 18, 292–307 (1993)
48. Noe, R.A., Wilk, S.L.: Investigation of the factors that influence employees' participation in development activities. *Journal of Applied Psychology* 78, 291–302 (1993)
49. Soliman, M.A., Beaudry, A.: Understanding individual adoption and use of social computing: A user-system fit model. In: *International Conference on Information Systems (ICIS) Proceedings* (2010)
50. Sykes, T.A., Venkatesh, V., Gosain, S.: Model of Acceptance with Peer Support: A Social Network Perspective to Understand Employees' System Use. *MIS Quarterly* 33, 371–393 (2009)
51. Goodman, P.S., Darr, E.D.: Computer-Aided Systems and Communities: Mechanisms for Organizational Learning in Distributed Environments. *MIS Quarterly* 22 (4), 417–440 (1998)
52. Olivera, F., Goodman, P.S., Tan, S.S.: Contribution Behaviors in Distributed Environments. *MIS Quarterly* 32, 23–42 (2008)
53. Kankanhalli, A., Tan, B.C.Y., Wei, K.-K.: Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation. *MIS Quarterly* 29, 113–143 (2005)
54. Jarvenpaa, S., Staples, D.S.: The use of collaborative electronic media for information sharing: an exploratory study of determinants. *The Journal of Strategic Information Systems* 9, 129–154 (2000)
55. Hardie, W.F.R.: *Aristotle's ethical theory*. Clarendon Press, Oxford (1980)
56. Horn, J.L.: A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185 (1965)
57. O'Connor, B.P.: SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc.* 32, 396–402 (2000)
58. Kelly, A.E.: Research as Design. *Educational Researcher* 32, 3–4 (2003)

59. Herrington, J., Reeves, T.C., Oliver, R.: A guide to authentic e-learning. Routledge Chapman & Hall (2010)
60. Dubin, R.: Theory Building. Free Press, London (1978)
61. Gregor, S., Jones, D.: The Anatomy of a Design Theory. *Journal of the Association for Information Systems* 8, 312–335 (2007)
62. Hevner, B.A.R., March, S.T., Park, J., Ram, S.: Design Science in information systems research. *MIS Quarterly* 28, 75–105 (2004)
63. Walls, J., Widemeyer, G., El Sawy, O.: Building an Information System Design theory for Vigilant EIS. *Information Systems Research* 3, 36–59 (1992)

Resistant Use of Project Management Methodologies – Using Psychology to Rethink the Influence of Methodology Attributes

Kunal Mohan and Frederik Ahlemann

EBS Business School, Oestrich-Winkel, Germany
{kunal.mohan, frederik.ahlemann}@ebs.edu

Abstract. Even though practitioners and researchers generally agree that project management methodologies are very useful in managing IT-related projects, organizations are rarely able to motivate their staff to use them. Even when employees use these methodologies, the question of how they are being used still remains. To better understand the way in which employees use project management methodologies, we develop the construct “resistant use.” Applying a diffusion of innovations perspective, we develop a conceptual model to examine how methodology attributes interact with users psychological needs to influence a user’s resistant use behavior. Based on a sample size of 2645 participants, and using the structural equation modeling technique, we find that a user’s need for achievement and need for cognition moderate the impact of project management methodologies’ attributes (relative advantage, complexity, and compatibility) on their resistant usage behavior.

Keywords: Project management, Resistant use, Behavioral science

1 Introduction

In the search for systematic and predictable ways to find replicable, pragmatic, cost-effective, and timely solutions to IT development problems, organizations either adopt or customize and adaptively apply project management methodologies (PMMs), which consist of tested bodies of methods, rules, and procedures. Despite the overwhelming advantages of using a PMM, organizations are often unable to compel their staff to use such methodologies. For example, a software development project survey conducted by Russo et al. [1] shows that only 6% of organizations claim that their methodologies are always used as specified. Cicmil et al. [2] found that resistance to the acceptance of project management methodologies is high because users do not have faith in the concept, fear a loss of power, or lack adequate training and support from upper management. In light of this evidence it is clear that methodologies’ potential benefits cannot be realized if organizational members who can benefit from their use resist using them. Moreover, even when employees use these methodologies, the question regarding the nature of their use remains, i.e. *how*

are they using them? This question is critical because *use* alone does not mean that employees are dedicated to using a PMM. Because the use of organization specific PMM is not volitional i.e. employees have no choice regarding the use of a PMM, their act of using a PMM might be superficial. Deep within they might be actually using the PMM in a resistant, counterproductive manner. Understanding employees resistant usage behavior is critical, because although a particular methodology is developed, implemented, and forced upon employees by an organization, the way in which it is used is only determined by the methodology's users. Reasons why PMM usage might be challenging, causing resistance, derive from the tacit, organizational, and individual problems caused by the introduction of a new methodology. The stress associated with learning a new methodology, fear, the impact on self-esteem and identity associated with an organizational restructuring, the emotional costs of role conflict, ambiguity, and workplace transformation might be serious inhibitors of commitment motivations [3]. Based on existing literature, we therefore attempt to capture the impervious, non-compliant, unwilling nature of PMM usage behavior by conceptualizing it as *resistant use*. As discussed in a later section, this specific construct captures the negative, destructive, and counterproductive character of "use", the mysteries of which both organizations and researchers are constantly trying to unearth and prevent. This approach might help organizations to not simply make employees use a PMM but also to reduce the often passive and covert resistant use behavior.

The second issue concerning the methodology adoption and use domain pertains to the study of usage antecedents from singular perspectives. Some of these studies have attempted to examine individual usage behavior regarding IS methodologies from a technology adoption perspective (e.g., [4-5]), while others apply sociological models, such as the theory of planned behavior (TPB) [6] to examine the development of individuals' intention to use methodologies (e.g., [7-8]). Unfortunately, the various disciplines, generally concentrating on their individual variables, have neglected to incorporate methodology and personality attributes into their understanding of the methodology use problem. Little is known about the interactive effects of methodologies' attributes and non-technical, psychological personality characteristics, and it seems reasonable to consider variables from both sets important in explaining the problem at hand [9].

Applying a *methodology-centric* perspective, based upon the diffusions of innovation theory (DOI) [10], we develop and test a conceptual model to examine how methodology attributes (technical perspective) interact with users' psychological needs (psychological perspective) thereby influencing their resistant use behavior. The integration of human needs is important because, even though past research has shown that IT artifacts have certain technical attributes and influence a person's usage behavior, the results across different domains and contexts have been inconsistent and sometime contradictory (e.g. effects are found to be insignificant or to have a negligible effect [4-5]). Human needs might be able to explain this variation and show us that, based upon an individual's intrinsic psychological characteristics, different methodology attributes influence a person's usage behavior in different ways. Murray's [11] theory of psychogenic needs provides a comprehensive theoretical basis to help us

understand how, when, and which specific needs are more important to what type of people.

Our study contributes to filling the gap in the methodology development, adoption, and implementation literature, which, until now, has neither developed a theoretically and practically complete nor relevant taxonomy of potential methodology characteristics. Moreover, no prior studies have investigated the way in which these characteristics interact with people's need to affect their methodology usage behavior. This leads to fundamental questions regarding the nature of employees' methodology use: a) How do PMM attributes affect an individual's resistant usage behavior?; b) How do human needs interact with the PMM attributes to explain resistant use? The remainder of the paper is organized as follows: Section 2 provides an overview of the theoretical foundations that provide the framework for our conceptual model. In Section 3, we present our research model and hypotheses. In Section 4 we discuss the methodology and results, while we discuss the study's implications and contributions in Section 5.

2 Theoretical Background

A PMM is viewed as a framework that is used to structure, plan, and control the process of developing or managing IS artifacts (e.g. software, hardware, infrastructure, processes etc.). The main idea behind PMMs is to facilitate the management of projects in a very deliberate, structured, and methodical way, requiring each stage of the project life cycle – from the inception of the idea to the handover of the final deliverables – to be carried out rigidly and sequentially. Some of the most fundamental characteristics and advantages that justify the use of such structured PMMs, as identified by Fitzgerald, (1998) are the following: i) They reduce complexity by subdividing the project development and management process into plausible and coherent steps; ii) They increase transparency and therefore facilitate control over the activities, thus reducing the risk and uncertainty of projects; iii) They provide a goal-oriented framework that helps to direct the application of techniques and resources at appropriate times during the project.

2.1 Nature of Use

Resistant Use (RU). The resistance concept has been a core focus of the management science and organizational behavior literature with regard to employees' resistance to accepting management-initiated changes. This resistance is generally negatively viewed as an obstacle or barrier to change and progress. In short, resistance is something to be resisted. Long before management sciences discovered "resistance" as a potent construct, it had already had a long history, characterized by rich theoretical developments, in the psychology literature. From psychology we learn that resistant behavior is a means by which the subject of the resistance attempts to acquire external and/or internal benefits. According to research conducted in the domain of cognitive-affective psychology, resistant behavior could be driven, consciously or unconsciously, by either a) cognitive or b) affective processes. For example, in case of *cognitively*

driven resistant behavior, a person would rationally analyze a task at hand and conclude that it would take too long to master it and was not worth the effort, consequently reject it. With regard to *affectively* driven resistant behavior, the task at hand might unconsciously generate intolerable emotions of anxiety, anger, fear, etc. (e.g. based upon past failures, experiences, phobias, etc.) due to which the person might automatically avoid the task without any cognitive deliberation or logic.

Based upon psychoanalytic theory, especially the work of Freud [12], resistant behavior is considered to be also driven by desires and emotions that occur without an individual being consciously aware of them i.e. through unconscious processes such as defense mechanisms that emerge involuntarily whenever an individual perceives psychic danger. In Freudian psychoanalytic theory, these defense mechanisms are psychological strategies to protect an individual's mind from anxiety, which is an aversive psychological inner state. Anxiety is a core concept in psychoanalytic theory and, when experienced in an intense or acute form, is the most unpleasant feeling an individual can experience. It arises from internal conflicts between one's primitive desires, from the constraints of reality, and from one's values and beliefs, or when an external threat is perceived. Consequently, when anxiety becomes too overwhelming, individuals deploy defense mechanisms that distort, transform, or falsify reality is some way, to protect themselves from unpleasant feelings. Resistant behavior allows individuals to eliminate psychological threats by avoiding actions or blaming the object of behavior that could be contributing to anxiety.

Following the discussion above, we propose that *resistant use* occurs when an individual is consciously or unconsciously opposed to the usage behavior, which is either based on a rational cost-benefits analysis or on feelings of anxiety, and actively, passively and overtly or covertly tries to avoid *usage* by, for example, refusing, arguing, delaying, or seeking to have the request or order to use the methodology nullified.

2.2 Diffusion of Innovations Theory

Research in an array of academic disciplines, such as anthropology, communication, geography, sociology, marketing, political science, public health, economics, social psychology, sociology, and political science, has applied DOI to understand the effect that new ideas, processes, and technologies have on people's adoption and usage decisions. Over the past five decades, DOI theory has been used to study how innovations diffuse into and are adopted by wider social networks [10]. While early research employing DOI theory concentrated on the diffusion and acceptance of products, the research community has recently reached consensus that *ideas* and *practices*, such as methodologies, can also be regarded as innovations if the potential adopter perceives them to be new [10]. According to Rogers [10], one of the most influential factors determining an innovation's adoption rate is the innovation itself, i.e. its *characteristics*. Based on the DOI theory, a methodology's characteristics play a crucial role in how potential users use it. The more attractive the attributes of a methodology are perceived to be, the more readily potential users accept that methodology, and the more dedicated they are when using it. Extensive empirical research has found that some of the attributes are more important than others. After conducting a meta-

analysis of 75 articles pertaining to innovation characteristics, Tornatzky and Klein [13] found that relative advantage, complexity, and compatibility are the only innovation characteristics consistently related to innovation adoption and implementation. Although extensive empirical evidence in various fields suggests that these influences are applicable in the context of methodology use – except for relative advantage – they have either been neglected or have been considered insignificant. Applying a least-square regression analysis, Riemenschneider et al. [5] individually tested five theoretical models regarding individuals' intention to accept information technology tools. Their aim was to understand why software developers accept or resist methodologies. They came to the following conclusions: *Perceived usefulness* was the only significant variable across all five models ($p < 0.001$), *voluntariness* was found to be not significant (or was not included) in three models, *compatibility* was found to be not significant (or was not included) in four models, and result *demonstrability*, *complexity*, *observability*, and *image* were found to be not significant (or were not included) in all five models. Hardgrave et al. [4] also investigate software developers' intentions to use methodologies, and find *usefulness* to be significant (although comparatively weaker), *complexity* to be not significant, and *voluntariness* and *compatibility* to be significant, but weak.

Recognizing the large gap in the innovation attributes proposed by the DOI theory and those studied in the context of methodology use, we identify two areas that require further attention: Future research needs to examine a) which of the wide range of innovation characteristics apply to the methodology domain, and b) how these different attributes affect resistant behaviors. As mentioned earlier, while the DOI theory provides a comprehensive list of attributes with which to examine the former issue, the latter problem is virgin territory.

2.3 Psychogenic Human Needs

Following Murray's [11] theory of psychogenic needs in humanistic psychology, individuals are expected to use a PPM based on their perceptions that it will enable them to fulfill their specific needs. Of the many definitions of basic needs that have been proposed, the one by Ryan and Deci [14] is the most applicable to this study. They indicate that "a basic need, whether it be a physiological need or a psychological need, is an energizing state that, if satisfied, conduces toward health and well-being but, if not satisfied, contributes to pathology and ill-being". As such, "what a person does" is determined by "what a person needs," i.e. people's behavior is determined by the needs they attempt to examine/fulfill either consciously or subconsciously. This implies that the factors that will be most influential in inducing resistance against PMM use are those that fail to satisfy people's basic needs. The inability to satisfy psychological needs results in serious discomfort. Moreover, this dissatisfaction might lead to the individual's rejection of the particular methodology. An individual's needs are thus expected to play a *moderating* role and influence the explanatory power of the determinants of resistant PMM use.

3 Conceptual Framework

Based on our discussion above, we hypothesize that a) specific methodology attributes influence employees' resistant PPM usage behavior, and b) employees' individual needs affect the explanatory power of the methodology attributes' effect. As outlined before, we combine both theories towards a comprehensive model explaining resistant behavior based on the insight that decision-making behavior regarding methodology use does not only depend on an assessment of a methodology's attributes but on a person's individual needs. A combination of both theories is useful and appropriate for the following reasons: (1) Both theories relate to the individual as the unit of analysis, (2) all relevant constructs can be measured through a survey, (3) the psychological assumptions underlying both theories are not contradictory, and (4) the philosophical assumptions of both theories are compatible. Moreover, it is to be expected that the explanatory power of the resulting model is significantly higher than that of the isolated theories.

3.1 Attributes of a Methodology

Relative advantage (RA) is the degree to which potential users perceive a methodology to be superior to its precursor, which is either the previous way of doing things (if there is no current way), the current way of doing things, or doing nothing. A methodology's superiority is not only measured in economic terms, but also in terms of a reduced or increased status and other benefits, for example, an increase in productivity and efficiency. Relative advantage is always measured in terms that matter to methodology users. Here, relative advantage mainly pertains to the workplace where the methodology is used. For example, a relative advantage may be perceived when the user can plan his project faster, he can make better decisions or he is able to steer his team in a more efficient way. All other factors being equal, the higher the relative advantage, the lower the resistance. Consequently, we propose that relative advantage will have a negative effect on resistant PPM use.

Complexity (CL) is the degree to which a methodology is perceived as difficult to understand and use. Complexity can e.g. be the result of a methodology's high number of procedural steps, the diversity of the activities included, the comprehensibility of its documentation, or the cognitive abilities required to master it. Complexity can result in a significant effort required to learn and master a methodology. Complexity may also increase the likelihood of mistakes when applying the methodology. E.g. a project management methodology may be error-prone because of highly complex business case documents or scheduling approaches. The more complex a methodology is perceived to be, the more resistance it is expected to generate. The more complex a methodology is perceived to be, the more an individual doubts his or her own ability to use a methodology properly. Complexity has been addressed in the technology adoption literature by the *ease of use* construct (which is also based on the concept of self-efficacy), which refers to the degree to which a person believes that using a particular methodology would be a) free of physical and mental effort, and b) easy to learn. Numerous empirical evaluations of *self-efficacy* and *ease of use* consider

these constructs an important predictor of human behavior and therefore provide substantial justification for including the *complexity* construct in our model. As such, we propose that complexity will have a positive impact on resistant PPM use.

Compatibility (CA) is the degree to which a methodology is perceived to be consistent with potential adopters' existing social cultural values and past experiences [15]. DOI's assumption is that any innovation may also be perceived as a risk. A high degree of compatibility with what is already known and mastered reduces an individual's risk to experience adverse effects. Adverse effects may be increased learning effort or frustration. Thus, the lower the compatibility, the higher the resistance to use the methodology. The roots of this lie in assumption that individuals in organizations might be reluctant to change their habits, which they have learned unconsciously through past repetitions, and might therefore be unwilling to adopt new methodologies that may cause radical change. In matters of radical change, such as new methodology adoption (see Section 2), the methodology might not be compatible with potential users' habits and could therefore evoke negative feelings and emotions and, consequently, resistance.

H1: *Relative advantage (RA) is negatively associated with the resistant use of a PMM*

H2: *Complexity (CL) will be positively associated with the resistant use of a PMM*

H3: *Compatibility (CA) will be negatively associated with the resistant use of a PMM*

3.2 Personal Needs

Need for achievement (nAch) refers to an individual's desire to do things better, accomplish difficult tasks, overcome obstacles, become an expert, achieve high performance standards, or need for a significant task-related accomplishment [11]. Such individuals are focused on internal motivation and personal achievement rather than external rewards and recognition. People with a high nAch aspire to accomplish difficult tasks in which success depends primarily on their efforts. Individuals with a high nAch are driven by their desire to have their success attributed to internal factors, i.e. their skills and competencies, rather than external factors such as luck or outside support, since attributions to internal factors produce stronger self-esteem related affective reactions than attributions to external factors. Such individuals are most satisfied when they know that they alone are responsible for a successful outcome. Empirical evidence confirms this.

However, according to the DOI perspective, relative usefulness is a characteristic of the methodology. Therefore, successful outcomes of a job that were achieved using a highly beneficial methodology are directly attributed to the methodology's usefulness (i.e. is external) rather than to the skills of the person who applied it. It generates the impression that anyone can be successful if they use that particular methodology. Therefore, psychologically, high achievers are dismayed by the notion that, even after putting in all the hard work, they might not get credit for the success. They would thus be more inclined to use a less useful methodology, because success in such a case might be directly attributed to their own abilities and contributions, giving them a heightened sense of achievement and pride (e.g. "even though the methodology was useless, I still got the job done"). We therefore propose that:

H4: The negative influence of *relative advantage (RA)* on resistant use will be moderated by *need for achievement (nAch)* such that the effect will be weaker for individuals with high nAch.

The more complex a methodology is perceived to be, the more gratification/satisfaction people with a high *nAch* are expected to feel, since being successful at using methodologies, which others fail to master, symbolizes and communicates personal competence. Studies have also shown that individuals with a high *nAch* are more committed to achieving difficult goals [16]. Individuals with a high *nAch* are expected to put *more effort* into tasks, *persevere longer* when confronted with obstacles, and show *resilience* in the face of complex methodology use. On the other hand, individuals with a low *nAch* avoid difficult tasks characterized by a high level of complexity, because their fear of failure greatly outweighs their expectation of success. We therefore propose that:

H5: The positive influence of *complexity (CL)* on resistant use will be moderated by *need for achievement (nAch)* such that the effect will be weaker for individuals with nCog.

Need for cognition (nCog) represents a desire for knowledge and reasoning [11], as well as the need to explore and discover. It represents the extent to which people engage in and enjoy *effortful* cognitive activities. Individuals with a high *nCog* tend to naturally seek, acquire, think about, and reflect on information by experimenting and exploring, to make sense of a problem at hand. Empirical studies have shown that people with a high *nCog* are generally more intelligent, conscientious, and open-minded, and therefore actively seek out challenging tasks. Consequently, people with a high *nCog* are more likely to want to use new and complex methodologies, as they would find it intellectually stimulating.

H6: The positive influence of *complexity (CL)* on resistant use will be moderated by *need for cognition (nCog)* such that the effect will be weaker for individuals with high nCog.

4 Research Methodology

Data Collection. The entire development process, leading to the final survey instrument, was conducted according to Straub's [17] recommendations. An initial pool of reflective measures was selected, based on their empirical validation in prior research. Instrument refinement was conducted based on interviews with 2 subject matter experts, Q-sorting exercise in 2 rounds with 7 and 8 participants respectively, and a web-based pre-test with 65 participants. Finally, all items were embedded in survey questions using a 7-point Likert scale anchored at *strongly disagree* (1) and *strongly agree* (7). Throughout the entire instrument development process, three researchers from different disciplines, nationalities, and institutions were always involved, discussing every issue and formulating improvements. Data was collected via an online survey for a period of four months. Participants for the study were collected through two approaches (personalized and anonymous): 1) They were randomly chosen utilizing databases of professionals (e.g. XING, Viadeo, CompetenceSite), with keyword

search (e.g. 'project manager'), and 2) International project management organizations (e.g. PMI, IPMA) sent out open invitations to all their members. 30 cases were excluded, because during data cleaning we noticed that the participants had responded in a similar manner for all questions (e.g. all questions answered with the same liker scale value). Personalized survey URLs were administered to a total of 7982 individuals, of whom 1246 completed the survey, representing a 16% response rate. In total, 1399 individuals responded anonymously, bringing the total number of participants to 2645. After the survey, we contacted all individuals who had been invited but had not participated in the personalized survey via email to inquire about their reason for non-participation. Overall, we received feedback from 613 *nonparticipants* and the most cited reasons for nonparticipation were: 1) the individual was the wrong contact person for the survey (45.68%). 2) a lack of time (39.8%). 3) no interest in topic (3.42%). 4) no interest in participation (2.28%). 5) the questionnaire was too long (2.21%). and 6) data confidentiality concerns (1.14%).

Data Analysis and Results. The research model was tested and the psychometric properties of the scales were assessed using the software SmartPLS (version 2.0 M3), which is based on partial least squares (PLS). We used PLS because, compared to covariance-based approaches, it is beneficial when the research model is relatively complex with a large numbers of indicators and multiple moderation effects, and the data is not normally distributed (Chin et al. 1996, 2003; Fornell and Bookstein 1982). Additionally, it has been argued that our chosen approach to analyze the moderation effects is far more difficult to implement in a covariance-based SEM context than in PLS path modeling [18]. The statistical significance of the parameter estimates was assessed using a bootstrapping procedure with 1000 resamples. In order to provide an overview of the survey instrument, detailed demographics, and additional statistical analysis results, which cannot be reported here due to limited space, we have compiled a document, which is available at <http://tinyurl.com/WII2013>.

Validation of the measurement model. We used reflective indicators for all the constructs. The adequacy of the measurement model was assessed by means of the individual items' reliability, the internal consistency between items, and the model's convergent and discriminant validity (see additional information file). Cronbach's alpha (CAP) and Dillon-Goldstein's rho (D.G.) were used to measure the internal consistency reliability. However, according to Chin [19], Dillon-Goldstein's rho is a much better indicator of reliability than Cronbach's alpha, since D.G. is derived directly from the model (i.e. the loading) instead of the correlations observed between the manifest variables in the dataset. In this study, the CAP and D.G. of each construct are greater than the recommended respective values of 0.50, and 0.70, which indicate the strong reliability of all the constructs in our model. Moreover, we followed Chin's [20] suggestion and calculated the composite reliability (CR) as an alternative to CAP. All the constructs' CR values are higher than 0.90, which is above the recommended minimum of 0.70. Convergent validity is demonstrated as a) the AVE (average variance extracted) values of all the constructs are higher than the suggested threshold value of 0.50, and b) all item-loadings are higher than the 0.70 guideline and statistically significant at the 0.001 level. Evidence of discriminant validity was found, since a) the

square root of all the AVEs was larger than the interconstruct correlations, and b) all the construct indicators loaded on their corresponding construct more strongly than on other constructs (see additional information file). Moreover, and the cross-loading differences were much higher than the suggested threshold of 0.1. For a variable to be a moderator, the variable should preferably have a low correlation with the predictor (independent) variable, since multicollinearity (r_{XZ}) can lead to researchers falsely concluding that a moderation effect is present, when a nonlinear effect in disguise is actually present (Baron & Kenny 1986). Carte and Russell [21] consider r_{XZ} ranging from 0.008 to 0.05 very low. In our study, the inter-correlations between CL and nCog, as well as CL and nAch are only 0.01 and 0.02 respectively (i.e. practically absent). This suggests that this error and result contamination are unlikely.

We evaluated the common method bias (CMB) using the exploratory method of Harman's one-factor test. The results from this test show that five factors are present, which explains 76.5% of the variance, while the most variance explained by one factor is only 37.7%, indicating that common method biases most likely did not contaminate the results. Furthermore, we applied a confirmatory method to analyze CMB using SmartPLS, as explained by Liang et al. [22]. We added a common method factor to the PLS model. The indicators of all the constructs were reflectively associated with the method factor. Thereafter, each indicator variance was computed to explain the principle construct and the method factor. The results (see additional information file) show that, while the indicators' average substantively explained variance is 0.818, common method-based variance is only 0.005. The ratio of substantive variance to method variance is about 167:1. Owing to the above evidence and the method variance's small size, we maintain that common method bias is unlikely to be a significant concern for this study.

Structural model results. After the validation of the measurement model, the structural model was independently analyzed and the proposed relationships between the constructs were tested. Using a blindfolding approach, we measured the cross-validated communality and redundancy using a Stone and Geisser test. The Q^2 results of both cross-validated communality and redundancy were greater than 0, which suggests that the model has good predictive validity. A *post-hoc* power analysis using the software G*Power 2 resulted in a value greater than .80, which implies that our model can detect small effect sizes. Finally, we calculated our model's goodness of fit (GoF) as proposed by Tenenhaus et al. [23] and emphasized by Wetzels et al.[24], who define GoF as the square root of the product of AVE and R^2 . The application of this formula leads to a GoF of 0.48, which exceeds the cut-off value of 0.36 for the large effect size of the squared multiple correlations (R^2), as proposed by Cohen [25] and allows us to conclude that our model performs well. In assessing the PLS model, we examined the squared multiple correlations (R^2) for the endogenous latent variable. The structural paths were evaluated for their significance. Proposed relationships were considered supported if the corresponding path coefficients (β) had the proposed sign and were significant.

Table 1. PLS path analysis results' endogenous variable: resistant use

LV	PLS (Stage I) Main Effects	PLS (Stage II) Individual analysis of Moderation effects			PLS (Stage III) Total variance explained
		Model 1: <i>nAch*RA</i>	Model 2: <i>LC*nCog</i>	Model 3: <i>nCog*CL</i>	
RA	-.28*(.07)	-.25*	-.26*	-.26*	-.26*(.06)
CL	.15*(.03)	.15*	.16*	.16*	.16*(.04)
CA	-.29*(.08)	-.28*	-.29*	-.30*	-.28*(.08)
nAch*RA		.19* (.05)			.14*(.02)
nAch*CL			-.16* (.04)		-.07**(.01)
nCog*CL				-.14* (.03)	-.05***(.003)
R² of RU	0.31	0.34	0.34	0.33	0.35

Path coefficients with Effect size: (f^2) in parentheses; *** $p < 0.05$; ** $p < 0.01$, * $p < 0.001$; Effect size (f^2) using the F-test [25]

Results in the grey cells are used for evaluation and interpretation

To provide a deeper analysis, we calculated the effect size using the F-test, since this is the most common and widely accepted measure of effect size in tests of moderation. We used the difference between the squared multiple correlations to assess the overall effect size f^2 for the variables. Cohen [25] classifies effect sizes of 0.02, 0.5, and 0.35 as *small*, *medium*, and *large*.

We applied a three-stage approach based upon Chin et al.'s [26], as well as Carte and Russell's [21] guidelines and recommendations to estimate the model. In the first stage, we entered the main effects. The results indicate that, for the first stage model hypotheses, H1 ($\beta = .28$, $p < .001$), H2 ($\beta = .15$, $p < .001$), and H3 ($\beta = 0.29$, $p < .001$) meet the criteria of both statistical, as well as practical significance and explain 31% of the variance in the dependent variable RU. Amongst the three variables, compatibility is found to have the strongest effect size with $f^2 = .08$, followed by relative advantage ($f^2 = .07$), and complexity ($f^2 = .03$). In the second stage, for each moderation effect we estimate stand-alone models in the presence of the main effects. As hypothesized, the need for achievement positively moderates ($\beta = 0.19$, $p < .001$, $f^2 = .05$) the relationship between relative advantage and resistant use (H4). The need for achievement negatively moderates ($\beta = 0.16$, $p < .001$, $f^2 = .04$) the relationship between complexity and resistant use, thus weakening its effect on individuals with a high need for achievement (H5). Furthermore, the need for cognition negatively moderates ($\beta = 0.14$, $p < .001$, $f^2 = .03$) the relationship between complexity and resistant use, thus weakening its effects on individuals with a high need for cognition (H6). In the third stage, we included all the moderation effects in addition to the main effects, and find that, compared to Stage I, the overall R^2 increased by 4% from 31% to 35%, which is attributed to the moderation effects. Figure 1 shows the results of the PLS stage III structural model. This three-staged approach is more appropriate when the goal is to understand the impact of each moderation effect. When estimating all the

effects in a single model containing highly complex multiple two/three-way moderation effects, the path coefficients and the effect sizes become contaminated and uninterpretable at even with the slightest degree of multicollinearity, which is caused by the underlying product-indicator approach [21]. However, when the goal is the evaluation of the model's overall performance via R^2 , the inclusion of all the main and moderating effects, as done in the stage III, does not distort the interpretation.

To facilitate a better understanding of the moderation effects, we drew up an appropriate visualization of the results, following Cohen et al.'s [27] recommendation, and calculated a simple regression equations for the RU and CL at low (-1 SD) and high (1 SD) levels of the moderator variables nAch and nCog. The obtained regression lines for high, and low values of the moderator variable are then plotted to determine whether there is an effect. We find that, compared to people who have a low nAch, i) an increase in the PMM's relative advantage has a weaker effect on suppressing the resistant usage behavior of people who have a high nAch, and ii) an increase in the PMM's complexity has a weaker effect on suppressing the resistant usage behavior of individuals who have high nAch and nCog.

In order to examine the heterogeneity in the data, we conducted *ex post* a permutation-based multigroup comparison in the framework of PLS path modeling. This approach is better suited to the PLS technique, since, in contrast to bootstrapping (t-test), permutation is non-parametric, i.e. it does not require the two samples to be normally distributed. The test was conducted for the variable "gender," which comprises the categories male and female. The results of the permutation test (see additional information file) show that, regarding Hypothesis 2 (CL \rightarrow RU), there is a significant difference between the path coefficients for males and females (Difference = .13, t-value = 2.83, $p < 0.01$). This implies that PPM complexity has a stronger effect on increasing resistant use in women ($\beta = 0.26$, $p < .001$), than in men ($\beta = 0.13$, $p < .001$).

5 Discussion and Conclusion

Our work seeks to further research on individuals' acceptance and use of PMMs by unifying the theoretical perspectives on the intrinsic *needs of individuals* and the *methodology attributes* within a single model. Based on validated theories, we develop a conceptual model, which maintains that individuals' psychological needs determine how a methodology's attributes impact their usage behavior.

This study's contributions and their implications lie in that, first, our use of the newly developed *resistant use* construct is a departure from traditional operationalization of the usage construct. It reveals more complex and still unknown interaction effects on human behavior, especially with regard to the use of new methodologies. Since researchers have generally relied on rather simple and straightforward ways to operationalize the use of IS artifacts and have linked it to a number of desirable outcomes, such as user satisfaction and productivity, the construct and the relationships have remained a black box. We still know very little about how IS artifacts are used and whether the differences in their usage style might be a better predictor of the nu-

merous proposed positive/negative effects of IS use. It is thus becoming important to understand resistant employee behavior because, if people do not use methodologies in an appropriate, committed manner, related benefits might not be realized. This lack of understanding might unjustifiably lead to IT becoming the “scapegoat” for organizational failures – for example, a lack of IT system effectiveness could be attributed to the bad system rather than employees' counterproductive resistant system usage behavior – casting doubt on the contribution of the MIS domain in question. By adopting a methodology-centric perspective, our findings suggest that, while a methodology's *relative advantageousness* and its *compatibility* reduce resistant use, its *complexity* induces resistance. It is interesting to see that compatibility has a stronger resistance curbing effect than relative advantageousness as past research, driven largely by the use of the theory of planned behavior and the technology acceptance model (TAM), has mostly shown RA to be, by far, the most dominant predictor of general usage behavior.

Our study's second contribution lies in the development of a deeper, context-specific, and relevant understanding of the role of employees' deep rooted psychological needs play in determining the effect that methodology attributes have on resistant usage behavior. While past research has repeatedly discovered and discussed the harmful nature of complexity in various contexts and domains (e.g. information complexity, system complexity, website complexity, innovation complexity, job/task complexity, etc.), our study provides a different perspective. Our findings suggest that, while employees generally resist using complex methodologies, this is not true for everyone. Employees who are driven by a strong need for achievement and need for cognition are found to be positively motivated by the inherent complex characteristic of a methodology. For individuals with a high *nAch*, complexity fulfills their preference for success under conditions of competition. Complex methodologies allow high achievers to satisfy a need for self-actualization through accomplishments that others in their social environment find difficult to achieve, because, to them, easily attained success is not a genuine achievement. Similarly, individuals with a high need for cognition find complex methodologies to be an intellectual challenge. Complexity in methodology provides such employees with a platform to engage in effortful cognitive activities, evaluate ideas, and analyze problems and their solutions. It forces individuals to “think out of the box,” be open to experiences associated with unconventional thoughts, as well as to consider solving problems and thinking an end in itself. These results might prove to be instrumental for management in ensuring that employees use PMMs in the proper manner – in particular, with regard to staffing issues. It is plausible that employees with a *high* *nAch* and *nCog* will be less resistant to using a PMM when assigned to large projects, since such projects, involving a multitude of stakeholders, goals, deadlines, and deliverables are generally governed by comprehensive and complex methodologies. On the other hand, it would be advisable to assign smaller projects that are usually managed with simple, less complex methodologies to employees with a *low* *nAch* and *nCog*.

In general, our findings might not only have important implications for the MIS research community, but also for related fields. Human needs have always played a key role in organizational development, and the study is an attempt to “humanize” organi-

zational methodologies; that is, to enable organizations to be more responsive to human concerns when developing and implementing new methodologies. Only when we understand and acknowledge that a diverse list of actions and feelings are typical of human behavior, do we view the acceptance and use of methodologies as a complex process and realize that research needs a fresh perspective to understand the *nature* of use and its antecedents.

References

1. Russo, N.L., Hightower, R., Pearson, J.M.: The Failure of Methodologies to Meet the Needs of Current Development Environments. In: Proceedings of the British Computer Society's Annual Conference on Information System Methodologies, pp. 387–393. (1996)
2. Cicmil, S., Đorđević, Z., Zivanovic, S.: Understanding the adoption of project management in Serbian organizations: Insights from an exploratory study. *Project Management Journal* 40 (1), 88–98 (2009)
3. Vickers, M.H.: Information technology development methodologies. *Journal of Management Development* 18 (3), 255 (1999)
4. Hardgrave, B.C., Davis, F.D., Riemenschneider, C.K.: Investigating Determinants of Software Developers' Intentions to Follow Methodologies. *Journal of Management Information Systems* 20 (1), 123–151 (2003)
5. Riemenschneider, C.K., Hardgrave, B.C., Davis, F.D.: Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models. *IEEE Transactions on Software Engineering* 28 (12), 1135–1145 (2002)
6. Ajzen, I.: The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50 (2), 179–211 (1991)
7. Hardgrave, B.C., Johnson, R.A.: Toward an information systems development acceptance model: the case of object-oriented systems development. *IEEE Transactions on Engineering Management* 50 (3), 322–336 (2003)
8. Khalifa, M., Verner, J.M.: Drivers for Software Development Method Usage. *IEEE Transactions on Engineering Management* 47 (3), 360–369 (2000)
9. Warner, K.E.: The Need for Some Innovative Concepts of Innovation: An Examination of Research on the Diffusion of Innovations. *Policy Sciences* 5 (4), 433–451 (1974)
10. Rogers, E.M.: *Diffusion of Innovations* 5th Edition. Free Press (2003)
11. Murray, H.A.: *Explorations in Personality*. John Wiley & Sons Inc (1938)
12. Freud, S.: *The Psychopathology of Everyday Life*. W. W. Norton & Company (1990)
13. Tornatzky, L.G., Klein, K.J.: Innovation Characteristics and Innovation Adoption-Implementation: A Meta-Analysis of Findings. *IEEE Transactions on Engineering Management* EM-29 (1), 28–45 (1982)
14. Ryan, R.M., Deci, E.L.: Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist* 55 (1), 68–78 (2000)
15. Chigona, W., Licker, P.: Using Diffusion of Innovations Framework to Explain Communal Computing Facilities Adoption Among the Urban Poor. *Information Technologies & International Development* 4 (3), 57–73 (2008)
16. Hollenback, J.R., Williams, C.R., Klein, H.J.: An Empirical Examination of the Antecedents of Commitment to Difficult Goals. *Journal of Applied Psychology* 74 (1), 18–23 (1989)
17. Straub, D.W.: Validating Instruments in MIS Research. *MIS Quarterly* 13 (2), 147–169 (1989)

18. Henseler, J., Chin, W.W.: A Comparison of Approaches for the Analysis of Interaction Effects Between Latent Variables Using Partial Least Squares Path Modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 17 (1), 82–109 (2010)
19. Chin, W.W.: The Partial Least Squares Approach for Structural Equation Modeling. In: Marcoulides, G.A. (ed.): *Modern Methods for Business Research*. Lawrence Erlbaum Associates (1998)
20. Chin, W.W.: Issues and Opinion on Structural Equation Modeling. *MIS Quarterly* 22 (1), vii-xvi (1998)
21. Carte, T.A., Russell, C.J.: In Pursuit of Moderation: Nine Common Errors and Their Solutions. *MIS Quarterly* 27 (3), 479–501 (2003)
22. Liang, H., Saraf, N., Hu, Q., Xue, Y.: Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. *MIS Quarterly* 31 (1), 59–87 (2007)
23. Tenenhaus, M., Amato, S., Esposito Vinzi, V.: A global goodness-of-fit index for PLS structural equation modelling. In: *Proceedings of the XLII SIS Scientific Meeting*, pp. 739–742. CLEUP (2004)
24. Wetzels, M., Odekerken-Schroder, G., van Oppen, C.: Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration. *MIS Quarterly* 33 (1), 177–195 (2009)
25. Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Routledge Academic (1988)
26. Chin, W.W., Marcolin, B.L., Newsted, P.R.: A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research* 14 (2), 189–217 (2003)
27. Cohen, J., Cohen, P., West, S.G.: *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum Assoc Inc (2002)

Predicting The Disclosure of Personal Information on Social Networks: An Empirical Investigation

Thomas Buckel and Frédéric Thiesse

University of Würzburg, Chair of IS Engineering, Würzburg, Germany
{thomas.buckel, frederic.thiesse}@uni-wuerzburg.de

Abstract. The present study considers factors that motivate users of social networks to publish different types of privacy-related information to friends or even the public. In contrast to prior research, we do not limit our research scope to an individual's decision-making process (i.e., the formation of behavioral intentions) but also include actual behavior as observed among a group of real Facebook users. Our objective is to test to what extent existing theory is not only capable of explaining self-disclosure decisions but also to predict subsequent behavior. We test our model using a combination of structural equation modeling and logistic regression with questionnaire data and data collected from the Facebook platform. Our results indicate that the way self-disclosure was operationalized in prior research shows low predictive power, especially when compared to predictions based on simple questions regarding an individual's sensitivity to the disclosure of personal information.

Keywords: Social networks, privacy, self-disclosure, risk perception

1 Introduction

With more than 800 million active users, Facebook is the largest and most popular social network worldwide. More than 50 percent of its members log into the network every day in their personal profile; even the least active users interact at least once per month with the website. Around 100 billion interpersonal relations are currently maintained and more than 250 million photos are uploaded onto Facebook servers per day. Not least, the enormous success the company established is also reflected by its IPO in the range of US\$ 100 billion, which poses a new record in the history of Internet-based companies [14], [36].

However, the rise of Facebook was also associated with several discussions surrounding the impact that the publication of personal profiles on the network might have on the privacy of individuals and their perceptions and valuations of privacy in general. A particular phenomenon that has attracted the interest of researchers is the 'Privacy Paradox' [34], that is, the apparent gap between the personal attitude towards the protection of privacy on the Internet and the actual behavior of social network users [3]. Facebook itself argues that the growing popularity of social networks should be interpreted as an early sign of a long-term trend towards more openness

regarding personal information, which might eventually become the new social norm [15]. Indeed, in recent years some mindlessness among social network users could be observed regarding the disclosure of private information if the expected outcomes outweigh the potential dangers [28], [38].

Against this backdrop, the present study considers the factors that motivate users of social networks to publish different types of privacy-related information to friends or even the public. In recent years, a number of prior studies were presented that developed and tested models for explaining user behavior by the example of Facebook or other platforms. However, as we argue in the following, prior research was limited to the investigation of behavioral intentions, whereas actual user behavior was usually not observed. As a consequence, there remains a problematic gap in the literature with regard to the predictive power of these models, which we aim to fill. For this purpose, we build upon the model presented by Krasnova et al. (2010) and extend it by new elements related to privacy preferences and information disclosure [26]. We test our model using a combination of structural equation modeling and logistic regression using a sample of questionnaire data and data collected from the Facebook platform. Our results confirm the original results by Krasnova et al. (2010) to a large extent, but also indicate that the model's ability to predict actual behavior is rather low. This especially holds in comparison to predictions based on simple questions regarding an individual's sensitivity to the disclosure of particular types of personal information.

The remainder of the paper is organized as follows. The next section provides an overview of prior research on social networking and the disclosure of privacy-related information. Next, we develop our research model and formulate a set of testable hypotheses. The following two sections describe the data collection process and the results of the hypothesis tests. The paper closes with a discussion of our findings, implications for practice and theory, and limitations.

2 Related Work

A number of prior studies can be found in the existing body of literature, which deals with the apparent gap between privacy preferences and disclosure of privacy-related information on the Internet. Berendt et al. (2005) presented a study with 171 respondents, which compared the actual behavior of internet users to previous statements on various privacy aspects. The authors reported that the participants revealed significantly more information on the net than they were willing to share according to the survey. They conjecture that expected benefits from the disclosure of information might be the predominant reason for this phenomenon [3].

Acquisti et al. (2006) found that users are increasingly aware of privacy risks associated with social media. In their study more than 500 Facebook members were surveyed and 196 profiles were analyzed. The objective of the study was to compare the congruence of survey responses with the actual profiles. Among others, the respondents were asked whether a particular type of information had been published within their Facebook profile or not. The results indicate an 80% match. Notwithstanding

some confusion with regard to the the complexity of the Facebook privacy settings, most of the respondents were aware of their level of self-disclosure [1].

Dwyer et al. (2007) developed a model, which explains information disclosure and the emergence of relationships on Facebook and Myspace by the users' trust in the platform and its members. According to the answers given by the study participants, the more information is shared the more users trust the other parties. Moreover, the authors tried to predict this behavior by the factor 'Internet Privacy Concern', but the empirical results do not indicate any significant influence [13].

Krishnamurthy & Wills (2008) found that 55% to 90% of social network members make their profiles publicly available. At least 80% provide confirmed contacts with complete access to their personal profile. In addition, their results indicate a negative relation between the size of the network and the use of functions for limiting the public visibility of profiles [27].

Christofides et al. (2009) conducted a survey and came to the counterintuitive conclusion that the willingness to publish personal information may not be negatively correlated with the user's perceived control of their data. Both aspects were defined as independent behaviors, with the general tendency to disclose, self-esteem and trust in the platform being the strongest predictors for information control [10].

Another study by Debatin et al. (2009) confirmed the conjecture by Youn (2009) that the expected benefit of being a Facebook user outweighs the perceived risks [11], [38]. In the same year, Bonneau et al. (2009) studied the manifold options for extracting profile data from other users by circumventing the network's security mechanisms [4]. The security issues associated with third-party applications, which are integrated into Facebook, were also confirmed by Felt (2008) [16]. Data protection problems in social networks in general were discussed by Fung et al. (2010) [18].

The study by Hoy & Milne (2010) set the focus on the differences between men and women in the context of privacy risks on the Facebook platform. The authors found that female users are slightly more sensitive to privacy risks [21].

Krasnova et al. (2010) investigated the motivation for disclosing personal information. In their model, self-disclosure is traced back to the negative influence of perceived risk as well as the positive influence of enjoyment, convenience, and the desire for self-presentation and relationship building. The empirical results show strong support for their model with the independent variables explaining about half of the variance in the self-disclosure variable [26].

In sum it can be said that prior research on self-disclosure in social networks has contributed substantially to our understanding of the decision-making process that eventually leads to people publishing privacy-related information. However, as with many other studies following the behaviorist paradigm, the scope of the empirical analysis is all too often limited to the investigation of behavioral intentions without observing actual behavior. Although behavioral intention is known to be a strong predictor of behavior, there is an evident gap in the literature regarding the ability of prior research models to predict actual behavior of social network users.

3 Research Model

The research model underlying the present study is depicted in Figure 1. Following the model proposed by Krasnova et al. (2010), we posit that perceived risks associated with the disclosure of personal information exert a negative influence on self-disclosure of an individual. We assume that risk perception is to a large extent determined by perceived control, trust in the social network operator, and trust in other network members. Self-disclosure, again, is assumed to be influenced by convenience, enjoyment, self-presentation, and the opportunity to build relations with others [26]. We extend this original model by an additional construct, which reflects an individual's preferences regarding the disclosure of specific types of privacy-related information. We hypothesize that this additional construct acts as mediator between risk perception and self-disclosure. Furthermore, we assume that the three dependent variables in the model may allow for predicting actual disclosure of personal information on the social network.

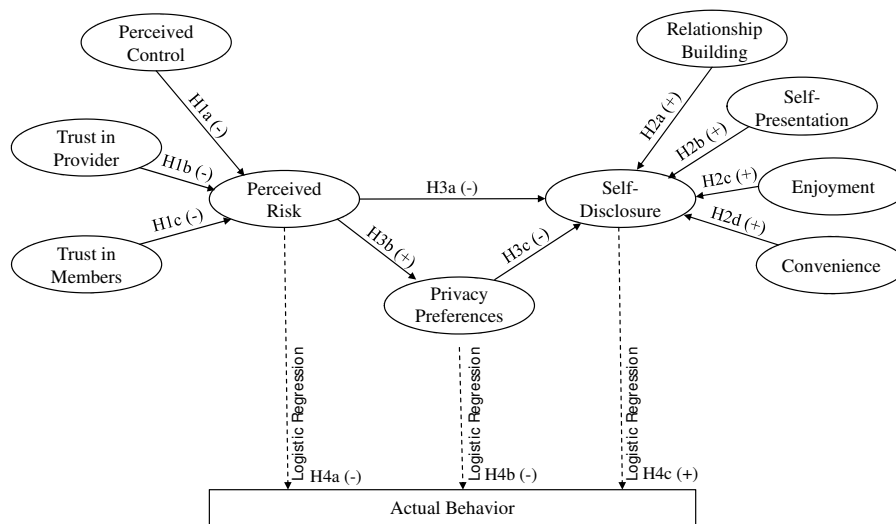


Fig. 1. Structural research model and hypotheses

Hypotheses H1a/b/c and H2a/b/c/d in our model were adapted from Krasnova et al. (2010). Due to space considerations, we refer the reader to this prior study for the underlying theoretical rationale. In the following, we limit ourselves to a mere enumeration of the hypotheses:

H1a: Perceived control exerts a negative influence on the perceived risk associated with the use of a social network platform.

H1b: Trust in other members of the network exerts a negative influence on the perceived risk associated with the use of a social network platform.

H1c: Trust in the operator of the social network exerts a negative influence on the perceived risk associated with the use of a social network platform.

H2a: Users' beliefs regarding relationship-building opportunities exert a positive influence on their self-disclosure on a social network platform.

H2b: Users' perceived benefits of self-presentation opportunities exert a positive influence on their self-disclosure on a social network platform.

H2c: Users' perceived enjoyment of platform use exerts a positive influence on their self-disclosure on a social network platform.

H2d: Users' beliefs regarding a network's ability to aid them in conveniently maintaining relationships exerts a positive influence on their self-disclosure on a social network platform.

The model by Krasnova et al. (2010) also posits a negative causal relation between the perceived risk and self-disclosure based on prior research, among others, by Malhotra (2004) [26], [29]:

H3a: Users' perceived risk associated with the use of the network exerts a negative influence on their intention to disclose privacy-related information.

An extension that we make to this original model refers to the modeling of risk beliefs and privacy preferences. While Krasnova et al. (2010) do not distinguish between the two and posit a direct effect of perceived risk on self-disclosure, we conjecture a more complex relation including a mediating variable. This view is supported by prior studies, which indicate that general risk beliefs materialize in the form of more specific concerns and/or preferences regarding the disclosure of personal information [12], [26]. These concerns/preferences may then have an impact on the intention to disclose information on the social network [12], [25], [38]. Consequently, we hypothesize that the relation between risk beliefs and self-disclosure might be mediated by an individual's privacy preferences regarding specific types of privacy-related information:

H3b: Perceived risks associated with the use of the network exert a positive influence on the formation of an individual's preferences regarding the disclosure of privacy-related information.

H3c: Privacy preferences regarding the disclosure of privacy-related information exert a negative influence on self-disclosure on the social network.

A second extension of the original model refers to our objective of investigating not only behavioral intentions but rather to predict actual behavior. Prior research shows inconsistencies between an individual's statements on self-disclosure and the information provided by the very same person in social media [1], [13]. Norberg et al. (2007) conclude that in the context of privacy research generalized constructs might be too coarse-grained to be applicable to the prediction of actual behavior [30]. We hence aim to test whether the two dependent variables from the original model as well as the new introduced privacy preferences pose predictors of actual behavior. The underlying rationale is that general constructs and the corresponding measurement scales are inferior to more specific questions regarding individual preferences when it comes to the prediction of self-disclosure behavior. If our assumption holds true, this might explain the apparent privacy paradox observed in prior research by a measurement issue rather than a theoretical gap.

H4a: Users' perceived risk associated with the use of the network exerts a negative influence on actual self-disclosure behavior.

H4b: Users' privacy preferences with regard to the disclosure of privacy-related information exert a negative influence on actual self-disclosure behavior.

H4c: Users' intention to self-disclose on the network exerts a positive influence on actual self-disclosure behavior.

Table 1. Questionnaire and construct operationalization

Construct	Items
Perceived Control (PC)	<ol style="list-style-type: none"> 1. I feel in control over the information I provide on Facebook 2. Privacy settings allow me to have full control over the information I provide 3. I feel in control of who can view my information on Facebook
Trust in Provider (TP)	<ol style="list-style-type: none"> 1. Facebook is open and receptive to the needs of its members 2. Facebook makes good-faith efforts to address most member concerns 3. Facebook is also interested in the well-being of its members, not just its own 4. Facebook is honest in its dealings with me 5. Facebook keeps its commitments to its members 6. Facebook is trustworthy
Trust in Members (TM)	<ol style="list-style-type: none"> 1. Other members will do their best to help me 2. Other members care about the well-being of others 3. Other members are open and receptive to the needs of each other 4. Other members are honest in dealing with each other 5. Other members keep their promises 6. Other members are trustworthy
Perceived Risk (RISK)	<ol style="list-style-type: none"> 1. Overall, I see no real threat to my privacy due to my presence on Facebook 2. I fear that something unpleasant can happen to me due to my presence on Facebook 3. Overall, I find it risky to publish my personal information on Facebook 4. Please rate your overall perception of privacy risk involved when using Facebook
Relationship Building (RB)	<ol style="list-style-type: none"> 1. Through Facebook I get connected to new people who share my interests 2. Facebook helps me to expand my network 3. I get to know new people through Facebook.
Self-Presentation (SPR)	<ol style="list-style-type: none"> 1. I try to make a good impression on others on Facebook 2. I try to present myself in a favourable way on Facebook 3. Facebook helps me to present my best sides to others
Enjoyment (EN)	<ol style="list-style-type: none"> 1. When I am bored I often login to Facebook 2. I find Facebook entertaining 3. I spend enjoyable and relaxing time on Facebook
Convenience (CON)	<ol style="list-style-type: none"> 1. Facebook is convenient to inform all my friends about my ongoing activities 2. Facebook allows me to save time when I want to share new stuff with my friends 3. I find Facebook efficient in sharing information with my friends
Self-Disclosure (SD)	<ol style="list-style-type: none"> 1. I have a comprehensive profile on Facebook 2. I find time to keep my Facebook-profile up-to-date 3. I keep my friends updated about what is going on in my life through Facebook 4. When I have something to say, I like to share it on Facebook
Privacy Preferences (PP)	1.- 21. How crucial do you rate the availability of your personal information ('Name', 'Home address', 'Hometown', 'Date of birth', ...) on the internet?

4 Data Collection

Constructs from the research model were operationalized using multi-item measurement scales with at least three items per construct. All questions were measured by 7-

point Likert scales (see Table 1). The corresponding scales were adapted from Krasnova et al. (2010) with the exception of the privacy preferences construct, which was modeled as a formative construct based on the different types of personal information included in a Facebook profile. The questionnaire was discussed with external experts and tested before the actual data collection with a group of students in order to ensure the comprehensibility of our questions. The questionnaire was prepared in the form of a PDF file, which avoids some issues with data quality associated with online surveys and allows the respondents to interrupt the process at any time and to continue later.

The data were collected from undergraduate and graduate students of business administration, economics and MIS. In total, 650 students were contacted. In addition to the questionnaire, we also asked respondents to accept 'friend requests' sent from our institute's Facebook profile. The latter step was necessary in order to access not only the respondents' public information but also information that is restricted to friends. In total, 182 students accepted our friend requests, among which 105 filled out the questionnaire. Five questionnaires were incomplete and had to be excluded from further analysis. As a result, we received 100 usable responses, which equals a response rate of 15.4%. The respondents' age was in the range of 17 to 30 years. On average, they were members of Facebook since 2009 and spend about 1.7 hours per day in the network. Genders were almost equally distributed with 51 male and 49 female respondents. The number of Facebook friends per respondent lies in the range of 150 to 300 persons. Most of the respondents were also members of other social networks.

5 Data Analysis

5.1 Structural Model Test

In a first step, we tested our research model excluding actual disclosure behavior using structural equation modeling techniques. For this reason, we followed the original procedure applied by Krasnova et al. (2010), who used the Partial Least Squares (PLS) method with Bootstrapping and Blindfolding [35]. A second reason for the use of PLS was the fact that our model included a formative construct, which cannot be modeled using covariance-based techniques [8]. Furthermore PLS is preferred instead of LISREL as the method is better suited for optimizing predictive power [6-7].

Validity and reliability of our scales were verified in a factor analysis. Bartlett's test of sphericity was accepted in all cases, which already indicates independent constructs. Items with factor loadings smaller than 0.5 were removed [23]. The number of items per factor was nevertheless three or greater for all factors. We continued with verifying the fit of the internal structure of our model by examining Cronbach's Alpha, composite reliability and the average variance extracted (AVE) per factor. The only exception was the 'Trust in Members' factor with an AVE of 0.46. Since the value is only slightly below the threshold and the factor did not contribute substantially to the overall explanatory power of the model by Krasnova et al. (2010), we decided not to exclude it. All factors surpass the widely accepted thresholds [2], [31]. Dis-

criminant validity was ensured by verifying that the square root of the respective AVE values was also larger than the correlations between the construct's validity [17].

We then calculated model fit indicators for the structural model. The PLS method does not offer a set of global fit indicators in the same way as covariance-based techniques. The fit of the model to the data can hence only be judged by R^2 and Q^2 values for the dependent variables. In our case, the R^2 values are greater than 0.2 for all relevant factors, which indicates substantial explanatory power [8]. Furthermore, we examined the Stone-Geisser criterion (Q^2), which indicates the predictive relevance [19], [33]. We consider both the cross-validated redundancy (Q^2 cvr) as well as the cross-validated communality (Q^2 cvc) [8-9], [24], [35]. Both types of Q^2 values are greater than 0 for all relevant factors. An overview of all considered factor and model criteria is given in Table 2.

The results of the actual hypothesis tests are given in Figure 2. In the following, we not only provide path coefficients and confidence levels, but also the f^2 statistic, which indicates the effect size [20], [32]. The test results show that PC exerts a significant negative influence on RISK (beta = -0.367, f^2 = 0.15, p < 0.001). In contrast, the influence of TP (beta = -0.186, f^2 = 0.04, p > 0.05) and TM (beta = -0.108, f^2 = 0.01, p > 0.05) is not significant. RB exerts a significant positive influence on SD (beta = 0.306, f^2 = 0.14, p < 0.001); the same holds for the influence of SPR (beta = 0.221, f^2 = 0.08, p < 0.01) and EN (beta = 0.229, f^2 = 0.06, p < 0.05) on SD. In contrast, the influence of CON on SD (beta = 0.070, f^2 = 0.00, p > 0.05) remains insignificant. In sum, our results are partly consistent with those presented by Krasnova et al. (2010), who were able to confirm a significant influence of TP on RISK. Furthermore, the results by Krasnova et al. (2010) indicate a significant influence of SP on SD, whereas the influence of CON was insignificant [26]. Our results show the opposite. We can thus confirm H1a and H2a/b/c, whereas H1b/c and H2d must be rejected.

Table 2. Validity, reliability, and model fit indicators

Factor	Type	Alpha	CR	AVE	R^2	Q^2 cvr	Q^2 cvc
<i>Threshold</i>		≥ 0.7	≥ 0.7	≥ 0.5	≥ 0.2	≥ 0	≥ 0
Perceived Control (PC)	R	0.88	0.93	0.81	-	-	0.59
Trust in Provider (TP)	R	0.86	0.88	0.55	-	-	0.37
Trust in Members (TM)	R	0.87	0.83	0.46	-	-	0.14
Perceived Risk (RISK)	R	0.77	0.85	0.59	0.23	0.12	0.31
Relationship Building (RB)	R	0.77	0.86	0.68	-	-	0.36
Self-Presentation (SPR)	R	0.89	0.93	0.82	-	-	0.61
Enjoyment (EN)	R	0.76	0.85	0.67	-	-	0.36
Convenience (CON)	R	0.72	0.82	0.61	-	-	0.24
Self-Disclosure (SD)	R	0.86	0.90	0.70	0.57	0.35	0.50
Privacy Preferences (PP)	F	-	-	-	0.26	0.04	0.14

Notes: R = reflective; F = formative

With regard to the interrelations between the dependent variables, we observed a significant positive influence of RISK on PP (beta = 0.541, f^2 = 0.37, p < 0.001). Furthermore, the negative influence of PP on SD (beta = -0.294, f^2 = 0.16, p < 0.05) was significant, too. In contrast, the results did not indicate a significant relation between

RISK and SD (beta = -0.041, $f^2 = 0.00$, $p > 0.05$), which can be explained by the inclusion of PP as mediator between the two factors. We can thus confirm hypotheses H3b/c, whereas H3a must be rejected. The R^2 value for SD shows that the model is able to explain 57.1% of the observed variance in this pivotal factor.

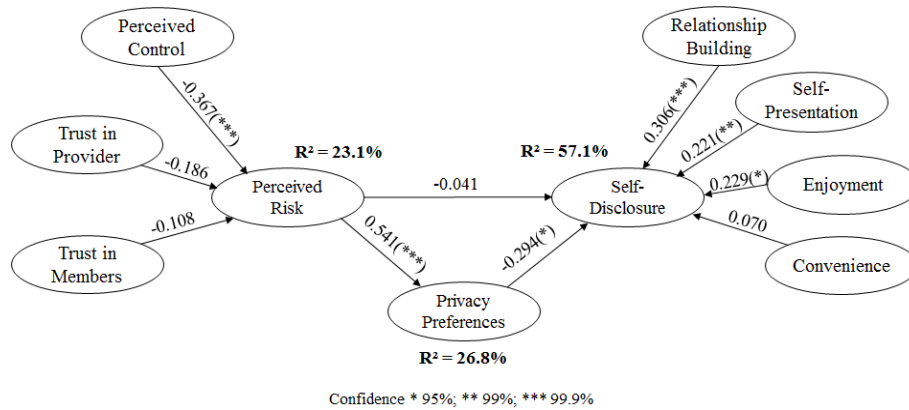


Fig. 2. Results of the Structure Model Test

5.2 Logistic Regression

In a second analysis step, we investigated to what extent the model is not only able to explain behavioral intentions (i.e., the variance observed in the SD construct), but also actual behavior. It is important to note that the explanatory power of the model measured by the R^2 value for SD should not be mistaken for the model's predictive power. For this reason, we considered actual behavior in the form of personal information published in an individual's Facebook profile. For each information type (e.g., name, address, religion), we investigated whether the dependent variables from the structural model act as predictors. Since actual behavior is measured by a set of binary variables, we use the logistic regression method for analysis. We consider both profile information that is publicly available as well as information that is restricted to friends. The occurrence of the 20 different information types on the Facebook profile pages of our survey respondents is summarized in Table 3. Note that 'Name' and 'Profile picture' are always publicly available. We considered only information types, which were included in at least 10% (average of public and friends profile) of the respondents' profiles. As an exceptional case, the mobile number was included though this information type did not reach the 10% threshold. The 'restriction ratio' denotes the proportion of people, who restricted a particular type of information to their friends.

Table 3. Occurrence of Facebook profile information among the survey respondents

Information type	Visibility	%	Restriction ratio	Information type	Visibility	%	Restriction ratio
Name	Friends / Public	74%	-	Current school	Friends	74%	43%
					Public	42%	
Home address	Friends	64%	47%	Former school	Friends	61%	51%
	Public	34%			Public	30%	
Home-town	Friends	53%	43%	Internships	Friends	2%	100%
	Public	30%			Public	0%	
Date of birth	Friends	83%	95%	Profile picture	Friends / Public	92%	-
	Public	4%					
E-Mail	Friends	70%	97%	Other pictures	Friends	84%	89%
	Public	2%			Public	9%	
Mobile number	Friends	4%	100%	Interests	Friends	86%	59%
	Public	0%			Public	35%	
Phone number	Friends	0%	-	Sports	Friends	39%	64%
	Public	0%			Public	14%	
Home-page	Friends	5%	80%	Politics	Friends	5%	80%
	Public	1%			Public	1%	
Marital status	Friends	60%	88%	Religion	Friends	9%	100%
	Public	7%			Public	0%	
Current profession	Friends	12%	42%	Relationships	Friends	37%	86%
	Public	7%			Public	5%	
Former profession	Friends	9%	67%	Avg. visible for friends: 42%		Avg. ratio: 74%	
	Public	3%		Avg. visible for public: 19%			

We developed separate logistic regression models for each type of profile information (= dependent variable) and the RISK, PP, and SD constructs from our structural model (= independent variables). In the case of the PP construct, we considered each item separately as predictor to the corresponding information type. In total, we investigated the prediction of 13 out of 21 information types among the restricted profiles and 10 out of 21 among the public profiles. The results of this analysis step are given in Table 4. We provide information on Nagelkerke's R^2 , the regression coefficients, and the corresponding significance value. Based on the R^2 value, we also determine which predictor works best for the respective information type. Our results show that the simple question for an individual's privacy preference works best as predictor. In contrast, the more general SD factor shows much lower R^2 values and, in some cases, exerts no statistically significant influence at all. The same holds for the RISK factor, which is the best predictor in only one case. In total, the logistic regression analysis indicates that the structural model, which performs reasonably well with regard to the explanation of behavioral intentions, is limited when it comes to the prediction of actual behavior. We therefore concluded that H4b can be confirmed, while H4a/c should rather be rejected.

Table 4. Logistic Regression Analysis Results

Independent Variable	Dependent Variable	Profile	Nagelkerkes R ²	Regression Coefficient	Sig.	Best Predictor
PP_Name RISK SD	Name	F/P F/P F/P	0.289 0.067 0.078	-0.616 (***) -0.576 (*) 0.633 (*)	0 0.037 0.029	PP
PP_HomeAddress RISK SD PP_HomeAddress RISK SD	Home address	F P	0.122 0 0.061 0.036 0 0.085	-0.354 (***) -0.037 0.505 (*) -0.185 0.005 0.577 (*)	0.004 0.873 0.042 0.109 0.983 0.015	PP SD
PP_Hometown RISK SD PP_Hometown RISK SD	Hometown	F P	0.159 0 0.137 0.043 0.004 0.077	-0.41 (***) 0.021 0.768 (*) -0.205 -0.126 0.557 (*)	0.001 0.926 0.002 0.083 0.604 0.02	PP SD
PP_DateOfBirth RISK SD	Date of birth	F	0.163 0.142 0.06	-0.523 (**) -0.986 (**) 0.611	0.003 0.007 0.075	PP
PP_Email RISK SD	E-Mail	F	0.117 0.092 0.001	-0.448 (**) -0.666 (*) 0.072	0.007 0.013 0.761	PP
PP_Handy RISK SD	Mobile number	F	0.185 0.008 0.056	-0.609 (***) -0.251 0.616	0.005 0.592 0.149	PP
PP_MaritalStatus RISK SD	Marital status	P	0.097 0.053 0	-0.46 (*) -0.647 0.04	0.05 0.157 0.925	PP
PP_CurSchool RISK SD PP_CurSchool RISK SD	Current school	F P	0.241 0.147 0.084 0.124 0.012 0.021	-0.605 (***) -0.905 (**) 0.661 (*) -0.409 (**) -0.218 0.276	0 0.003 0.024 0.004 0.34 0.211	PP PP
PP_ForSchool RISK SD PP_ForSchool RISK SD	Former school	F P	0.074 0.012 0.184 0.047 0.004 0.059	-0.294 (*) -0.215 0.969 (***) -0.255 -0.136 0.481 (*)	0.021 0.355 0.001 0.079 0.576 0.043	SD SD
PP_ProfPicture RISK SD	Profile picture	F/P	0.138 0.058 0.232	-0.595 (*) -0.696 1.988 (*)	0.027 0.132 0.013	SD
PP_OthPictures RISK SD PP_OthPictures RISK SD	Other pictures	F P	0.087 0.063 0.069 0.134 0.005 0.097	-0.496 (*) -0.623 0.675 -0.649 (*) -0.182 0.762 (*)	0.046 0.063 0.061 0.016 0.64 0.037	PP PP
PP_Interests RISK SD PP_Interests RISK SD	Interests	F P	0.062 0.100 0.007 0.130 0.001 0.015	-0.338 -0.836 (*) 0.199 -0.409 (**) -0.076 0.234	0.073 0.026 0.541 0.003 0.747 0.297	RISK PP
PP_Sports RISK SD PP_Sports RISK SD	Sports	F P	0.077 0.001 0.067 0.092 0.007 0.068	-0.303 (*) 0.065 0.505 (*) -0.428 (*) -0.199 0.588 (*)	0.019 0.778 0.028 0.034 0.536 0.051	PP PP
PP_Relationships RISK SD	Relationships	F	0.189 0.006 0.057	-0.477 (***) 0.151 0.464 (*)	0 0.517 0.043	PP

6 Implications for Theory and Practice

In this section, we compare the results from the structural model test to the results by Krasnova et al. (2010) and discuss our original results, which go beyond this piece of prior research. With regard to the interrelations between factors of the structural model, we found that neither the influence of trust in other members nor trust in the network operator showed a statistically significant influence on the users' risk perception. This phenomenon might be explained by the increasing understanding on the part users that the use of Facebook poses a privacy risk and that the platform operator makes money from the processing of personal information. The only factor that could be confirmed to exert a significant influence is perceived control. This might be attributed to the growing knowledge among users regarding the practical application of different privacy settings [5].

Similar to Krasnova et al. (2010), we were able to confirm significant positive influences of the ability to maintain relationships and enjoyment from Facebook usage on the intention to self-disclose. In contrast to this prior study, we also observed that the opportunity for self-presentation has a positive impact on self-disclosure, too. On the other hand, we could not confirm an influence of convenience, which had been identified by Krasnova et al. (2010) as a significant factor. Reasons for these different results might be that the ease of use of Facebook has become a standard among social network websites and that users' have become used to it. Moreover, the increasing awareness of the benefits of self-presentation might explain the impact of the users' interest in self-presentation on self-disclosure [5], [28]. Furthermore, we were able to confirm the relation between risk perception and self-disclosure. However, we extended the research model by an additional mediator variable, which reflects the users' preferences regarding the publication of private information. Our results indicate that privacy preferences are to a large extent shaped by the perceived risk and determine an individual's willingness to self-disclose. In contrast to that, the direct effect between perceived risk and self-disclosure did not turn out to be significant. In addition, we saw that our refined model shows a higher explanatory power with $R^2 = 0.571$ for the self-disclosure variable compared to 0.472 in the study by Krasnova et al. (2010).

In order to test the predictive power of the dependent variables with regard to actual behavior, we conducted a logistic regression analysis for each of the personal information types, which are part of any Facebook profile. The results show that the predictive power of the general factors from prior research is low and particularly inferior to simple questions regarding an individual's privacy preferences. This finding reflects a problematic assumption made in many behaviorist studies, which assume that actual behavior is in virtually any case determined by behavioral intentions and may hence be excluded from the respective investigation. In contrast, our results highlight the need for collecting data on actual behavior as well, even though the costs for acquiring such data are usually much higher than conducting a survey alone.

From a practitioner's perspective, the study also allows for a number of implications. Empirical studies of self-disclosure behavior are important means for judging the relevance of social network design decisions. Our results confirm some of the factors identified by prior research. We have seen that self-presentation and enjoy-

ment are important influence factors, which must be supported by any social network platform. We have also seen that perceived control exerts a significant influence on the perception of privacy risks. Social network operators should implement mechanisms that allow for fine-grained control of information disclosure. However, we could also show that investigations of general intentions and attitudes are not sufficient for analyzing the future behavior of users and that more specific measurement instruments are needed in order to capture their real privacy preferences.

7 Summary and Outlook

The purpose of the present study was to address a common weakness in many behaviorist studies considering human decision-making processes. Despite the fact that behavioral intentions are known to be strong predictors of actual behavior, researchers should not take this relation for granted and include data on their respondents' behavior in their analysis. We considered the specific example of self-disclosure on social networks, which has been the subject of a substantial number of prior studies. Based on our review of the literature, we found that the majority of existing studies relied only on survey data, whereas actual self-disclosure was usually ignored.

In order to fill this gap in the literature and to evaluate the predictive power of existing models, we extended and tested the research model by Krasnova et al. (2010), which showed high explanatory power with regard to the self-disclosure variable. We complemented the data set collected via a survey among 100 Facebook users by data on the same individuals' personal profiles on the social network. We evaluated the predictive power of the self-disclosure variable and compared it to a set of simple questions related to the respondents' privacy preferences. The results show that the factor measuring behavioral intentions is a rather weak predictor for actual behavior. This result could be interpreted as a sign of the so-called 'Privacy Paradox' phenomenon. However, our results also show the stated privacy preferences allow for much better predictions of the same data. This leads us to the conclusion that the Privacy Paradox may to some extent be attributed to a measurement issue in contrast to a more fundamental flaw in our theoretical understanding of self-disclosure.

Limitations of the study could be seen in the relatively small data sample, which we nevertheless consider sufficient for an analysis using PLS and logistic regression. Second, it should be noted that the respondents' survey answers and profiles might have been influenced by their participation in the study. The latter is a common issue with any study that depends on respondents who are aware of being observed. Further research will be necessary in two regards. On the one hand, researchers should strive to confirm our results using different and larger data samples. On the other hand, further research will be necessary in order to develop refined models and measurement scales to improve the explanatory and predictive power, for example, by introducing constructs such as 'social norm' [22], [37] into the respective models.

References

1. Acquisti, A., Gross, R.: Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In: Proceedings from Privacy Enhancing Technologies Workshop. Cambridge, UK (2006)
2. Barker, C., Pistrang, N., Elliott, R.: Reserach methods in clinical and counseling psychology. John Wiley, Chichester (1994)
3. Berendt, B., Günther, O., Spiekermann, S.: Privacy in E-Commerce: Stated Preferences vs. Actual Behavior. *Communication of the ACM* 48 (3), 101-106 (2005)
4. Bonneau, J., Anderson, J., Danezis, G.: Prying Data Out of a Social Network. In: Proceedings of the International Conference on Advances in Social Network Analysis and Mining, pp. 249-254, Washington, DC, USA (2009)
5. Boyd, D.: Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. *The International Journal of Research into New Media Technologies*, (14) 1, 13-20 (2008)
6. Chin, W.W., Marcolin, B.L., Newsted, P.R.: A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research* 14 (2), 189-217 (2003)
7. Chin, W.W., Newsted, P.: Structural Equation Modeling Analysis with small Samples using PLS. In: Hoyle, R.H.: *Statistical Methods for small sample research*. Sage Publications, Thousand Oaks et al. (1999)
8. Chin, W.W.: Issues and Opinion on Structural Equation Modeling. *MIS Quarterly* 22 (1), vii-xvi (1998)
9. Chin, W.W.: The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (eds.): *Modern methods for business research*. Lawrence Erlbaum, Mahwah (1998)
10. Christofides, E., Muise, A., Desmarais, S.: Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? *Cyberpsychology & Behavior* 12 (3), 341-345 (2009)
11. Debatin, B., Lovejoy, J., Horn, A., Hughes, B.N.: Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication* 15 (1), 83-108 (2009)
12. Dinev, T., Hart, P.: An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research* 17 (1), 61-80 (2006)
13. Dwyer, C., Hiltz, S.R., Passerini, K.: Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: Proceedings of the Thirteenth Americas Conference on Information Systems, Keystone (CO), Paper 339 (2007)
14. Facebook, <http://www.facebook.com>
15. Farber, A.: Facebook's Zuckerberg calls online openness the 'social norm.' *New Media Age*, 1/14/2010, 05 (2010)
16. Felt, A., Evans, D.: Privacy Protection for Social Networking Platforms. In: Proceedings of W2SP 2008: Web 2.0 Security and Privacy, pp 1-8, Oakland (CA) (2008)
17. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with unobservable variables and measurment error. *Journal of Marketing Research* 18 (2), 39-50 (1981)
18. Fung, B.C.M., Wang, K., Chen, R., Yu, P.S.: Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)* 42 (4), Article 14 (2010)
19. Geisser, S.: A Predictive Approach to the Random Effects Model. *Biometrika* 61, 101-107 (1974)

20. Henseler, J., Fassott, G., Dijkstra, T.K., Wilson, B.: Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. *European Journal of Information Systems* 21 (1), 99-112 (2012)
21. Hoy, M.G., Milne, G.: Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. *Journal of Interactive Advertising* 10 (2), 28-45 (2010)
22. Johnston, A.C., Warkentin, M.: Fear Appeals and Information Security Behaviors: An Empirical Study. *MIS Quarterly* 34 (3), 549-566 (2010)
23. Kaiser, H.F.: An index of factorial simplicity. *Psychometrika* 39 (1), 31-36 (1974)
24. Karim, J.: Emotional Labor and Psychological Distress: Testing the Mediator Role of Work-Family Conflict. *European Journal of Social Sciences* 11 (4), 584-598 (2009)
25. Krasnova, H., Kolesnikova, E., Günther, O.: It Won't happen to me!: Self-Disclosure in Online Social Networks. In: *Proceedings of the Fifteenth Americas Conference on Information Systems*, Paper 343, San Francisco (2009)
26. Krasnova, H., Spiekermann, S., Koroleva, K., Hildebrand, T.: Online social networks: why we disclose. *Journal of Information Technology* 25, 109-125 (2010)
27. Krishnamurthy, B., Wills, C.E.: Characterizing privacy in online social networks. In: *Proceedings of the Workshop on Online Social Networks in conjunction with ACM SIGCOMM Conference*, pp. 37-42, Seattle (WA) (2008)
28. Lucas, M.M., Borisov, N.: FlyByNight: mitigating the privacy risks of social networking. In: *Proceedings of the 7th ACM workshop on Privacy in the electronic society*, pp. 1-8, Alexandria (VA) (2010)
29. Malhorta, N.K., Kim, S.S., Agarwal, J.: Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15 (4), 336-355 (2004)
30. Norberg, P.A., Horne, D.R., Horne, D.A.: The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *The journal of consumer affairs* 41 (1), 100-126 (2007)
31. Nunnally, J.C., Bernstein, I.H.: *Psychometric Theory*. Mc Graw-Hill, New York (1994)
32. Ringle, C.M., Sarstedt, M., Straub D.W.: A Critical Look at the Use of PLS-SEM. *MIS Quarterly* 36 (1), iii-xiv (2012)
33. Stone, M.: Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2), 111-147 (1974)
34. Stutzman, F., Vitak, J., Ellison, N., Gray, R., Lampe, C.: Privacy in Interaction: Exploring Disclosure and Social Capital in Facebook. In: *International Conference on Weblogs and Social Media (ICWSM '12)*, Dublin, IE (2012)
35. Tenenhaus, M., Vinzi, V.E., Chatelin, Y., Lauro, C.: PLS path modeling. *Computational Statistics and Data Analysis* 48 (1), 159-205 (2005)
36. Townsend, A.: This Is Your Life (According to Your New Timeline). *Time International (South Pacific Edition)* 179 (6), 24-27 (2012)
37. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27 (3), 425-478 (2003)
38. Youn, S.: Determinants of Online Privacy Concern and Its Influence on Privacy Protection Behaviors Among Young Adolescents. *Journal of Consumer Affairs* 43 (3), 389-418 (2009)

Personal Information Markets AND Privacy: A New Model to Solve the Controversy

Alexander Novotny and Sarah Spiekermann

Vienna University of Economics and Business, Vienna, Austria
{alexander.novotny, sarah.spiekermann}@wu.ac.at

Abstract. From the early days of the information economy, personal data has been its most valuable asset. Despite data protection laws, companies trade personal information and often intrude on the privacy of individuals. As a result, consumers feel out of control and lose trust in electronic environments. Technologists and regulators are struggling to develop solutions that meet businesses' demand for more personal information while maintaining privacy. However, no promising proposals seem to be in sight. We propose a 3-tier personal information market model with privacy. In our model, clear roles, rights and obligations for all actors re-establish trust. The 'relationship space' enables data subjects and visible business partners to build trusting relationships. The 'service space' supports customer relationships with distributed information processing. The 'rich information space' enables anonymized information exchange. To transition to this model, we show how existing privacy-enhancing technologies and legal requirements can be integrated.

Keywords: informational privacy, personal data markets, privacy regulation

1 Introduction

The digital economy is in a deep crisis. From the inception of the digital economy, personal information (PI) has emerged as its core asset. PI is "any information relating to an *identified or identifiable* natural person" [1]. Abundantly leveraged as a free commons, PI is at the core of Internet economics and is considered the motor for online innovation. "Personal data is the new oil of the Internet and the new currency of the digital world" [2]. It finances the Internet's free content. It strengthens an Internet company's competitive stance. In many cases, it is even the only ingredient that brings an online service to life (e.g., social networking).

However, PI is also at the core of three facets that constitute humanity: Identity, dignity and privacy. And as PI is increasingly collected, used, packaged, and sold, more conflict arises around how people can retain control of their identities and protect their dignity and privacy. Under the umbrella terms "data protection" and "privacy" - the ability to control both the circulation of PI (out-flowing information) and the access of others to the self (in-flowing information) [3] - a global political debate has

emerged. This debate centers on whether people shall be enabled to control their PI and which aspects companies shall be allowed to use.

Unfortunately, the economic realities of personal data markets on one side and data protection efforts on the other are drifting apart. Companies capitalize on opportunities to collect and trade PI at an unprecedented scale. Uncontrolled PI trading has evolved [4]. Every time users surf online, an average of 56 parties track their activities on a website, largely without their consent or knowledge [5]. Companies claim “legitimate” business interests in the data they collect. The digital marketing association thinks that “marketing fuels the world” [6]. Major self-regulatory efforts of the industry, such as the Safe Harbor Agreement and the Do-Not-Track initiatives, are failing [6]. As a result, almost every regulatory privacy framework in the world (EU data protection directive 95/46/EC, Convention 108, OECD Data Protection Guidelines, US Bill of Rights Proposal, and more) is now being overhauled with the goal of strengthening consumer rights.

However, will regulation and self-regulation initiatives achieve what they say they aim for? With increasing business interest in personal information and an escalating conflict between privacy rights groups, regulators and industry, we believe that the time is ripe to develop a tenable vision of PI markets. This vision must allow for an innovative information-rich world while maintaining privacy. Fruitful streams of research and innovation depend on data about individuals. However, harm to human dignity and privacy must be avoided, and people must remain masters of their identities. What if we had digital markets that used and traded PI but let people control their information and identities?

Because of incongruous technical, economic and legal assumptions, it seems as if we are far from shaping such a future. Technology scholars have developed valuable privacy-enhancing technologies (PETs) that could put PI management back into consumers’ control [7-8]. However, their technical proposals often build on the assumption that people prefer anonymity in transactions with companies [9-10]. Consumers, in contrast, often don’t mind being identified in transactions with business partners, and companies are keen to foster ‘personal’ relationships [11]. While most PET proposals imply that consumers will invest time into privacy management, people simply expect regulators to protect them and companies to behave in an ethical way [12]. Finally, the PET community insists on terms such as “data minimization” [13], which are hardly realistic in times when users submit 95 million tweets on Twitter and send about 47 billion (non-spam) e-mails on an average day. The result is a patchwork of PET solutions that are adopted by neither industry nor governments.

Besides the difficulties to deploy easy to use PETS, economists disagree on the effects of privacy on welfare [14]. Chicago school proponents argue that PI disclosure benefits society because information asymmetries are reduced [15-16]: As companies learn more about their customers, they can serve customer preferences better. In contrast, critics contend that privacy protection generates social welfare [17]. Everyone acknowledges that people need control over the use of their PI [18-19]. But no consensus is reached yet on whether people should legally own their PI and get a property right [1], [19]. Many want to view privacy exclusively as a human rights issue because they are concerned that people could be ‘propertized’ [3], [20]. But giving peo-

ple control over their PI has driven human-rights based privacy regulation so far [18]. As a result, only a few scholars have theorized about how PI markets could be organized with privacy [3], [19], [21-22]. Where this has been done, models fail to consider the current technological landscape and legal environment.

This article makes an attempt to fill the visionary gap we need to make privacy efforts work in the economic environments we face. Based on insights about consumer behavior, market mechanisms, existing regulation and privacy technologies, we propose a 3-tier model for PI markets. Our model embraces information richness as the future of a digital economy. ‘Social data’ originating from people will inevitably be an important resource. We acknowledge that many transactions will be identified. However, the market we propose aims to empower people as much as companies. People and companies are assigned a few core rights and obligations resulting in a new and simple market structure. Many of these rights and obligations are already established; however, they are either weakly enforced or their importance is not recognized by policy makers. In our model, company obligations vis-à-vis consumers are enforced by the law and supported through privacy-enhancing technologies. The design of our model is guided by the principle of combining legal and technical enablers which mutually complement each other in asserting market rule enforcement. Our model is limited to the private commercial PI sphere, excluding government activity.

In the next section, we describe our vision of a functioning PI market in which privacy can be preserved and consumer trust in PI handling can be re-established. In the subsequent sections, this hypothetical market model is described in detail, including the derivation of technical and legal requirements to enforce it. The paper closes with a critical discussion of our model’s benefits and challenges.

2 A Three-tier Model for PI Markets

The model builds on the existing PI ecosystem. Currently, this system is complex and opaque, and its players engage in many secondary data use activities that undermine consumer privacy and trust [4]. We create transparency and simplicity by assigning existing players to a simple three-tier market structure (see Figure 1). The *first market tier*, which we call ‘relationship space’, includes the business relationship between data subjects and 1st tier partners. The *second market tier*, ‘service space’, includes the distributed computing and service infrastructures that enable today’s business relationships. It integrates all those processors who need to receive customers’ PI to directly enable and enrich 1st tier services. We distinguish between service delivery providers, which are necessary to perform the principal service, and service enhancement providers, which contribute to the 1st tier business relationship. The *third market tier*, ‘rich information space’ encompasses those players who do not directly support the 1st tier relationship. Participants in this part of the market can process as much data as they want, but the data they work on needs to be anonymized - to a degree that it cannot be linked with reasonable effort to 1st or 2nd tier transactions or data subjects. Each time PI is transferred to ‘rich information space’, it has to pass what we call the

‘anonymity frontier’. When information passes the frontier, it loses its ‘personal’ nature.

The stakeholders in our model are connected by contractual relationships. For any given relationship, market actors are unambiguously assigned to one of the three tiers. Table 1 summarizes the rights (Right 1-3) and obligations (Obl. 1-9) of all actors in our model. Usually, the data subject and 1st tier partner agree on a contract governing the exchange of service, compensation and PI. 1st tier partners arrange service-level agreements with service delivery and enhancement providers specifying the expected service quality. In exchange, service delivery and enhancement providers receive monetary compensation, or the right to use and sell anonymized information (AI). Market participants in the 3rd tier close sales contracts over AI with other actors.

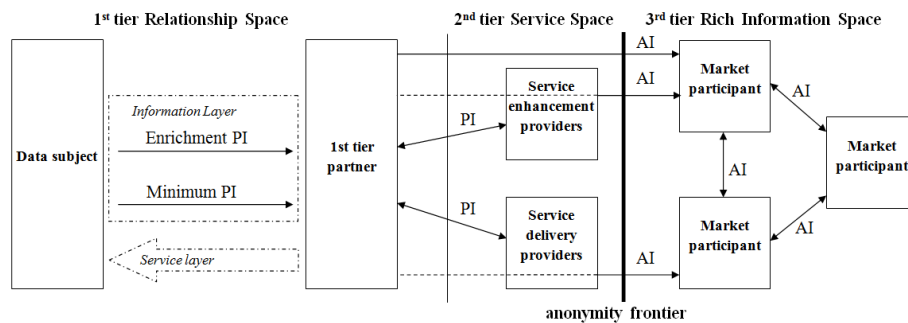


Fig. 1. Three-tier model for PI markets

2.1 The 1st Market Tier: Relationship Space

The 1st market tier is termed a “relationship space”: visible 1st tier partners maintain identified one-to-one relationships with their customers. All PI they receive is the recognized property of their customers and can be used only for purposes set down in PI usage policies, which accompany every PI exchange. The 1st tier is characterized by identified business relationships between users and one visible company, a separation of service and information exchange and the right to a privacy-friendly service, legitimized information collection, people’s property rights in their personal information and liability of the 1st tier partner for any PI abuse. The next paragraphs justify these characteristics from an economic and human rights perspective and point to their technical and legal implementation.

Identified Business Relationships and a Unique, Visible 1st Tier Partner. Because personalized customer relationships have proved effective, companies have invested in CRM solutions. Companies need and want identified customer relationships [11]. And many customers are willing to provide their PI in the service context if they receive appropriate benefits. Therefore, we depart from traditional data protection visions that promote the idea of total anonymity vis-à-vis companies [9].

However, users want predictable relationships in which they can control the use of their PI [23]. Predictability is supported when users are dealing with only one visible

PI-collecting business partner. We define *partner visibility* as a state in which data subjects visiting a physical or electronically-enabled premise can unambiguously and effortlessly name the commercial entity that they are transacting with. Customers in a physical retail store such as Waldepot see Waldepot as the 1st tier partner (and not, for example, the shelf suppliers).

Table 1. Actors in the three-tier model and their rights and obligations

Role	Definition	Rights			Obligations									
		1	2	3	1	2	3	4	5	6	7	8	9	
Data subject	Natural person disclosing PI in the course of a service transaction in a business relationship with the 1 st tier partner.	x	x											
1 st tier partner	Visible and primary opposite party in the service transaction and, from the viewpoint of the data subject, the party that is responsible for the PI.			x	x	x	x	x	x		x	x	x	
Collector	Party that gathers the PI from the data subject either by interrogation or observation.			(x)	x					x	x	x		
Controller	„Natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing“ of PI (Art 2 Directive 95/46/EC).			(x)	x					x	x	x	x	
Service delivery provider	Entity authorized by the 1 st tier partner that is necessary to perform the principal service.			(x)	x					x	x	x	x	
Service enhancement provider	Entity authorized by the 1 st tier partner that is not the service delivery provider but contributes by sufficiently close enrichment to the relationship between the 1 st tier partner and the data subject.			(x)	x					x	x	x	x	
Market participant	Any party including businesses, private persons, and governments who exchanges AI with other entities in the marketplace.			x									x	

All parties having a contractual agreement with data subjects need to be visible. Otherwise, they are not allowed to collect any PI through mechanisms such as cookies or uploaded software daemons. If data aggregators and brokers want to collect PI from users, they need to establish a distinct and visible relationship with data subjects.

The reason for this one-partner rule is that people lose control when multiple parties invisibly collect their PI at the same time. This loss of control promotes distrust on the web [23]. From a company perspective, the one-partner rule enables companies to regain the monopoly on PI collection in their transactions. This increases the power that they get from competitive information, as unrelated data traders will not have access to identified information anymore.

Selected technical enablers: identity and claim mechanisms; graphical user interface design. *Legal enablers:* use of standardized symbols for signaling 1st tier partner; mandatory ‘one-visible-partner’ rule; legal liability of 1st tier partner for PI use.

Separation of Service and Information Exchange and the Right to a Privacy-Friendly Service. Today, most online transactions are of a composite nature. Information is collected as a service spin-off [24] without making the ‘information deal’ visible to the customer. In our model, companies are asked to distinguish an information layer and a service layer within a business relationship. The service layer embraces the delivery of the principal service to the data subject, such as the sale and delivery of a book. Within the information layer, PI is split into necessary information for delivering a service (“minimum information”) and additional information that is used to enrich and enhance the service experience (“enrichment information”). Minimum information can be defined as the set of PI that is necessary and sufficient to perform the principal service. For the online book retailer, the minimum information is the name, delivery address and payment information. The individual’s purchase history, date of birth, and affinity profile, in contrast, is what we consider enrichment information. Additionally, partners are obliged to offer one service option (Obl. 3) that requires data subjects to disclose only the minimum amount of their PI. Thus, people always have the right to a privacy-friendly service (Right 2). This right repackages the existing concept of “data minimization” [13] but limits its scope to users preferring data minimal over information rich services.

Consider, for example, a web search engine, look.com, which offers three service options. Selected by default, the privacy-friendly option requires the individual to pay a subscription fee of, let’s say, € X per month; this option neither records search queries of the data subject nor shows any personalized ads. In contrast, the € Y priced second option collects more PI and uses it for an agreed time period to provide a richer service experience, such as individualized search results. The third option commercially leverages users’ PI for an agreed time period, such as for the targeted placement of ads. This option may be free. The user trades his or her PI in exchange for the free search service. The ‘free’ mentality governing online business relationships today would make room for a more realistic view of what digital services are actually worth. The result of such a separation of service options would benefit all market participants: Competition in the market for PI may be improved because the salience of the information transaction increases [24]. In addition to service quality, marketers could compete on PI usage rights and privacy. They could realize new revenue streams from privacy-friendly service options. And people would finally get a true choice of PI disclosure options.

A market challenge is that 1st tier partners could deliberately create opacity by providing myriad options, with variations on factors such as retention times or usage purposes for the PI. We therefore see the need for standardized PI usage policies that are adopted by regulators, at least for the privacy-friendly baseline offer (Obl. 4). *Technical enabler:* standards for the presentation of minimum PI service options. *Legal enabler:* mandatory separation of the service deal from the PI deal; obligation

to offer one service option with minimum information use at reasonable quality and price (Obl. 3); mandatory compliance with standardized privacy policies.

Legitimized Information Collection and Liability. The legitimization of data collection is probably the most important bridge between US American and European data protection frameworks [25]. Legitimization justifies the collection and use of PI. It can be obtained either through the active consent (Obl. 1) of a data subject or by legal empowerment; for example, mobile operators are legally required to preserve some connection data. Consent is a voluntary, timely agreement of the data subject to the 1st tier partner's PI terms, which should be explicitly communicated to the data subject [26].

Reconsider the search engine example from above: The default option would be the privacy-friendly version of a service. At one click, customers can explicitly opt into the free version. Whatever service option an individual chooses, all parties handling PI shall respect the agreement between data subjects and their 1st tier partners manifested in electronic PI usage policies (Obl. 2). Software agent solutions, such as P3P agents, enable people to initially configure their privacy preferences in their clients once (i.e., in the browser); for example, people might object to data processing for marketing purposes. A client-based architecture choice gives users more control over settings [27]. The user's software agent matches PI usage preferences with companies' standard usage policies (cf. 'Privacy Bird' presented in [7]) and supports the negotiation of an agreed PI usage policy. People are empowered to actively take advantage of their legal rights in every transaction and companies benefit from better data quality and compliance.

Moreover, the 1st tier partner is legally liable for any collection and use of PI. Liability safeguards the data subjects' property right and a contractually agreed PI usage policy. Liability of the 1st tier partner is natural from a customer perspective. The 1st tier partner acts as the single point of contact for the data subject. Most importantly, we envision that the 1st tier partner is responsible for implementing a technical accountability system that ensures that PI usage rights set down in electronic PI usage policies are obeyed (Obl. 5). Accountability ensures that any access, use, disclosure, alteration, and deletion of PI can be traced back to the party who has done so by using technical means. The 1st tier partner shall therefore have a technical infrastructure that can demonstrate PI usage rights to authorities and auditors at any time (Obl. 8).

Technical enablers: standards for the presentation and content of PI usage policies, software agent-supported privacy policy negotiation, use of an accountability system to enable and monitor policy-compliant use of PI (e.g., sticky policies, audit logs).

Legal enablers: 1st tier partners obtain legitimization for PI usage; handling of PI in accordance with electronic PI usage policies (Obl. 2); legal obligation to have and regularly audit an accountability system; 1st tier partner liable for all PI transactions.

Property Rights to Personal Information. A core component of our model is that data subjects have property rights for their PI (Right 1). The property right to PI cannot be alienated [1], [19]. Because of its personal rights character - similar to moral rights in copyright - seizing PI-related rights shall be prohibited. It is the characteristic

of identifiability that inseparably binds PI to an individual. Identifiable information can never be an object separable from a beholder, such as a book can be divided from its owner. However, usage rights to PI can be transferred. From a human rights perspective, data subjects have the biggest interest in the PI asset. Thus, they are the natural holders of this property right.

The main reason for proposing property rights to PI is a psychological one: Property rights would create stronger asset awareness in the minds of all stakeholders. The awareness that PI is an asset of economic value makes data subjects more informed when deciding about disclosing PI [28]. Equally, companies will probably be more cautious and reflective in collecting and using it. To make people aware of this asset, we must label information self-determination rights as “property rights”.

Technical enabler: policy repository on the client side. **Legal enabler:** recognition of a property right to PI (for an elaborate discussion of this proposal, see [18-19]).

2.2 The 2nd Market Tier: Service Space

Typically, the 1st tier partner is assisted by subcontractors, outsourcers, and strategic alliances to deliver services and products. This complex service web adds to the insecurity of today’s personal information markets. In fact, consumers are most concerned about secondary uses of their data by invisible partners [29]. For this reason, we create a ‘market chunk’ where this web of invisible service providers is organized. The 2nd tier includes all companies that contribute to the services delivered in the 1st tier.

PI abuses arise when parties at greater distance from the initial service perceive less responsibility for the PI they use. To extend the context-based trust between data subjects and 1st tier partners, 2nd tier service providers must be legally tied to the initial business relationships. This tie is created via a chain of accountability that ensures authorization, non-repudiation, separation, and auditability. Since all 2nd tier providers need to serve the 1st tier business relationship with the customer, our model ensures contextual integrity of PI use. PI is used within the boundaries of contextual integrity when the applicable social norms of appropriate PI collection and distribution are upheld in a given situation [30]. The following characteristics enable the 2nd tier:

Tying the Service Space to 1st Tier Relationships. We distinguish between service delivery and service enhancement providers (see Table 1). Service delivery providers such as parcel services delivering book orders are necessary to perform the principal service. They are always immediately involved in the 1st tier relationship and, for instance, include entities supporting the accountability and security of transactions. Service enhancement providers might also need to receive PI. These providers are parties that *directly* or *immediately contribute* to the 1st tier business relationship. For instance, they provide advertisements matching the interests of book purchasers. In case a data subject chooses such an enhanced service option, the service delivery providers can also handle enrichment information and service enhancement providers can process minimum information. **Technical enablers:** privacy policy language; accountability system to enable and monitor policy-compliant use of PI. **Legal enabler:** legal obligation to have and regularly audit an accountability system.

Authorization, Nonrepudiation, Separation and Auditability. For 2nd tier parties, an accountability system must comply with the requirements of authorization, nonrepudiation, separation, and auditability. First, authorization requires that access to PI by the service provider is approved by the 1st tier partner on an individual transaction basis (Obl. 6). When a customer purchases a book, the online shop must explicitly authorize a credit-scoring agency to use customer data for a credit check. Second, nonrepudiation prevents service providers from falsely denying that they have accessed, used, altered or deleted PI. Third, separation requires that PI units stemming from different service transactions, data subjects, and 1st tier partners are kept in strict isolation unless the legitimized purpose allows for the combination of PI (Obl. 7). This safeguards contextual integrity. Fourth, auditability ensures that compliance can be demonstrated at any time to authorities and auditors (Obl. 8). **Technical enabler:** use of an accountability system to monitor policy-compliant use of PI (e.g., sticky policies, audit logs). **Legal enabler:** separation of PI from multiple data subjects or 1st tier partners; legal obligation to and auditing of the accountability system.

2.3 The 3rd Market Tier: Rich Information Space

The 3rd tier is a market space where businesses, individuals, governments, and other parties not contributing to an identified business relationship freely exchange and trade information. They, however, need to ensure *anonymity* according to state-of-the-art technical standards. PI may originate from data subjects, but when the anonymity frontier is passed, this information becomes a *freely exchangeable* good. Innovation can be vividly spurred on the basis of this data. We assume that the marginal utility from identification outside of business relationships is so minimal that it does not justify the ensuing privacy risks. Severe *sanctions* should be imposed on 3rd tier market players who distort competition by holding identifiable or re-identifiable data.

Anonymity and Sanctions. Data subjects want to retain control over the distribution of their PI and want to share in good peace of mind. A straightforward way to create control and peace of mind is to legally enforce anonymity of all data except in situations where identification is needed or desired by the customer (1st and 2nd tier). People are granted a privacy commons, a shared space of anonymity [19]. In our model, this space is created by ensuring that PI cannot leave the contextual boundaries of the 1st and 2nd tier. When it does, it must be anonymized. What constitutes sufficient anonymization is a dynamic concept dependent on the current state-of-the-art of technology. Regulators should document and update current standards for anonymization in so called “BREF”s, best available techniques reference documents, which have been applied successfully for integrated pollution prevention and control (IPPC, Directive 2010/75/EU). Currently, the concepts of k-anonymity [10], l-diversity [31] and t-closeness [32] suggest that it is sufficient to have a large anonymity set of individuals, diverse attribute values and similar attribute value distributions. Each market participant in the 3rd tier is obliged to respect these anonymity mechanisms (Obl. 9) and is regularly audited for the fulfillment of this requirement. Finally, damages and

penalties for the illegal acquisition, possession, use or sale of identifiable information are necessary to protect a trustworthy market regime. **Technical enabler:** anonymization. **Legal enabler:** legal obligation and auditing of anonymity requirement in 3rd tier; sanctions for breaking the anonymity rule.

Free Exchange. Free trade of anonymized information increases the amount of exchanged information. Any market participant shall have free access to the 3rd tier market, including data subjects who may want to sell their anonymized information directly. Compensating for the costs 1st tier partners incur in our model, they have the right to anonymize and sell any PI collected independent of the data subjects' consent. Market participants can resell anonymized data once they acquire it (Right 3). **Technical enabler:** privacy-preserving data mining. **Legal enabler:** right to alienate AI.

3 Implementing the Three-tier Model

As has been outlined throughout Section 2, technical and legal enablers must support the implementation and enforcement of our model. Many of these technologies and legal enablers already exist. This section outlines how our model builds on these existing enablers and which need to be developed or changed.

3.1 Technical Enablers

Well-established privacy-enhancing and security technologies enable the enforcement of our model [33]. Table 2 gives an overview of selected technologies and assigns them to the relevant market tiers. To implement the requirement of accountability in the 1st and 2nd tiers, different systems based on sticky PI usage policies and audit logs are available [8], [34-35]. Most accountability systems suitable for ensuring contextual integrity are based on cryptographic technologies that can be easily applied in distributed environments [36-37]. Determining the responsible party for a data breach can be achieved by available identity technology. Existing security mechanisms, such as SAML, can identify the 1st tier partner and the data subject [38]. To specify the content of PI usage policies privacy policy languages are necessary [39]. Some privacy policy languages have already been standardized by the W3C consortium (P3P). Since negotiating these policies is a laborious and complex task for the data subject and 1st tier partner, architectures can make the task easier by employing software agents that semantically understand policy content [40]. The usability of privacy functionality and user agents at the interface between human individuals and machines is more and more improved [7]. Although data subjects are possibly identified on the application layer, they might want to be anonymous to third parties on the communication layer; to ensure their anonymity, data subjects can employ existing web anonymity technologies that protect the interaction between data subject and business partner [9]. Anonymity on the web can be supported by the “do not track” functionality that many web browsers will offer; this functionality indicates to the communication partner that no PI shall be collected. Additionally, anonymization technologies

are needed to realize sufficient anonymization of PI in the 3rd tier [10], [31-32]. Privacy preserving data mining technologies can guarantee endured anonymity [41].

Table 2. Assortment of existing technologies to support enforcement in the market tiers

Relationship Space (1 st tier)	Service Space (2 nd tier)	Rich Information Space (3 rd tier)
Accountability system Sticky policy, Privacy injector, Privacy-aware access control, Distributed auditing logs		Anonymization k-anonymity, l-diversity, t-closeness, graph anonymity
Identity mechanisms SAML, OAuth, OpenID		
Contextual integrity-compatible cryptography Identifier-based encryption, NOYB		
Privacy policy languages POL, PrimeLife policy language, E-P3P, EPAL, Rei, EnCoRe, PERFORM, Ponder, Contextual Integrity language		
Privacy policy negotiation P3P, PISA		Privacy-preserving data mining Randomization, Perturbation, Differential privacy, KD cycle-based data mining
Web anonymity and pseudonymity agents LPWA, Crowds, Hordes, Onion Routing, Mixminion		
Do Not Track		
Human-computer interface Privacy pictograms, User privacy agent interface design, Visual tagging		

3.2 Legal Enablers

Our model shall not only be technically feasible, but shall also be meaningful to public policy. Policy makers need to know which of the rights and obligations we propose already exist in the current legal framework. One important idea is to consider PI as the private property of data subjects [18]. A property right to PI (Right 1) is reflected in the principles of informed consent (Art 7 Directive 95/46/EC, Para. 7 OECD, Art 2 FTC Fair Information Practices (FIP)) and the right to object (Art 14 Dir. 95/46/EC). Missing is the recognition of full property rights to PI in civil law, however.

So far, a data subject's right to a privacy-friendly service (Right 2) exists only at a very limited scope. For example, Art 8 Directive 2002/58/EC mandates service providers to offer an option preventing the presentation of calling line identification. All of our model's other obligations already exist in legal frameworks. For instance, the obligation of anonymizing any information exchanged in the 3rd tier (Obl. 9) already exists to some extent in the principle of data quality (Art 6 Directive 95/46/EC). PI should "be kept in a form which permits identification of data subjects for no longer than is necessary [...]". To this vague formula, our model adds a clear anonymity frontier that unambiguously determines when anonymization takes place. Best available technique reference documents (BREF), kept current by data protection authorities, prescribe state-of-the-art anonymization technologies. Thus, only minor adaptations to the current legal framework are needed.

4 Discussion

Our vision for a personal information market establishes compromise between players in the current PI ecosystem and data protection proponents. Our model embraces the fact that data richness is the future of a digital economy and creates room for information-rich services and data trading as well as identified customer relationships. At the same time, our technical and legal suggestions empower people to participate in PI markets and protect their privacy. To help people understand their transactions with companies and the value of their PI, we create a new and simple market structure that assigns clear rights and obligations to all market players. Trust built by a clear allocation of rights also aids companies and legal enforcers.

We are aware that many of the rights, obligations and legal and technical enablers we propose are not new. They have been proposed for over two decades by researchers in privacy, identity, security, and legal studies and debated by companies and regulators. We do not need new security mechanisms which can, for instance, identify the 1st tier partner, but build on existing technologies which have been outlined in Section 3.1. However, no one has demonstrated how all of the puzzle pieces could be arranged in a market model to benefit both people *and* companies.

Personal information markets working to the benefit of people and the economy require that the enforcement of market rules is improved. The main design principle of our market model is to combine legal and technical mechanisms which mutually overcome its weaknesses. A legal property right to PI backed up by technical accountability of data usage simplifies law enforcement access for data subjects.

A sour apple that companies have to swallow is to finally provide people with a privacy-friendly default service option. But, as we show in this article, the apple isn't that sour. Companies can finally re-enter competition on the basis of service qualities. Furthermore, our model meets the privacy preferences of different individuals: Access to content at potentially lower cost for those who are willing to 'pay' with their PI and alternative versions for customers that are concerned about their privacy. Privacy rights proponents may argue that this preference-based market structure disadvantages the poor, who may be forced into selling their PI. This argument is true only if marketers choose to have people pay for the privacy-friendly version. Marketers could also make the data-rich version more attractive from a service perspective – with greater functionality and no ads – while offering a baseline service with non-personalized ads in a privacy-friendly way.

Finally, even if individuals opted into the usage of their PI in exchange for the service, our market proposal provides privacy protection: After all, companies would be accountable and liable for how they use PI. Limitless reuse and repackaging out of context would be outlawed. Privacy risks would hence be limited even for those who share. As data subjects will have property rights to their PI, they will also be brought back to the negotiating table. Property rights, a right to privacy-friendly service options and defaults, company accountability and a transparent market structure promise to re-establish the trust we need to see information services flourish.

A core benefit of our model is also its main technical challenge: the creation of a free market space that ensures anonymity. Ensuring anonymity becomes more diffi-

cult as technology becomes more powerful, facilitating identification. Anonymization could reduce the entropy of information to such an extent that the utility for information users would vanish. For multidimensional PI that contains many attributes about data subjects, the ‘curse of dimensionality’ forces that information is extensively aggregated to guarantee reasonable anonymity [41]. Utility-based privacy preservation, however, guarantees that the utility of anonymized data does not drop by more than a defined threshold ϵ , known as ϵ -differential privacy [42-43]. Data protection authorities define the “BAT” (Best Available Techniques) (Directive 2010/75/EU) that guarantee sufficient anonymity. Flourishing service spaces based on “non-identified, social data” instead of “personal data” may be the result. Information buyers want to obtain a representative sample of a population of individuals, not the information of identified single data subjects [41].

Finally, two more fundamental challenges of our model must be considered: concerns of ‘monopolizing’ information and the international enforceability of our model. The idea that personal data could be recognized as property originated in the US; this idea has been met by the criticism that people shouldn’t be ‘propertized’ [3], [20] as well as a series of other arguments (for an overview see [19]). Ralph Waldo Emerson once remarked, “As long as our civilization is essentially one of property, of fences, of exclusiveness, it will be mocked by delusions.” For these reasons, we view the idea of property rights to PI critically. However, because markets already treat PI as property, we ask only that people get the same rights that companies have already claimed for themselves. Moreover, a property right would not substitute, but rather enhance the human rights basis of privacy [18]. In Europe, it would provide people with an additional legal instrument, giving them easy access to existing, well-proven enforcement structures. Data subjects would be enabled to effectively claim their rights to PI on their own instead of calling on data protection authorities.

Another challenge of our model is its international practicability. Recent years have shown how difficult it is to reach international consensus on data protection or privacy. Even more difficult is enforcement. The Safe Harbor Agreement between the US and Europe on data handling practices is a good example of failure. A more effective path could be to implement and enforce binding hard-law for data protection. For example, property rights are enforceable as well-recognized legal instruments within both the European and US legal orders. If Europe and the US applied property rights to PI [25], the rest of the world would potentially follow suit.

Acknowledgement

We would like to thank Julian Cantella for the editing of this paper.

References

1. Bergelson, V.: It's Personal But Is It Mine? Toward Property Rights in Personal Information. UC Davis L. Rev. 37, 379 (2003)
2. Personal Data: The Emergence of a New Asset Class, World Economic Forum, Jan (2011)

3. Noam, E.M.: Privacy and Self-Regulation: Markets for Electronic Privacy. In: Wellbery, B.S. (ed.): *Privacy and Self-Regulation in the Information Age*, pp. 21-33. NTIA (1997)
4. *Rethinking Personal Data: Strengthening Trust*, World Economic Forum, May (2012)
5. Angwin, J.: *Online Tracking Ramps Up - Popularity of User-Tailored Advertising Fuels Data Gathering on Browsing Habits*. WSJ, June 18, B1 (2012)
6. Bott, E.: *The Do Not Track Standard has Crossed into Crazy Territory*, <http://www.zdnet.com/the-do-not-track-standard-has-crossed-into-crazy-territory-7000005502/>
7. Cranor, L.F., Guduru, P., Arjula, M.: *User Interfaces for Privacy Agents*. *ACM Transactions on Computer-Human Interaction* 13, 135-178 (2006)
8. Karjoth, G., Schunter, M., Waidner, M.: *Privacy-Enabled Services for Enterprises*. In: 13th *International Workshop on DEXA*, pp. 483-487, Aix-en-Provence (2002)
9. Gritzalis, S.: *Enhancing Web Privacy and Anonymity in the Digital Era*. *IMCS* 12, 255-287 (2004)
10. Sweeney, L.: *k-Anonymity: A Model for Protecting Privacy*. *IJUFKS* 10, 557 (2002)
11. Spiekermann, S., Dickinson, I., Günther, O., Reynolds, D.: *User Agents in E-commerce Environments: Industry vs. Consumer Perspectives on Data Exchange*. In: Eder, J., Missikoff, M. (eds.): *LNCS*, Vol. 2681, pp. 696-710. Springer, Berlin (2003)
12. *Personal Data in the Cloud: A Global Survey of Consumer Attitudes*, Fujitsu Res. Inst. (2010)
13. Borcea-Pfzmann, K., Pfzmann, A., Berg, M.: *Privacy 3.0 := Data Minimization + User Control + Contextual Integrity*. *IT* 53 (1), 34-40 (2011)
14. Acquisti, A.: *The Economics of Personal Data and the Economics of Privacy. 30 Years after the OECD Privacy Guidelines*. OECD (2010)
15. Posner, R.A.: *The Economics of Privacy*. *Am. Econ. Rev.* 71, 405-409 (1981)
16. Calzolari, G., Pavan, A.: *On the Optimality of Privacy in Sequential Contracting*. *Journal of Economic Theory* 130, 168-204 (2006)
17. Acquisti, A., Varian, H.R.: *Conditioning Prices on Purchase History*. *Marketing Science* 24, 367-381 (2005)
18. Purtova, N.: *Property Rights in Personal Data: a European Perspective*. Dissertation, Uitgeverij BOXPress, Oisterwijk (2011)
19. Schwartz, P.M.: *Property, Privacy, and Personal Data*. *Harv. L. Rev.* 117, 2056 (2003)
20. Cohen, J.E.: *Examined Lives: Informational Privacy and the Subject as Object*. *Stanford Law Review* 52, 1373-1437 (1999)
21. Laudon, K.C.: *Markets and Privacy*. *Communications of the ACM* 39, 92-104 (1996)
22. Aperia, C., Huberman, B.: *A Market for Unbiased Private Data: Paying Individuals According to their Privacy Attitudes*. HP Working Paper (2012)
23. Smith, H.J., Milberg, S.J., Burke, S.J.: *Information Privacy: Measuring Individuals' Concerns about Organizational Practices*. *MIS Quarterly* 20, 167-196 (1996)
24. Jentsch, N., Preibusch, S., Harasser, A.: *Study on Monetising Privacy: An Economic Model for Pricing Personal Information*. ENISA (2012)
25. Purtova, N.: *Property Rights in Personal Data: Learning from the American Discourse*. *CLSR* 25 (6), 507-521 (2009)
26. Art29WP: 01197/11/EN WP 187 - Opinion 15/2011 on the Definition of Consent, Article 29 Data Protection Working Party, Adopted on 13 July 2011 (2011)
27. Spiekermann, S., Cranor, L.F.: *Engineering Privacy*. *IEEE Transactions on Software Engineering* 35, 67-82 (2009)
28. Spiekermann, S., Korunovska, J., Bauer, C.: *Psychology of Ownership and Asset Defense: Why People Value their Personal Information Beyond Privacy*. In: *International Conference on Information Systems (ICIS 2012)*, Orlando, FL (2012)

29. Culnan, M.J.: "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MISQ* 17, 341-363 (1993)
30. Nissenbaum, H.: Privacy as Contextual Integrity. *Wash. L. Rev.* 79, 119 (2004)
31. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkatasubramanian, M.: l-diversity: Privacy Beyond k-anonymity. *TKDD* 1, 3 (2007)
32. Li, N., Li, T., Venkatasubramanian, S.: t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In: 23rd IEEE ICDE '07, pp. 106-115 (2007)
33. Shen, Y., Pearson, S.: Privacy Enhancing Technologies: A Review. Report HPL-2011-113, Hewlett-Packard Laboratories (2011)
34. Ringelstein, C., Staab, S.: DIALOG: Distributed Auditing Logs. In: IEEE ICWS '09, pp. 429-436 (2009)
35. Mont, M.C., Pearson, S., Bramhall, P.: Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services. In: 14th Int. Workshop on DEXA, pp. 377-382, Prague (2003)
36. Mont, M.C., Bramhall, P.: IBE Applied to Privacy and Identity Management. Technical Report HPL-2003-101. Hewlett-Packard Laboratories (2003)
37. Guha, S., Tang, K., Francis, P.: NOYB: Privacy in Online Social Networks. In: 1st Workshop on Online Social Networks, pp. 49-54. ACM, Seattle, WA (2008)
38. Recordon, D., Reed, D.: OpenID 2.0: a Platform for User-Centric Identity Management. In: 2nd ACM Workshop on DIM, pp. 11-16, Alexandria, VA (2006)
39. Berthold, S.: Towards a Formal Language for Privacy Options - Privacy and Identity Management for Life. In: Fischer-Hübner, S., Duquenoy, P., Hansen, M., Leenes, R., Zhang, G. (eds.): *Privacy and Identity 2010*, Vol. 352, pp. 27-40. Springer, Boston (2011)
40. The Platform for Privacy Preferences 1.1 Spec., W3C, 13 Nov (2006)
41. Aggarwal, C.C., Yu, P.S.: A General Survey of Privacy-Preserving Data Mining Models and Algorithms. In: Aggarwal, C.C., Yu, P.S. (eds.): *Privacy-Preserving Data Mining*, Vol. 34, pp. 11-52. Springer, New York (2008)
42. Ghosh, A., Roth, A.: Selling Privacy at Auction. In: 12th EC, pp. 199-208. ACM, San Jose (2011)
43. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating Noise to Sensitivity in Private Data Analysis Theory of Cryptography. In: Halevi, S., Rabin, T. (eds.): *TCC 2006*. LNCS, Vol. 3876, pp. 265-284. Springer, Berlin (2006)

Legal Literacy and Users' Awareness of Privacy, Data Protection and Copyright Legislation in the Web 2.0 Era

Katharina Steininger and David Rückel

Johannes Kepler University Linz, Department of Business Informatics – Information Engineering, Linz, Austria
{katharina.steininger,david.rueckel}@jku.at

Abstract. Web 2.0 offers an easy way for individuals to share any kind of content. The users' new role of being a content producer is associated with the responsibility to observe the relevant law. In this article, relevant regulations of the German and Austrian law are presented. In an attempt to understand the factors that influence users' lawful behavior, the presented research examines the effects of legal literacy, awareness and lawful attitude on lawful and unlawful actions. The empirical study shows results from a survey of 1,134 students from three different faculties. Regression and mediation analyses were used to analyze the effects. The results indicate that legal literacy has a negative direct influence on lawful behavior, while lawful attitude shows a positive effect among Austrian students. Furthermore, legal awareness has a stronger effect on lawful attitude than legal literacy.

Keywords: Legal Literacy, Data Protection, Copyright, Lawful Behavior, Mediation Model.

1 Introduction

In recent years, much research has been done to investigate behavioral antecedents of ethical Internet usage. Ethical boundaries often go along with legal permissiveness. While in the early years of commercial Internet usage most legal aspects were only relevant to editorial content producers, the rise of the web 2.0 concept shifts more legal responsibility to the common user. With the increasing amount of user generated content, web 2.0 users face current and forthcoming challenges concerning lawful behavior. Currently, there are several types of application and platforms that follow the principles of web 2.0 – most of them relevant regarding privacy, copyright, and data protection issues.

Copyright infringements, privacy violations, as well as extensive information disclosure have flourished in recent years. Copyright holders, especially the music and software industries, have responded with a series of lawsuits. Public and private initiatives have started to raise web 2.0 users' awareness on responsible and lawful usage of personal and third party information. The effects of those measures seem unclear [1]. Copyright infringements and invasion of privacy remain prevalent. Thus, to

explore ways to effectively approach the problems, it is essential to investigate the underlying factors associated with web 2.0 users' choice to engage in unlawful behavior. Prior research has focused on specific forms of unlawful behavior within those regulations, for example on music piracy [1-2], software piracy [3-5], or privacy issues [6-8]. An integrated research approach that addresses the problem on a more abstract level can help gain a more holistic view on factors influencing web 2.0 users' unlawful behavior.

The aim of this study was to identify the legal literacy of web 2.0 users as well as their awareness concerning certain legal domains - furthermore to find out about the lawful attitude arising from their knowledge and awareness that finally leads to lawful behavior. Therefore, hypotheses were derived from relevant literature and tested empirically. The data was collected from an online survey. Regression and mediation analyses were used to test the hypotheses.

The paper is organized as follows. Relevant literature concerning legal aspects on web 2.0 platforms, legal Internet literacy, legal awareness, and lawful attitude is reviewed and four hypotheses are derived in section 2. Next, the research model is presented and the methodology is described (section 3). In section 4, the results are analyzed. Finally, the findings are discussed along with future avenues for research in section 5.

2 Theoretical Considerations and Propositions

2.1 Legal Aspects on Web 2.0 Platforms

Even for private users, there are several legal and regulatory aspects that need to be considered when actively posting content on web 2.0 platforms. The main legal domains are

1. Data Protection;
2. Privacy Aspects and General Terms and Conditions;
3. Copyright.

Data Protection. The Austrian as well as any other European Data Protection Act is based on the extensively regulated European Data Protection Directive (Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data). Austrian Data Protection Law might not be applied in most cases of web 2.0 usage since the operating company of the platform is hardly ever based in Austria. Once the state in which an operating company is registered is part of the European Union, the home state regulation applies and the law of the country of residence is applicable. Since data protection regulations are highly harmonized within the European Union, the data protection principles of the Europe-

an Directive are applied.¹ Therefore, it can be assumed that the Data Protection Directive serves as minimum standard. If the operating companies are not registered in the European Union, data protection law of the home country of the user applies [9]. The main principles of the Data Protection Directive claim that personal data should not be processed except when certain conditions are met. One important exception is when the data subject agrees to the processing.

One main characteristic of web 2.0 platforms is the fact that a platform user is controller of the data entered to the platform and at the same time data subject concerning his personal data. The platform operator generally demands the right to access and process all user generated data for its own purpose. In that case the operator is seen as the controller and the user has to give an unambiguous, specific, and informed consent to the data processing. In most cases of web 2.0 platforms, such a consent does not exist or meet the legal demands [9-10]. Furthermore and strictly speaking, a web 2.0 user should obtain other users' consents when publishing their personal data on the platform (e.g. tagging a person on a photo).

Another major problem in web 2.0 is the common loss of control of data after a user deletes some of his content, the most famous social network Facebook is known for not removing deleted data such as status updates, photos, friend connections, or even whole profiles. That practice is not in accordance with Data Protection Law since the consent for the processing ceases to apply as soon as the data is deleted.

Finally there is an ongoing debate about the legal admissibility of social plugins by e.g. Facebook, Google+, or Twitter on private websites. The German data protection commissioner stated that the integration of social plugins on a website is not allowed without further ado if the exact form of data procession cannot be explained to the website's users [11]. At this time there is no Austrian regulation on the handling of social plugins, but they should be handled with care since it is normally not possible to obtain a precise and informed consent to the involved processing of personal data (e.g. IP address and information from cookies).

Privacy Aspects and General Terms and Conditions. In social networks one of the major issues is the disclosure of personal information. Especially young people tend to reveal private information on the Internet, predominantly on social network platforms like Facebook. Some research on such liberal disclosing behavior has tried to find out, whether the perception of possible risks is lower in young people and what benefits they expect from giving away that much personal information [12]. From a legal point of view there are two important aspects: (i) the privacy culture of the platform typically regulated in the platform's general terms and conditions and (ii) relevant legal clauses dealing with the right to privacy.

In its general terms and conditions a service provider defines the rules effective on his platform. Users need to accept them in order to take advantage of the service. The

¹ For more information see the results of the legal procedure against "Facebook Ireland Ltd." at the Irish Data Protection Commission, after an Austrian law student raised more than 20 Complaints because of the violation of European Data Protection Principles by Facebook Ireland. (www.europe-v-facebook.org).

platform operator decides on the privacy culture, the acceptable behavior towards information and data of other users and the extent to which users can change their own privacy settings on the platform.

Among the relevant legal clauses to protect privacy are Art. 8 ECHR (right to respect for private and family life) and regulations that protect a person against compromising exposure in a medium (e.g. newspaper, TV, or Internet), defamation of character, or libel of business reputation [9].

The function of tagging people on photos or videos offered in some social networks (e.g. Facebook) is legally questionable. In that procedure a user can upload an image and tag another Facebook user by unambiguously identifying him and automatically adding the new information (person identified on image) to his profile. Though the user concerned can delete the tag afterwards, he cannot prevent other users from adding a tag in the first place. The tagging function is legally dubious [13-14]. Adding the name to an uploaded image means revealing personal data of another person to the platform operators as well as other users and therefore taking an action relevant to data protection rules.

Copyright. While data protection issues are highly relevant in social network platforms, social sharing applications like Youtube or Flickr often have to deal with aspects of intellectual property rights. Content provided by users on such web 2.0 platforms (e.g. photos, videos, text) are subject to the copyright, which generally belongs to the original data's producers. An important aspect is the fact, that users often upload data without being the copyright owner. Such utilization of content without the approval of the copyright owner is illegal [15-16]. Concerning the further usage of such data, especially the download by private users, there are different regulations in Germany and Austria. In Germany the legal private copy is regulated in § 53 sec.1 UrhG. The regulation shall apply to private copies notwithstanding their format but only from legal sources. The amount of legal copies is limited to seven [17]. So the German law states that each download of an illegal or an illegally uploaded content is a copyright infringement since it is not possible to create a legal private copy from an illegal source [16]. In Austria, the utilization of content without the approval of the copyright owner is usually illegal as well. The legal private copy is regulated in § 42 sec. 4 UrhG. In the Austrian copyright law the explicit mention of the need for a legal source is missing. Therefore, there is an ongoing debate whether a legal private copy can be made by downloading from an illegally uploaded source (e.g. film or sound file) [9], [18-19]. Obviously the Austrian legislative body consciously did not include the explicit regulation in the legal text. Furthermore, there is no elucidating final judgment by the Austrian Supreme Court. So currently, the download of digital content like film or music files (not software!) by individuals is not explicitly illegal in Austria. Practically, such cases are not prosecuted at the time.

Regarding the copyright of software, there are EU-based differences concerning the copying for private use. It is illegal to copy software for personal purposes or to pass on to relatives or close friends. According to EU regulations it is only allowed to make a backup copy of the software, but that copy cannot be used or passed on to

other persons [19]. For that reason it is illegal to copy and pass on operating systems or office software.

In social networks and social sharing platforms the sharing of photos and self-made videos showing other people in private situations is very common. The German law (§ 22 KunstUrhG) regulates that images of persons can only be taken and made available to the public with the consent of the person shown on the image. By comparison, the Austrian law (§ 78 UrhG) states that it is prohibited to make an image of any individual and publish it if it could affect their legitimate interests [9]. That does not mean that there is a general prohibition to make images of other people accessible, only to make them available to the public without their consent (Germany) or when personal interests are affected (Austria) is illegal. Being at a certain place at a certain time is not sufficient to meet the requirements of the defined legitimate interests, but when images are derogatory, vilified, implying untrue circumstances, or if they show very private situations the regulations of privacy rights apply [17].

2.2 Legal Internet Literacy

Internet users face challenges that may affect legally compliant online interactions. Challenges a typical web 2.0 user may encounter include reading and understanding a platform's general terms and conditions; dealing with privacy issues when entering personal data; knowing copyright conditions like owners and exploitation rights when uploading photos, videos, or texts; understanding data protection principles when revealing own or other people's personal information; and being aware of national legal practices concerning copyright infringements when downloading music, films or software. All these interactions require more than basic legal knowledge and competency to ensure that one is not incurring a penalty or compensation for damages.

Legal Internet literacy is a complex construct that is closely related to legal literacy. Legal literacy is claimed to be required for effective participation in modern society [20-21]. The opportunities to interact online are nearly unlimited and since the emergence of web 2.0 concepts and applications users can easily and without special skills take part in the production and manipulation of content. Along with these features comes the responsibility to know the regulatory framework and the legal boundaries of their action.

At present there is no standard definition for legal literacy. Originally, the term was used to describe an aspect of professional education in law. In that sense being legally literate meant to be able to read and write legal arguments, judgments and legislations that are part of the body of law [22]. Later, a broader meaning of legal literacy was established. White defined it as the "degree of competence in legal discourse required for meaningful and active life in our increasingly legalistic and litigious culture" [21]. Several other definitions emphasize the broader meaning of legal literacy that expands from the professional legal practitioner into the society. The Canadian Bar Association brought this idea to the point and defines legal literacy as "the ability to understand words used in a legal context, to draw conclusions from them, and to use those conclusions to take action" [23].

The present study defines legal Internet literacy as the ability of average skilled web 2.0 users to read, understand, and interpret legal text in order to align one's actions with these statutory specifications. The better the understanding of legal regulations on actions often associated with the use of web 2.0 platforms is, the higher the chance is to avoid infringements of data protection or copyright laws [24-26].

The above considerations suggest that there is a positive relationship between legal Internet literacy and intended lawful behavior.

In the theory of reasoned action (TRA) Azjen and Fishbein do not explicitly mention knowledge or literacy as a component but they argue that attitudes are a function of beliefs [27]. Those beliefs refer to beliefs about the consequences of certain behavior (e.g. prosecution) and are therefore closely related to knowledge and literacy [1]. Other studies dealing with environmental attitude found out, that changing the knowledge and beliefs (e.g. by increasing literacy) also changed the attitude [28-30]. Hence, increasing legal Internet literacy and therefore increasing knowledge about what is legal and what are the consequences of illegal behavior will lead to more lawful attitudes.

Based on the considerations outlined above, there is a positive relationship between legal Internet literacy and lawful attitude.

Hypothesis 1: Legal Internet literacy (LIT) has a positive effect on lawful behavior (BEH).

Hypothesis 2: Legal Internet literacy has a positive effect on lawful attitude (ATT).

2.3 Legal Awareness

The ability to read, understand, and interpret legal texts in order to act in a lawful way is one important factor that can have influence on a person's attitude towards lawful behavior. But legal literacy is only one aspect when it comes to the evaluation of potential consequences of one's actions. An individual may be literate to read and understand relevant regulations, but unaware about the existence of specific laws or of the consequences a certain behavior might have in a certain situation [3]. As the awareness is increased, a person's attitude might change. In particular, an individual might be able to read copyright laws, but the awareness of the specific consequences arising from copyright infringement might be low. The perceived prosecution risk and the risk of a lawsuit are important aspects of legal awareness and have been found to influence ethical decision making [2], [4], [31]. Any unlawful behavior on the web includes the user's risk of civil action by damaged companies or people as well as the legal prosecution for copyright or data protection infringements. Chiou et al. found out that the perceived prosecution risk, which is an important aspect of legal awareness, influences the attitude of unauthorized music downloads negatively. The authors suggested to promote more aggressive publicity of possible prosecution and to exemplify prosecution cases in order to heighten awareness of the risks associated with unlawful behavior [2].

As the awareness of laws and potential prosecutions is increased, the positive attitude towards unlawful behavior should become less positive. Consequently, these considerations lead to the following hypothesis:

Hypothesis 3: Legal awareness (AW) has a positive effect on lawful attitude.

2.4 Lawful Attitude

Several studies have shown that attitudes are an important predictor of unethical behavior such as cheating, stealing, or lying (see e.g. [32]). Therefore, the TRA has been developed to explain these correlations. The theory has been used by many researchers to describe ethical decision making behavior [33]. As mentioned above, attitudes toward a behavior correlate with the beliefs about certain consequences arising from the intended behavior. Therefore, attitude constitutes the sort and intensity of feelings one has for or against an object or behavior [1]. So attitude describes a phenomenon that combines a person's beliefs about the consequences of an intended behavior and the evaluation of these consequences. In particular, TRA predicts that if a person believes that consequences of a behavior are predominantly positive then the person's attitude towards behavior will be positive as well. In other words, if a person believes that lawful behavior on the Internet will have predominantly positive outcomes, he or she will have a positive attitude towards lawful behavior.

Both, the TRA as well as the theory on planned behavior (TPB) state that behavioral intention is determined by attitude. Several empirical studies confirm the significant relationship between attitudes and the intention to act [3], [34-35].

Therefore, the following hypothesis is suggested:

Hypothesis 4: Lawful Attitude has a positive effect on lawful behavior.

The research model is presented in Figure 1.

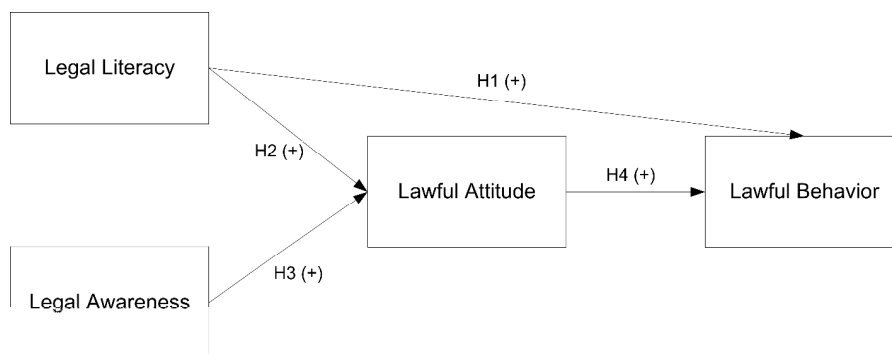


Fig. 1. Research Model

3 Research Method

The research hypotheses were tested empirically using data collected from an online survey. We chose a quantitative approach because a significant amount of research on the investigated constructs has already been conducted and several theories and hypotheses can be derived from prior work, as outlined in section 2. In contrast, a qualitative approach would be suitable for investigating topics that lack fundamental research and formal theories. Such nascent research areas are usually characterized by little knowledge on the research topic, a small number on prior work and the aim to engage in inductive theory development [36]. Since the research object of the presented study is lawful behavior on the Internet, people who do not use the Internet are not relevant to the presented study. Therefore, an online survey is a suitable instrument for the data collection and common critics on the usage of online questionnaires are not reasonable due to the target population. That is why the measure was a common choice because of the research subject (behavior on the Internet) [8]. The factors Legal Internet Literacy and Lawful Behavior were designed as manifest constructs to directly measure the knowledge of the respondents as well as their actual behavior instead of letting them rate the extent to which they believe to know facts or are likely to act. Therefore, test questions for legally relevant conditions were developed to measure Legal Internet Literacy. For Lawful Behavior legal and illegal situations were presented in the questionnaire and the students were asked to answer, whether they have already conducted these actions. For the analysis, only the illegal actions were taken into account. Legal Awareness and Lawful Attitude were defined as latent variables, measured with a five-point Likert scale and a two-level scale to express agreement/disagreement to a presented hypothetical situation. Table 1 presents a sample of four items for each construct.² For the following analysis for each construct the items were cumulated and transformed into interval scaled percentage rates, which indicated if an answer was correct (LIT); an action was legal (BEH); high awareness was shown (AW); or an intention was in accordance with the applicable law.

Regression and mediation analyses [37] were used to test the hypotheses. SPSS software was used for the regression analysis. The macro offered by Hayes to estimate the path coefficients in multiple mediation analyses and generate bootstrap confidence intervals for total and specific effects was used to test H1, H2, and H3 [38].

The regression analysis determines the effect of an independent variable on a dependent variable. The concept of a mediation hypothesis model is to test how, and to what extent an independent variable affects the dependent variable through one or more potential intervening variables, called mediators. Therefore, a model is tested in which the independent variable's effect can be separated into its direct effect on the dependent variable (without considering the mediator factor; c-path) and its modified effect on the dependent variable including the mediator (c'-path). Without the mediator, all of the paths would be quantified with the unstandardized regression coefficients and c-path is defined as the product of a and b. In mediation models the total

² See https://gw.ie.jku.at/dl/Construct_Item.pdf for the full list of items of the questionnaire per construct.

effect of the independent on the dependent variable can be expressed as the sum of the direct and indirect (= mediated) effects ($c = c' + ab$) [37].

Bootstrapping is a non-parametric procedure based on resampling with replacement of random samples of the data. (e.g. 1000 times). Each time, the indirect effect is computed. The mean of these indirect effects will not exactly equal the indirect effect and is therefore corrected [39]. In the presented study, bootstrapping is used to test the indirect effect from the mediation analysis.

Table 1. Constructs and item examples

Construct	Item
Legal Literacy	Total: 31
	<i>Indicate whether the statement is true or false:</i> Facebook does only unlink deleted user data. It is illegal to share mp3 files on file sharing networks.
Legal Awareness	Total: 6
	<i>Rate the extent to which you agree with the following:</i> I know the legal problems related to social media platforms. I am aware of the legal regulations related to copyright.
Lawful Attitude	Total: 11
	<i>Indicate whether you agree to the statement:</i> Copyright Infringement is a trivial offense. Copyright infringements need to be effectively prosecuted.
Lawful Behavior	Total: 8
	<i>Indicate whether you agree to the statement:</i> I have already uploaded mp3 files to a public platform. I have already given a copy of a copyrighted digital medium to a friend.

4 Results

4.1 Demographics and Correlations

The sample for this study is based on registered students from a university in Austria which is offering about 60 academic degree programs in the areas of (i) law (REWI), (ii) social sciences, economics and business (SOWI), and (iii) engineering and natural sciences (TNF). Although students are often criticized for not being representative for the population as a whole, they are a suitable target population in the presented study due to their age, because Austrian statistics show that people between 16 and 44 years are most likely to have Internet access [40]. Besides, one aim of the presented study is to investigate the effects of education on behavior. Students have a high educational background and are therefore suitable when investigating the effects of legal literacy on lawful behavior. Furthermore, similar samples have been examined to explain ethical behavior [41-43] and software piracy [44-45]. All of the nearly 18,000 students registered in January 2012 were asked to complete an online questionnaire, provided they named a valid e-mail address in the student administration system and

didn't opt out of the voluntary research information newsletter. A total of 1,624 students participated in the survey, 1,134 of them completed the questionnaire.

Table 2. Intercorrelations between control variables and constructs

	Mean	SD	Age	Gender	REWI	SOWI	TNF	Bachelor	Master	Diploma	PhD	Doctoral
Age	29.13	9.83	1									
Gender	0.51	--	-0.123***	1								
Faculty	REWI	0.27	--	0.155***	0.080**	1						
	SOWI	0.52	--	-0.036	0.136***	-0.629***	1					
	TNF	0.21	--	-0.123***	-0.253***	-0.314***	-0.541***	1				
	Bachelor	0.28	--	-0.219***	-0.018	-0.232***	0.073*	0.163***	1			
Level of studies	Master	0.12	--	0.080**	0.81**	-0.053	-0.045	0.112***	-0.232***	1		
	Diploma	0.51	--	0.025	0.114***	0.304***	-0.004	-0.324***	-0.636***	-0.376***	1	
	PhD	0.01	--	0.076*	-0.089**	-0.052	-0.074*	0.147***	-0.073*	-0.043	-0.117***	1
	Doctoral	0.08	--	0.194***	-0.047	-0.092**	-0.028	0.134***	-0.181***	-0.107***	-0.292***	-0.033
LIT	0.67	0.12	-0.054	-0.245***	-0.055	-0.065*	0.139***	-0.050	0.111***	-0.056	0.021	0.044
AW	0.68	0.14	0.068*	0.113***	0.138***	-0.052	-0.086**	-0.120***	0.037	0.068*	-0.046	0.050
ATT	0.52	0.13	0.204***	0.124***	0.168***	-0.092**	-0.070*	-0.077**	0.019	0.053	-0.052	0.030
BEH	0.80	0.18	0.068*	0.246***	0.155***	0.013	-0.183***	-0.030	-0.112***	0.108***	-0.017	-0.008

Note: n = 1134
* p < 0.05; ** p < 0.01; *** p < 0.001

Table 3. Intercorrelations between constructs

	Mean	SD	LIT	AW	ATT	BEH
LIT	0.67	0.12	1			
AW	0.68	0.14	0.059*	1		
ATT	0.52	0.13	-0.101**	0.267***	1	
BEH	0.80	0.18	-0.187***	0.204***	0.254***	1

Note: n = 1134
* p < 0.05; ** p < 0.01; *** p < 0.001

Table 2 and 3 show the means and standard deviations for the defined constructs (LIT, AW, ATT, BEH) and control variables (age, gender, faculty, level of studies). Gender, faculty, and level of studies are dichotomous variables and were 0-1 dummy coded with 1 indicating male / faculty or level of studies applied and 0 female / not applied. Age correlated significantly with three of the four focal variables (AW, ATT, BEH). Gender showed a significant correlation with all investigated constructs. Not all of the control variables representing the level of studies correlated significantly with the constructs. Only one of the three variables indicating faculty membership significantly correlated with all four constructs (TNF). Consequently, all control variables representing faculty and level of studies were excluded from further analyses. Since age and gender are prevalent control variables in research on effects on lawful behavior [46-48], both were not excluded. The Bravier-Pearson's correlation coefficient between LIT-ATT ($r = -0.101^{**}$), LIT-BEH ($r = -0.187^{***}$), AW-ATT ($r = 0.267^{***}$), and ATT-BEH ($r = 0.254^{***}$) were all significant. However, contrary to Hypothesis 1 and 2, the correlations between Legal Internet Literacy and Lawful Attitude (H2) as well as Lawful Behavior (H1) were negative rather than positive. Both positive correlation coefficients can generally be regarded as weak in size [32].

4.2 Regression and Mediation Model

Table 4 shows the results of the mediation analysis. As presented in the research model in figure 1, Legal Internet Literacy was defined as the independent variable, Lawful Behavior as the dependent variable, and Lawful Attitude as the mediator. Age and gender were included as control variables (CV). The adjusted R-squared of 0.1229 indicates that 12.3 percent of the effects on Lawful Behavior can be explained by the presented model.

Table 4. Results of the mediation analysis

	B	SE	Sig.	Adj. R-Sq	F	df	p
a-path (LIT-ATT)	-0.0613	0.0329	0.0625	0.1229	40.682	4.0 1129.0	< 0.001
b-path (ATT-BEH)	0.2833	0.0390	< 0.001				
c-path (LIT-BEH)	-0.1890	0.0441	< 0.001				
c' path (LIT-BEH)	-0.1716	0.0432	< 0.001				
CV1 (Age-BEH)	0.0008	0.0005	0.1347				
CV1 (Gender-BEH)	0.0720	0.0107	< 0.001				

Other than expected (H1), Legal Literacy had a significant negative (rather than positive) effect on Lawful Behavior (c-path). This effect decreased in size but remained significant (c' path) when the mediator variable, Lawful Attitude, was added to the prediction, which already indicates mediation. Furthermore, there was a significant *negative* effect of Legal Literacy on Lawful Attitude (a-path), which again was contrary to the hypothesis H2. However, in line with H4, Lawful Attitude showed a significant positive effect on Lawful Behavior (b-path). More importantly, the conducted bootstrap analysis revealed an indirect effect of LIT on BEH through ATT of $B = -0.0174$, $SE = 0.0097$. The 95% confidence interval (CI) did not include zero, 95% CI [-0.0387, -0.0001], which indicates that the indirect effect is significant at $p < .05$. This implies, that the negative effect of LIT on BEH can partially be explained by its effect on ATT.

The effect of Legal Awareness on Lawful Attitude (H3) was tested using linear regression analysis. The results showed a significant effect ($b = 0.231$ ($se = 0.027$), $\beta = 0.239$, $p < 0.001$; $adj. R-sq = 0.118$; $F(3,1130) = 51.43$, $p < 0.001$).

Table 5. Results of regression analysis (LIT/AW – ATT)

	β	Sig.	Adj. R-Sq	F	df	p
LIT-ATT	-0.080	0.006	0.1229	40.682	4.0 1129.0	< 0.001
AW-ATT	0.247	< 0.001				

Testing the comparative effects of Legal Awareness and Legal Literacy on Lawful Attitude, one can analyze which independent variable has higher effects on the dependent variable. As can be seen in table 5 Legal Awareness has a higher effect on

Lawful Behavior. This effect was positive and thus in line with H4. Figure 2 shows the research model with the non-standardized path coefficients

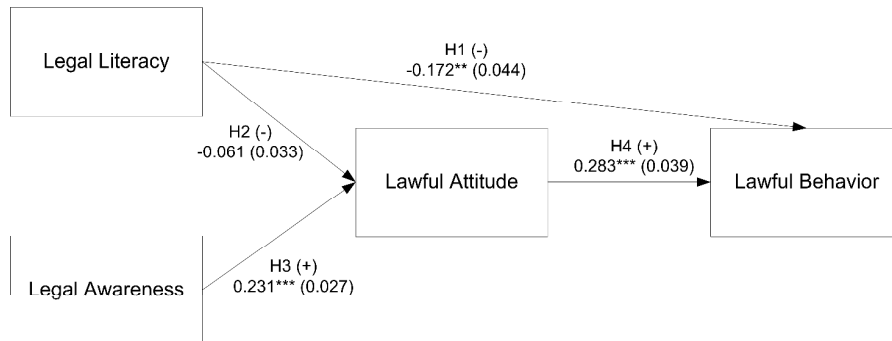


Fig. 2. Research model including results
(Numbers represent b-coefficient and standard error; ** p < 0.01; *** p < 0.001).

5 Discussion and Limitations

The purpose of this study was to better understand whether legal knowledge and awareness typically associated with Internet use are related to a lawful attitude and how all of these three antecedents are related to the actual behavior. The findings reported in the preceding section strongly support two of the four hypotheses. As shown in figure 2, the results of H3 and H4 are statistically significant at level 0.001, of H1 at level 0.01. The negative relationship between Legal Literacy and Lawful Behavior (-0.17) does not support H1 and therefore does not confirm previous findings [24-26]. Testing the literacy as manifest variable instead of letting the students rate the extent to which they believe they know legal regulations, might influence the results. As explained in the theoretical consideration section, there are some not explicitly regulated circumstances in the Austrian Internet laws. The knowledge about the legal limbos seems to come along with the ken of the specific regulations, since there is an ongoing legal debate in scientific and popular literature. Therefore, literate students seem to know that along with such unregulated aspects comes a tradition of lax prosecution. This explanation is supported by the fact, that within the presented survey students of technical programs, who have greater knowledge of the technical procedures on the Internet, are more likely to behave illegally although they are savvier when it comes to relevant legal clauses. Thus, in Austria individuals are even safer from civil prosecution compared to people who commit infringements for commercial purposes because of strict data protection clauses.

The results show no evidence that indicate a significant positive relationship between legal literacy and lawful attitude (as predicted in H2). While some research in other social science areas indicates that knowledge may positively influence the attitude [28-30], our field of research does not support that hypothesis.

As hypothesized in H3, legal awareness proves to be positively related to a lawful attitude. Indeed, observing and considerate Internet users are more likely to recognize legally relevant circumstances. Being more conscious of illegal actions and their possible negative consequences reduces the risk of Internet users to be seduced by the simplicity and deceived anonymity that comes along with illegal online actions.

There is a significant positive effect of lawful attitude and lawful behavior. This result supports H4 and is consistent with prior research on attitude and behavior [3], [34-35]. One important difference between the presented study and previous research is that earlier studies defined the dependent variable behavior as a latent construct representing the intended behavior. In the presented study, the behavioral aspect was measured directly using a manifest construct examining actual behavioral facts rather than intended behavior. So far as is known, this study is the first to measure the knowledge and behavior factors directly as manifest variables in the IS literature.

Due to the fact that the presented model explains 12.3 percent of the effects on lawful behavior, more influencing factors (e.g. experiences from past behavior, perceived anonymity) should be added. As with most empirical studies, the spectrum of respondents is a limitation. A statistically random sample would have increased confidence in the result or maybe have produced different outcomes. There is evidence that younger people are less aware of risks associated with Internet usage. Therefore, it could be of interest to investigate the model using a sample of considerably younger or older participants. Furthermore, because of the globalization of Internet platforms, it is important to shift the focus of the empirical study from the Austrian legal system and investigate the relationships in other, potentially stricter legal systems. Especially in case of data protection, an investigation in American societies might be interesting, since the European data protection regulations are much stricter.

References

1. Morton, N.: Understanding Attitudes toward Online Music Piracy. AMCIS 2004 (2004)
2. Chiou, J.-S., Huang, C., Lee, H.: The Antecedents of Music Piracy Attitudes and Intentions. *Journal of Business Ethics* 57, 161–174 (2005)
3. Goles, T., Jayatilaka, B., George, B., Parsons, L., Chambers, V., Taylor, D., Brune, R.: Softlifting: Exploring Determinants of Attitude. *Journal of Business Ethics* 77, 481–499 (2008)
4. Tan, B.: Understanding consumer ethical decision making with respect to purchase of pirated software. *Journal of Consumer Marketing* 19, 96–111 (2002)
5. Al-Rafee, S., Cronan, T.: Digital Piracy: Factors that Influence Attitude Toward Behavior. *Journal of Business Ethics* 63, 237–259 (2006)
6. Dinev, T., Hart, P.: Internet Privacy Concerns and Social Awareness as Determinants of Intention to Transact. *International Journal of Electronic Commerce* 10, 7–29 (2005)
7. Dinev, T., Hart, P.: Internet Privacy, Social Awareness, And Internet Technical Literacy. An Exploratory Investigation. *Proceedings BLED Conference 2004* (2004)
8. Tuunainen, V., Pitkänen, O., Hovi, M.: Users' Awareness of Privacy on Online Social Networking Sites – Case Facebook. *BLED 2009* (2009)
9. Burgstaller, P.: Soziale Netzwerke - Eine rechtliche Einführung. *lex:itec*. 16–22 (2012)

10. Europe-vs-Facebook: Complaint against Facebook Ireland Ltd. – 08 “Consent and Privacy Policy,” http://europe-v-facebook.org/Complaint_08_ConsentPrivacyPolicy.pdf, (2011)
11. Düsseldorf Kreis: Datenschutz in sozialen Netzwerken, http://www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/DuesseldorferKreis/08122011DSInSozialenNetzwerken.pdf?__blob=publicationFile (2012)
12. Palfrey, J.G., Gasser, U.: *Born digital: understanding the first generation of digital natives*. Basic Books, New York (2008)
13. Europe-vs-Facebook: Complaint against Facebook Ireland Ltd. – 02 “Shadow Profiles,” http://europe-v-facebook.org/Compalint_02_Shadow_Profiles.pdf (2011)
14. Europe-vs-Facebook: Complaint against Facebook Ireland Ltd. – 11 “Removal of Tags,” http://europe-v-facebook.org/Comlaint_11_RemovalOfTags.pdf (2011)
15. Wien, A.: *Internetrecht eine praxisorientierte Einführung*. Gabler, Wiesbaden (2009)
16. Völtz, G.: *Die Werkwiedergabe im Web 2.0 - Reformbedarf des urheberrechtlichen Öffentlichkeitsbegriffs* (2011)
17. Höhne, T., Jung, S., Koukal, A., Streit, G.: *Urheberrecht für die Praxis alles, was Sie wissen müssen. [Handbuch]*. Verl. Österreich, Wien (2011)
18. Kössler, M.: *Die Vervielfältigung zum privaten Gebrauch im Zusammenhang mit der Internetnutzung*. In: Aksam-Ratejczak, W., Stadler, A. (eds.): *Aktuelle Rechtsfragen der Internetnutzung*. Facultas, Wien (2011)
19. Sonntag, M.: *Einführung in das Internetrecht*. Linde, Wien (2010)
20. Bilder, M.S.: *The Lost Lawyers: Early American Legal Literates and Transatlantic Legal Culture*. Boston College Law School Faculty Papers (1999)
21. White, J.B.: *The Invisible Discourse of the Law: Reflections on Legal Literacy and General Education*. University of Colorado Law Review 143–159 (1983)
22. Zariski, A.: *What is legal literacy? Examining the concept and objectives of legal literacy.*, <http://www2.athabascau.ca/syllabi/lgst/LGST249sample.pdf>, (2011)
23. Canadian Bar Association. *Task Force on Legal Literacy: Report of the Canadian Bar Association Task Force on Legal Literacy*. The Association, Ottawa (1992)
24. Fillichio, S.: *Legal Literacy for Laypeople: What You - and Your Students - Should Know about Common Legal Situations*. *Update on Law-Related Education* 12, 34–38 (1988)
25. Correia, A.M.R., Teixeira, J.C.: *Information literacy: an integrated concept for a safer Internet*. *Online Information Review* 27, 311–320 (2003)
26. Livingstone, S.: *Internet Literacy: Young People’s Negotiation of New Online Opportunities*. In: McPherson, T. (ed.): *Digital Youth, Innovation, and the Unexpected*. 101–122 (2007)
27. Ajzen, I., Fishbein, M.: *Understanding attitudes and predicting social behavior*. Prentice-Hall, Englewood Cliffs, New York (1980)
28. Petzelka, P., Korsching, P.F., Malia, J.E.: *Farmers’ Attitudes and Behavior toward Sustainable Agriculture*. *Journal of Environmental Education* 28, 38–44 (1996)
29. Kuhlemeier, H., van den Bergh, H., Lagerweij, N.: *Environmental Knowledge, Attitudes, and Behavior in Dutch Secondary Education*. *Journal of Environmental Education* 30, 4–14 (1999)
30. Bradley, J.C., Waliczek, T.M., Zajicek, J.M.: *Relationship Between Environmental Knowledge and Environmental Attitude of High School Students*. *Journal of Environmental Education* 30, 17–21 (1999)
31. Fraedrich, J.P., Ferrell, O.C.: *The impact of perceived risk and moral philosophy type on ethical decision making in business organizations*. *Journal of Business Research* 24, 283–295 (1992)

32. Zimbardo, P.G.: The human choice: Individuation, reason, and order vs. deindividuation, impulse, and chaos. In: Arnold, W.J. (ed.): Nebraska Symposium on Motivation (1969)
33. Weber, J., Gillespie, J.: Differences in ethical beliefs, intentions, and behaviors. *Research in Ethical Issues in Organizations* 265–283 (2001)
34. Ajzen, I.: The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 179–211 (1991)
35. Sheppard, B.H., Hartwick, J., Warshaw, P.R.: The theory of reasoned action. *The Journal of Consumer Research* 15, 325–343 (1988)
36. Edmondson, A.N., McManus, S.: Methodological Fit in Management Field Research. *Academy of Management Review* 32 (4), 1155–1179 (2007)
37. Preacher, K.J., Hayes, A.F.: Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *BRM* 40, 879–891 (2008)
38. Hayes, A.F.: INDIRECT, <http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html>
39. Shrout, P.E., Bolger, N.: Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods* 7, 422–445 (2002)
40. Statistik Austria: IKT-Einsatz in Haushalten 2012, http://www.statistik.at/web_de/statistiken/informationsgesellschaft/ikt-einsatz_in_haushalten/index.html
41. Leonard, L.N.K., Cronan, T.: Illegal, Inappropriate, And Unethical Behavior In An Information Technology Context: A Study To Explain Influences. *Journal of the AIS*. 1 (2000)
42. Reiss, M.C., Mitra, K.: The Effects of Individual Difference Factors on the Acceptability of Ethical and Unethical Workplace Behaviors. *J. Bus. Ethics* 17, 1581–1593 (1998)
43. Kreie, J., Cronan, T.P.: Making ethical decisions. *Communications of the ACM* 43, 66–71 (2000)
44. Wagner, S.C., Sanders, G.L.: Considerations in Ethical Decision-Making and Software Piracy. *Journal of Business Ethics* 29, 161–167 (2001)
45. Kuo, F.-Y., Hsu, M.-H.: Development and Validation of Ethical Computer Self-Efficacy Measure: The Case of Softlifting. *Journal of Business Ethics* 32, 299–315 (2001)
46. Tittle, C.R., Ward, D.A., Grasmick, H.G.: Gender, Age, and Crime/Deviance: A Challenge to Self-Control Theory. *Journal of Research in Crime and Delinquency* 40, 426–453 (2003)
47. Meier, R.F., Weldon T. Johnson: Deterrence as Social Control: The Legal and Extralegal Production of Conformity. *American Sociological Review* 42, 292–304 (1977)
48. Moyano Dí az E.: Theory of planned behavior and pedestrians' intentions to violate traffic regulations. *Transportation Research Part F: Traffic Psychology and Behaviour* 5, 169–175 (2002)

A Household-Oriented Approach to The Benefits of Vehicle-To-Grid-Capable Electric Vehicles

Tobias Brandt, Stefan Feuerriegel, and Dirk Neumann

Albert-Ludwigs-University, Freiburg, Germany

{tobias.brandt, stefan.feuerriegel, dirk.neumann}@is.uni-freiburg.de

Abstract. In this paper we introduce a novel approach to exploring the benefits associated with Vehicle-to-Grid technology on an individual household level. We design an artifact that enables the implementation of different management strategies to utilize synergies between residential photovoltaic electricity generation and the energy storage provided by electric vehicles. The main advantage of this approach is that it does not rely on strong assumptions regarding the market penetration or social acceptance of electric vehicles. In a proof-of-concept case study we show that even a very simple management strategy derived from a household utility function provides additional revenues to the household, while simultaneously decreasing (peak) load on the distribution grid.

Keywords: Green IS, Energy Informatics, Electric Mobility, Energy Management, Sustainability

1 Introduction

The integration of intermittent renewable energy sources into the energy grid is one of the most pressing challenges many advanced nations face today. As the increasing scarcity of fossil fuels combined with a growing global demand for energy is likely to push resource prices upwards in the coming decades [1-2], societies are starting to embrace alternative means for generating electrical energy. This trend is reinforced by steadily decreasing production costs for solar panels and wind turbines and a growing awareness of man-made climate change.

The emergence of a new research frontier, Energy Informatics [3], caused scientists in information systems and computer science to consider their role in tackling the challenges associated with the power system of the 21st century. The central issue associated with an energy supply largely based on wind and solar power is its exogenously given intermittency, as energy is only generated when wind is blowing or the sun is shining. Out of several possible solutions to this problem, two stand out, both with their own inherent advantages and drawbacks, but both increasingly relying on information technology [4]. One is to align energy consumption with energy generation (demand-side management), thus flipping the traditional paradigm of adjusting energy generation to match demand. Often, this requires only small investments in infrastructure, mostly information technology for automation purposes, as it relies

primarily on rescheduling specific tasks. However, its limits are also evident – there is only so much work (both, in its colloquial and physical sense) that can be rescheduled before significantly encroaching on people’s lives. The second approach is the substantial expansion of energy storage systems. Most common methods of storing electrical energy rely on transforming it into other forms of energy that can be stored more easily, examples being chemical energy in the case of batteries and potential energy for pumped hydro storage. However, these systems have in common that they are to a varying degree quite expensive. Pumped hydro and any form of gas-based storage are also characterized by specific geologic requirements and, while they do not interfere with the day-to-day life of people, the former often requires significant alterations to landscapes and ecological habitats.

In recent years, electric vehicles (EVs) have been proposed as a possible – at least partial – solution to the problem of energy storage. While they do require electrical energy to function, creating their very own challenges for energy supply systems in the case of mass distribution, each one of them is supplied with its own battery. When equipped with technology that enables a two-way power flow – grid-to-vehicle and vehicle-to-grid (V2G) – they can serve as a swarm of small energy storage devices. The benefits and risks associated with V2G-capable EVs have since been the subject of several publications, which will be explored in Section 2. Most of them focus on large-scale utilization of EV fleets. Possible benefits of V2G-technology in the residential sector have been largely overlooked, since a single EV only offers a comparatively small amount of energy storage and is primarily used for mobility purposes.

This paper presents a novel perspective on the benefits of V2G-capable EVs in the residential housing sector. We introduce an energy management artifact that implements different strategies for coordinating energy generation from residential photovoltaic (PV) panels, household energy consumption, EV battery storage and mobility needs to derive additional benefits for the residents. We show that even the simplest management strategy can provide financial incentives for employing V2G-technology, while simultaneously decreasing peak demand and PV load on the distribution grid. This is especially relevant, since there are no assumptions about a widespread acquisition of EVs required – even a single household employing this management artifact can reap the benefit and take some burden off the power grid. Finally, we also shift a common approach in smart grid research: Instead of asking what is optimal from a technological point of view and subsequently analyzing how to “sell” this to the public, we first derive the optimum on an individual level and consider the aggregated effect on the grid following that.

This paper is structured as follows: In Section two we present publications related to our research and position our paper in this context. In Section three the requirements for the management artifact are analyzed, while the artifact itself is introduced within a case study in Section four. In Section five we discuss other possible management strategies and possible extensions to the underlying model. Section six concludes.

2 Related Work

Watson et al. [3] have stressed the importance of information systems and computer science in achieving sustainable solutions for a global economy. They emphasize that sustainable development supported by information technology goes beyond concepts like Green IS and Green IT and must include the role of these disciplines of shaping the power supply, transmission, and consumption systems of the future. Kossahl et al. [5] have since shown that Energy Informatics has gathered significant traction within the research community.

However, the steady rise of electric mobility as an alternative to traditional combustion engines has also caused research on energy and on mobility to become more and more intertwined. Battery-only Electric Vehicles (BEVs) and Plug-in-Hybrid Electric Vehicles (PHEVs) pose a promising option to address the problem of energy storage associated with the rise of intermittent renewable energy sources like wind and solar power. While a single vehicle can only store a comparatively small amount of energy (e.g. 16 kWh for the Mitsubishi i-MiEV or the Chevrolet Volt), the aggregated effect of a widespread adoption of EVs can provide substantial storage capabilities to the power grid. Beyond charging their batteries using the power grid, EVs require the ability to feed energy back into the grid – V2G-technology – to function as effective storage devices.

The engineering challenge associated with this technology has been addressed in several papers. Kempton and Tomic [6-7] summarize the technical and economic fundamentals related to V2G, while Cvetkovic et al. [8] extend this approach to include residential photovoltaic panels into the system design. The resulting energy system allows for smooth, uninterrupted transitions from grid-supplied to V2G-supplied energy provision for the households. Gurkaynak and Khaligh [9] design a residential photovoltaic control system to coordinate PHEV charging and regular residential requirements. On a larger scale, Lopes et al. [10] analyze grid-EV-interfaces to evaluate the effect of V2G-capable vehicles on the integration of wind power, showing that EVs can provide assistance in frequency control problems caused by intermittent energy sources.

The effects of a large number of EVs on the power grid have been the subject of a range of further publications, particularly in terms of adopted charging strategy. Lopes et al. [11] define a “dumb” strategy (charging whenever the vehicle owner desires to do so), a dual tariff strategy (lower energy price provides incentives to charge during night hours), and a “smart” strategy (charging centrally-controlled and grid-optimized), showing that the grid can only sustain a certain number of “dumb” EVs, but requires additional measures once the number of EVs increases. However, the financial incentives for vehicle owners to switch to a smart strategy are assumed to be given and not justified in detail. Similarly, Clement-Nyns et al. [12-13] illustrate the effect of an uncoordinated and a coordinated (dis)charging strategy of V2G-capable EVs on the power grid. However, these strategies are not examined on their popular incentive compatibility. Additionally, Flath et al. [14] have characterized protocols for smart charging from an Energy Informatics perspective.

A striking similarity in the literature on V2G-based business models is the reliance on EV-fleets or a widespread acquisition of EVs. This is justified in Guille and Gross [15], as well as Kempton and Tomic [7], with the superior revenues from entering the markets for ancillary services like frequency control. Most of these markets can only be entered when a certain amount of energy or power can be reliably provided, necessitating the aggregation of numerous EVs. Consequently, Kempton and Tomic [7] propose three business models distinguished by the aggregating player: (1) aggregation by the vehicle fleet owner, (2) aggregation by an electricity retail company, and (3) aggregation by a third party. The authors also identify the essential conflict between EVs as storage devices and their primary purpose, i.e. reliable mobility. White and Zhang [16] analyze the feasibility of an aggregated V2G program entering the markets for frequency regulation and peak-load reduction simultaneously. The authors conclude that there are additional payoffs for the individual participants of such a program, but do not propose a concrete contracting scheme to distribute these payoffs among the players in this scenario, thus not addressing the aforementioned essential conflict and the incentive compatibility of such a program. In a case study of PHEVs as providers of regulatory power in Germany and Sweden, Andersson et al. [17] simulate individual EVs as bidders in the regulatory power market in the respective countries. While this approach is one of the few to address individual vehicle owners, the authors mention a specific problem the implementation of such a project in a real-world setting would face. As the transmission system operators, who run the market for regulatory power, would be able to collect a massive amount of data on individual mobility, concerns about data protection could create substantial social resistance to such a V2G-program.

A broader analysis of barriers to a transition towards PHEVs and V2G is provided in Sovacool and Hirsh [18]. The authors argue that not only technical barriers are to be overcome, but also social, cultural and political hurdles. They highlight the basic acceptance problem of BEVs and PHEVs, the cultural aversion towards technological change and the interests of stakeholders in the current paradigm of mobility. Nevertheless, in an analysis on the long-term potential of V2G technology, Turton and Moura [19] conclude that V2G may cause a paradigm shift in both energy systems and transportation.

Summarizing, we conclude that V2G-technology offers an interesting perspective on managing future power systems in light of the rise of renewable energies. It would, however, also implicitly change the way we think about mobility and transportation. The majority of recent publications has analyzed the grid stabilizing effect of V2G-capable EVs. This approach relies heavily on a widespread adoption of electric mobility and only marginally considers popular acceptance of the coordination mechanisms, if at all. In this paper we present a novel approach to V2G research. Instead of optimizing on the aggregated level without regards to individual incentive structures, a control artifact implements different strategies to manage energy supply from residential PV installations and V2G-capable EVs on the household level. These strategies are derived from an individual utility function, thus addressing the incentive compatibility problem. While the aggregated effect on grid stability of such a household-level optimization may be inferior to other approaches in related works, it does

not rely on ambitious assumptions concerning the total number of EVs, popular acceptance or data security.

3 Requirements Analysis

The fundamental tradeoff associated with V2G for an individual car owner is between possible financial gains from using V2G-capabilities and unconstrained mobility. Subsequently, when constructing charging strategies, this tradeoff must be considered. In this paper we approach V2G in its fundamental sense as a technology complementary to renewable energy sources, and not primarily as a way to stabilize the power grid. The baseline scenario we analyze includes a single household, a residential photovoltaic panel and an EV. While houses with rooftop-PV panels have become common sights in many industrialized countries, there is still a low market penetration of EVs. Consequently, the notion of EV-fleets in the residential sector is still a long way off for the next years, perhaps even decades. Yet, the number of households that own both, a PV installation and an EV, is likely to steadily increase during that time, since both are currently targeted at the same type of customer (middle to upper income class, ecologically aware). By intelligently using V2G-technology, the owner can receive additional financial benefits from home-produced PV electricity without depending on a widespread adoption of EVs. This scenario is illustrated in Figure 1.

The basic idea is that the EV stores excess photovoltaic energy during times of low demand D , which would normally be fed into the grid (PV_G) and uses this energy to supply the household (EV_{HH}) during times of high demand, when otherwise energy would need to be procured from the distribution grid (G_{HH}). The often substantial price difference between energy fed into the grid and energy procured from the grid would then provide monetary incentives to the vehicle owner, as long as this gain is not offset by additional constraints on mobility. This allows us to formulate the fundamental tradeoff associated with V2G as follows:

$$U(t, M(t), s) = -C_E(t, s) + \phi(t, M(t), s) \quad (1)$$

The utility U of the household at time t with the mobility requirements $M(t)$ and under the V2G-strategy s is thus defined in Equation 1 as the degree to which $M(t)$ is satisfied under s , computed and translated into monetary terms by the satisfaction function ϕ , minus the total cost of energy procurement C_E in t under strategy s . While there are

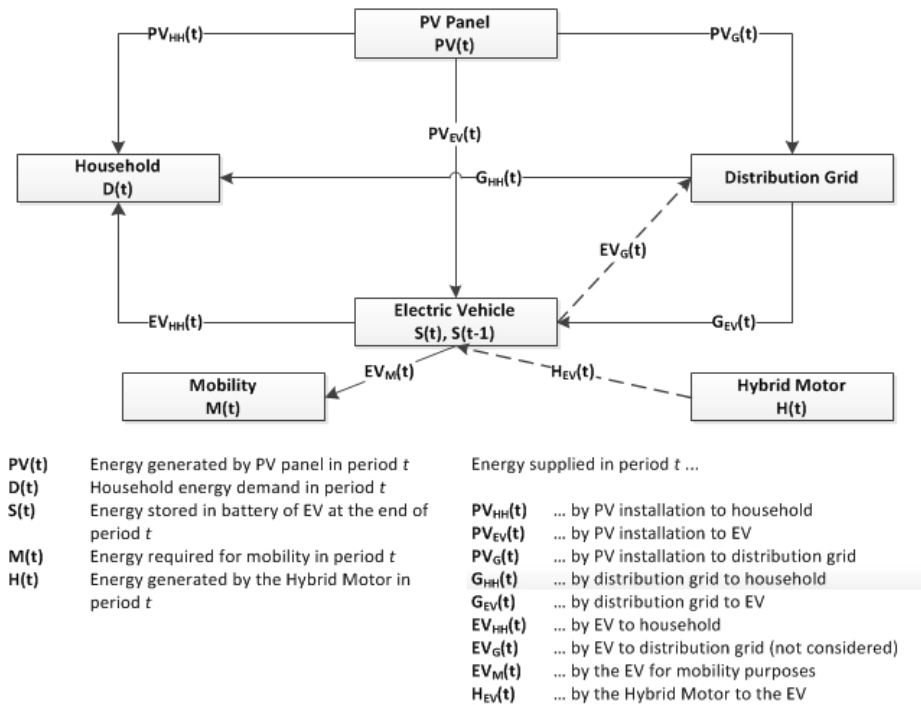


Fig. 1. Illustration of basic scenario

other variables that influence utility, they are assumed to be not affected by the V2G strategy and subsequently not considered.

The requirements the management artifact and the V2G strategies implemented need to fulfill are derived directly from the scenario and Equation 1. The requirements are as follows:

Requirement 1: Monitor and reflect the energy flows in the system and mobility behavior in the calculations

The reasoning behind this requirement should be evident, as the management artifact needs to observe the actual system to make informed decisions. This is even more important for management strategies that depend on predictions of future behavior (energy consumption or generation, mobility), since a historic data set of a higher quality is likely to improve forecasts.

Requirement 2: $\phi(t, M(t), s)$ must be computable for strategy s

This requirement addresses the problem of translating constraints on mobility into a value that can be compared to energy costs. Solving this problem is, however, essential for addressing the fundamental tradeoff of V2G.

Requirement 3: $\int_{t=0}^T U(t, M(t), s) dt$ of a strategy that includes V2G must be higher than the utility without V2G within a reasonable time interval $[0, T]$

This determines that any strategy that includes V2G must provide a higher utility than without this technology. What constitutes a reasonable time interval is largely in the eye of the beholder, but should at least span more than 24 hours (up to several weeks or months), since the EV energy storage takes advantage of daily cycles in PV-generation and mobility behavior.

Requirement 4: The distribution grid should not be adversely affected in terms of demand and feed-ins from PV

Although our approach aims primarily at the synergies between V2G and residential renewable energy sources, the distribution grid should at least not be negatively affected by the management artifact. We do not only consider a decrease in demand to be beneficial for the grid, but also a decrease in feed-ins from PV generation, since a large number of residential PV installations could pose challenges to the distribution grid in the future [20].

4 Strategy Modeling and Evaluation

In this section we present a simulation-based evaluation to illustrate the possible benefits that can be gained from managing synergies between residential renewable energy generation and a V2G-capable EV. As this study intends to serve as a proof-of-concept, we use a very simple management strategy to show that even with this simple strategy there are benefits for an individual household.

The simulation architecture has been implemented in MATLAB and enables the execution of different management strategies. In a real-world setting requirement 1, the monitoring of energy flows, would be addressed by using smart meters that supply the management artifact with data on the current system state. Within the simulation we use the profiles and parameters as specified and explained in Table 1. While we refer to the German market and German prices for electricity, which are heavily distorted by subsidies for renewable energies, this does not invalidate the applicability of our results to other countries without these subsidies. In fact, the spread between c_P and c_F (cost of procuring 1 kWh from the grid and cost of feeding 1 kWh into the grid, respectively) may be less than in an unsubsidized case. For example, an energy retail company may only be willing to voluntarily pay 4 or 5 cents per kWh, whereas the average retail price would only drop by about 2 cents, the current price effect of the solar subsidies. Finally, as recommended in Kempton and Tomic [6], we use a factor of 0.93 for AC/DC-inversion and vice versa. Battery degradation is not considered in our model, because its effect is likely to be negligible [21].

There are two possible ways to address requirement 2, evaluating mobility satisfaction in monetary terms. One is using a PHEV and computing the financial cost in regards to additional fuel required. The underlying assumption is that mobility needs

can always be satisfied, it might just be more expensive to do so. In our simple proof-of-concept strategy we use a different approach by assuming that the driver will never need the full battery capacity between charging cycles. More precisely, we define two decision strategies s_0 and s_{V2G} as the benchmark scenario and the V2G scenario as illustrated in Table 2.

Essentially, the EV owner determines a battery level S^{MIN} deemed necessary to fulfill all mobility needs. Up to this level all excess PV power (i.e. power exceeding household demand, Equations 2a and 3a in Table 2, respectively), and the distribution grid is used to charge the vehicle (subject to maximum charging power constraint, Equations 2e, 2f, 3e, 3f). In the benchmark case, once S^{MIN} is reached, only excess PV power is used to charge the EV (Equation 2e). All remaining PV power is fed into the grid (Equation 2g), which also supplies any additional power needed to satisfy demand (Equation 2d). Under the V2G-strategy, the EV uses any energy above S^{MIN} to power the household if the power generated from PV is not sufficient (Equation 3c). Hence, the central difference between the strategies and the potential for cost savings is how each strategy deals with stored energy above S^{MIN} .

Current BEV models still have a quite limited driving range and thus cannot fully replace cars powered by combustion engine. Hence, in our benchmark scenario we assume the EV to be the second car in the household. Specifically, we consider a four-person household in rural Germany with a household demand and PV generation as illustrated in Figure 2 for a sample day. This household consists of one full-time working parent (using the primary car), one stay/work-at-home parent (using the EV), a schoolchild and a toddler. While this seems to be a very strict selection, the artifact is not solely aimed at this small subset of the population. Basically, any household where one car is parked at home for a significant part of the day would achieve similar results. We use this very specific selection only to generate realistic patterns for driving behavior. Hence, our sample household and a household with a stay-at-home senior may both have a car being parked at home during the day, but they would undertake trips for completely different reasons.

We analyze the effects of our management artifact during one week of August with the following mobility requirements for the EV derived from our data set:

- **06.30 AM – 07.00 AM** Driving child to school and returning home (16 km)
- **12.30 PM – 02.00 PM** Shopping, picking child up from school and returning home (22 km)
- **05.30 PM – 05.45 PM** Short additional shopping trip (3 km)

Table 1. Variables and Parameters

Variable / Parameter	Value	Unit	Comment
$D(t)$	N/A	kW	This is simulated according to the standardized household demand profile of the German Association of Energy and Water Industries [22] for 2011 (15 minute intervals) with an annual demand of 3880 kWh. While this load trace is averaged over all households and flattens peaks, this does not present an issue for the proof-of-concept, since the flexible discharging of EVs would allow for a better management of demand spikes than provided by PV. The beneficial effect of our management artifact is thus at most underestimated.
$PV(t)$	N/A	kW	PV generation is based on the trace of a rooftop PV installation on a single-family home in eastern Bavaria, Germany (5.58 kWp installed).
$M(t)$	N/A	N/A	Mobility needs are constructed on the basis of the study “Mobility in Germany, 2008” [23], which documents several tens of thousands of trips in passenger cars taken in 2008 and links them to households and individuals. As mobility behavior is very persistent, there should arise no problem from linking these requirements with PV and demand data from 2011.
S^{MAX} OUT^{MAX} IN^{MAX} EV_M	16 1.5 2.4 0.187	kWh kW kW kWh / km	The technical parameters for the EV are based on the Mitsubishi i-MiEV electric car, as Mitsubishi is one of the first producers planning to introduce V2G-capable vehicles at least for emergency support [24] Additional parameters are taken from the “Fuel Economy Guide” [25]. They represent the maximum energy that can be stored in the battery, the maximum power the vehicle can feed into the household, the maximum power the vehicle can be charged with and the average energy required for 1 km of driving distance, respectively.
c_G c_P c_P	0.2495 -0.1243 -0.2443	€ / kWh € / kWh € / kWh	These parameters represent the cost of a single kWh procured from the distribution grid, PV-generated and used for private consumption, and PV-generated and fed into the grid, respectively. Values are according to the German subsidy scheme [26] and the average retail price in Germany in 2011 [27].

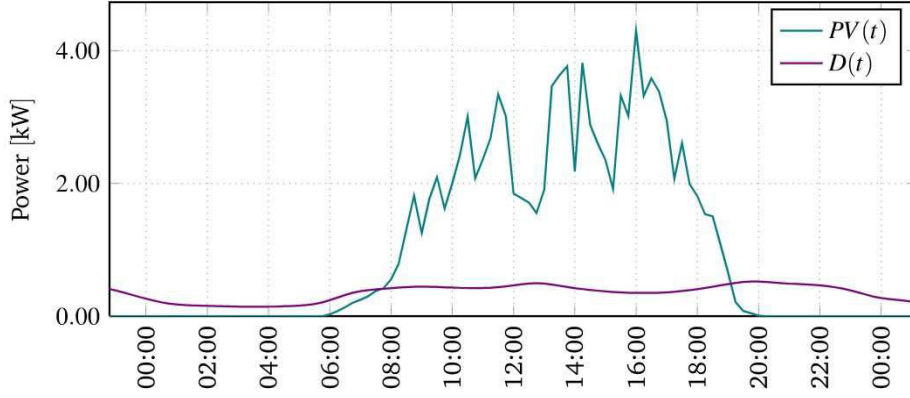


Fig. 2. Household demand and PV generation as on August 5, 2011

Table 2. Strategies

s_0 :	$PV_{HH}(t) = \min[D(t), PV(t)]$	(2a)
	$y = \begin{cases} 1 \\ 0 \end{cases}$	<i>if EV plugged in</i> <i>otherwise</i>
	$EV_{HH}(t) = 0$	(2c)
	$G_{HH}(t) = D(t) - PV_{HH}(t)$	(2d)
	$PV_{EV}(t) = \begin{cases} \min[PV(t) - PV_{HH}(t), IN^{MAX}] \\ 0 \end{cases}$	<i>if $y = 1$ and $S(t) < S^{MAX}$</i> <i>otherwise</i>
	$G_{EV}(t) = \begin{cases} IN^{MAX} - PV_{EV}(t) \\ 0 \end{cases}$	<i>if $y = 1$ and $S(t) < S^{MIN}$</i> <i>otherwise</i>
	$PV_G(t) = PV(t) - PV_{HH}(t) - PV_{EV}(t)$	(2g)
s_{V2G} :	$PV_{HH}(t) = \min[D(t), PV(t)]$	(3a)
	$y = \begin{cases} 1 \\ 0 \end{cases}$	<i>if EV plugged in</i> <i>otherwise</i>
	$EV_{HH}(t) = \begin{cases} \min[\max[D(t) - PV_{HH}(t), 0], OUT^{MAX}] \\ 0 \end{cases}$	<i>if $y = 1$ and $S(t) > S^{MIN}$</i> <i>otherwise</i>
	$G_{HH}(t) = D(t) - PV_{HH}(t) - EV_{HH}(t)$	(3d)
	$PV_{EV}(t) = \begin{cases} \min[PV(t) - PV_{HH}(t), IN^{MAX}] \\ 0 \end{cases}$	<i>if $y = 1$ and $S(t) < S^{MAX}$</i> <i>otherwise</i>
	$G_{EV}(t) = \begin{cases} IN^{MAX} - PV_{EV}(t) \\ 0 \end{cases}$	<i>if $y = 1$ and $S(t) < S^{MIN}$</i> <i>otherwise</i>
	$PV_G(t) = PV(t) - PV_{HH}(t) - PV_{EV}(t)$	(3g)

While this behavior is unlikely to repeat every day for the entire week (usually there is no school on the weekend), it still constitutes a reasonable benchmark case (the child might go to sports practice, instead). We set S^{MIN} to 70% of the maximum, i.e. 11.2 kWh. This guarantees that all trips can be realized and leaves some leeway for additional unplanned trips.

As our assumption is that this decrease of the effective battery capacity does not limit mobility behavior, $\phi(t, M(t), s_0)$ and $\phi(t, M(t), s_{V2G})$ would be equal, thus eliminating each other when calculating the difference in the total utilities of each strategy.

The difference in utilities reduces to the purely monetary difference in the electricity costs, resulting in the following equation:

$$\Delta U = -C_E(t, s_{V2G}) + C_E(t, s_0) \quad (4)$$

with

$$C_E(t, s_i) = \int_{t=0}^T [c_G(G_{HH}(t, s_i) + G_{EV}(t, s_i)) + c_P(PV_{HH}(t, s_i) + PV_{EV}(t, s_i)) + c_F PV_g(t, s_i)] dt. \quad (5)$$

Since the demand and PV data is divided in intervals of 15 minutes, we used a discrete approximation of these functions. Figure 3 illustrates the daily costs for electrical energy associated with each strategy, which were obtained through a computational experiment. The variance between the days is largely caused by the volatility of PV generation, but also by daily differences in the household demand. Negative costs implicate a profit for the household, resulting in an increase in profits of 9.15% over the week (27.17€ compared to 24.89€) from employing the V2G-strategy.

The actual impact of the V2G-strategy for a specific day is depicted in Figure 4. While the left graph shows which sources supply the electricity demand of the household given s_0 , the right graph illustrates this for s_{V2G} . With the latter strategy the EV relieves the distribution grid during the evening hours and at night, thus even dampening the load peak between 6 and 8 PM. Over the entire week the feed-ins from PV on the grid have been reduced by 14.02 kWh (14.07%), while the energy procured from the grid has been reduced by 15.87 kWh (55.83%, difference in absolute terms largely due to losses from inversion). This shows that our management artifact can provide additional revenues for the household and simultaneously reduce (peak) load on the distribution grid – even with a very simple management strategy.

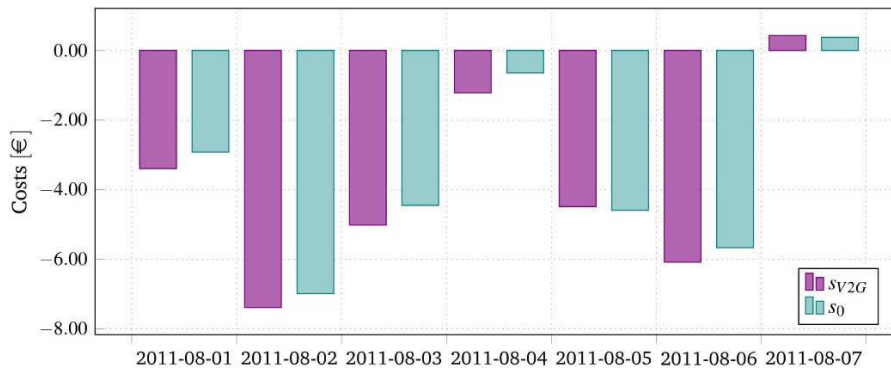


Fig. 3. Daily costs for electrical energy

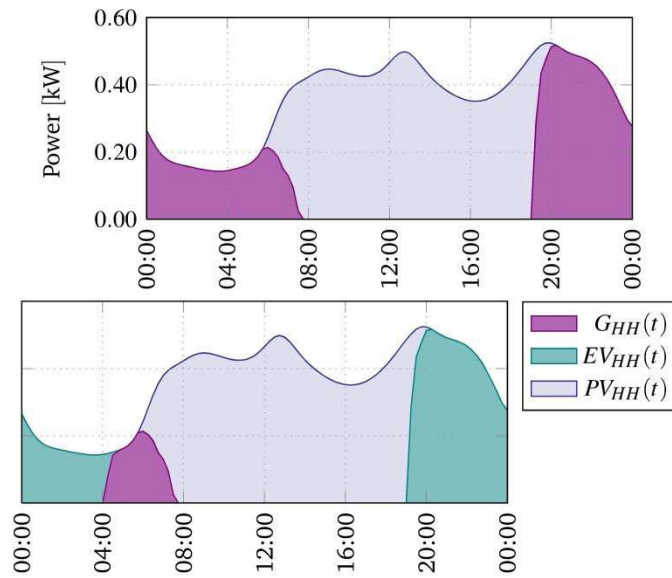


Fig. 4. Energy sources that supply household demand (top: s_0 / bottom: s_{V2G})

5 Discussion

It has been shown that using our management artifact households can realize monetary benefits through V2G without relying on (not yet existing) aggregation programs. While not every week is an August week, this sample household could gain revenues of around 60€ during a year, estimating that on average about 50% of the additional weekly income or savings is realistic for the entire year. This is almost an entire monthly rate of the average standardized three-person household in Germany [27].

However, this is just a lower boundary to the possible financial gains, as we used a very simple decision strategy that did not employ any optimization techniques and focused entirely on the synergies between residential PV and energy storage. In this section we briefly present three possible additions to our model that could substantially increase these revenues.

Strategies under Uncertainty (with or without Signaling)

In our case study, the mobility needs $M(t)$ are assumed to be equally satisfied by the strategy with V2G and without V2G. Consequently, the satisfaction function Φ was eliminated when taking the difference of utilities. However, the choice of a fixed S^{MIN} severely limits the ability of the management artifact for intraday optimization. This problem could be alleviated by handling ϕ differently and considering a PHEV instead of a BEV. The full satisfaction of $M(t)$ would be formulated as a constraint on the optimization problem, handling a low battery state by relying on the more expensive combustion engine. This would allow the management artifact to optimize V2G power supply over a future time interval, subject to forecasts on future demand, PV generation and mobility behavior. The results could additionally be improved by enabling the household to signal short-time mobility needs.

Non-uniform Energy Retail Pricing

Varying retail prices of electricity within a day could increase these revenues from intraday optimization even further. This does not necessarily require real-time pricing, as a dual-tariff structure (day / night) would suffice. Photovoltaic energy would then preferably be distributed during high price times, subject to mobility constraints.

Participation in V2G Program for Frequency Control

White and Zhang [24] explore the potential of V2G aggregation programs simultaneously entering the markets for peak-load reduction and frequency control and this concept should be adaptable to our approach, as well. Once such a program actually exists, it could be incorporated into the optimization calculus of the management artifact. The optimal strategy would then maximize the superior revenues from the V2G program and use any spare capacity for excess PV energy.

6 Conclusion

While most research focuses on the aggregated benefits of V2G-capable EVs, there exist potential revenues for individual households from synergies between residential renewable energy generation and EV energy storage. In this paper we introduced a management artifact that supports a household in realizing these revenues. The main advantage of this approach is that it does not rely on a high market penetration and

social acceptance of EVs – even a single household with a PV installation and an EV can profit.

Our study showed how simple information systems open up new possibilities for integrating renewable energies and electric mobility. However, this work should serve as a basis for future research on the role of information systems in this context. Systems that allow for signaling or prediction of driving behavior could substantially enhance the results produced in this study.

The strategy we implemented as a proof-of-concept was a simple decision strategy, but, nevertheless, produced non-negligible revenue increases. In our future research we will further extend this concept to PHEVs, thus enabling the computation of intra-day optimization strategies. We will also consider the effects of non-uniform energy retail pricing and the compatibility with V2G programs for frequency control.

References

1. Heinberg, R., Fridley, D.: The end of cheap coal. *Nature* 468, 367-369 (2010)
2. Owen, N.A., Inderwildi, O.R. and King, D.A.: The status of conventional world oil reserves – Hype or cause for concern?. *Energy Policy* 38 (8), 4743-4749 (2010)
3. Watson, R.T., Boudreau, M.-C., Chen, A.J.: Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community. *MISQ* 34 (1), 23-38 (2010)
4. Ipakchi, A., Albuyeh, F.: Grid of the future. *IEEE Power and Energy Magazine* 7 (2), 52-62 (2009)
5. Kossahl, J., Busse, S., Kolbe, L.M.: The Evolvement of Energy Informatics in the Information Systems Community – A Literature Analysis and Research Agenda. In: *ECIS 2012 Proceedings*, Paper 172 (2012)
6. Kempton, W., Tomic, J.: Vehicle-to-grid power fundamentals: Calculating capacity and net revenue. *Journal of Power Sources* 144 (1), 268-279 (2005)
7. Kempton, W., Tomic, J.: Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy. *Journal of Power Sources* 144 (1), 280-294 (2005)
8. Cvetkovic, I., Thacker, T., Dong Dong, Francis, G., Podosinov, V., Boroyevich, D., Wang, F., Burgos, R., Skutt, G., Lesko, J.: Future home uninterruptible renewable energy system with vehicle-to-grid technology. In: *IEEE Energy Conversion Congress and Exposition, 2009. ECCE 2009*, pp. 2675-2681. IEEE Press (2009)
9. Gurkaynak, Y., Khaligh, A.: Control and Power Management of a Grid Connected Residential Photovoltaic System with Plug-in Hybrid Electric Vehicle (PHEV) Load. In: *24th Annual IEEE Applied Power Electronics Conference and Exposition, 2009. APEC 2009*, pp. 2086-2091. IEEE Press (2009)
10. Lopes, J.A.P., Almeida, P.M.R., Soares, F.J.: Using vehicle-to-grid to maximize the integration of intermittent renewable energy resources in islanded electric grids. In: *2009 International Conference on Clean Electrical Power*, pp. 290-295. IEEE Press (2009)
11. Lopes, J.A.P., Soares, F.J., Almeida, P.M.R.: Identifying management procedures to deal with connection of Electric Vehicles in the grid. In: *2009 IEEE Bucharest PowerTech*, pp. 1-8. IEEE (2009)
12. Clement-Nyns, K., Haesen, E., Driesen, J.: The impact of vehicle-to-grid on the distribution grid. *Electric Power Systems Research* 81 (1), 185-192 (2009)

13. Clement-Nyns, K., Haesen, E., Driesen, J.: The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid. *IEEE Transactions on Power Systems* 25 (1), 371-380 (2010)
14. Flath, C., Ilg, J., Weinhardt, C.: Decision Support for Electric Vehicle Charging. In: *AMCIS 2012 Proceedings*, Paper 14 (2012)
15. Guille, C., Gross, G.: A conceptual framework for the vehicle-to-grid (V2G) implementation. *Energy Policy* 37 (11), 4379-4390 (2009)
16. White, C.D., Zhang, K.M.: Using vehicle-to-grid technology for frequency regulation and peak-load reduction. *Journal of Power Sources* 196 (8), 3972-3980 (2011)
17. Andersson, S.-L., Elofsson, A.K., Galus, M.D., Göransson, L., Karlsson, S., Johnsson, F., Andersson, G.: Plug-in hybrid electric vehicles as regulating power providers: Case studies of Sweden and Germany. *Energy Policy* 38 (6), 2751-2762 (2010)
18. Sovacool, B.K., Hirsh, R.F.: Beyond batteries: An examination of the benefits and barriers to plug-in hybrid electric vehicles (PHEVs) and a vehicle-to-grid (V2G) transition. *Energy Policy* 37 (3), 1095-1103 (2009)
19. Turton, H., Moura, F.: Vehicle-to-grid systems for sustainable development: An integrated energy analysis. *Technological Forecasting and Social Change* 75 (8), 1091-1108 (2008)
20. Paatero, J.V., Lund, P.D.: Effects of large-scale photovoltaic power integration on electricity distribution networks. *Renewable Energy* 32 (2), 216-234 (2007)
21. Peterson, S.B., Apt, J., Whitacre, J.F.: Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization. *Journal of Power Sources* 195 (8), 2385-2392 (2010)
22. E.ON Mitte AG: Standardized load profiles for 2011, <http://www.eon-mitte.com/index.php?parent=1019> (2012)
23. BMBVS (Federal Ministry of Transport, Building and Urban Development): Mobility in Germany 2008, data available upon request (2010)
24. Green Car Reports: Mitsubishi Turns 2012 'i' into Portable Emergency Power Station (2011), http://www.greencarreports.com/news/1060857_mitsubishi-turns-2012-i-into-portable-emergency-power-station
25. U.S. Department of Energy: 2012 Fuel Economy Guide (2012), <http://www.fueleconomy.gov/feg/pdfs/guides/FEG2012.pdf>
26. German Bundestag: Act on granting priority to renewable energy sources, as applicable on Jan 1, 2012, http://www.erneuerbare-energien.de/files/english/pdf/application/pdf/eeg_2012_en_bf.pdf
27. BDEW (German Association of Energy and Water Industries): Average retail price for household electricity (2011), http://bdew.de/internet.nsf/id/DE_20100311_PM_46_Prozent_des_Strompreises_sind_Steuern_und_Abgaben

A Decision Support Tool for the Risk Management of Offshore Wind Energy Projects

André Koukal and Michael H. Breitner

Leibniz University Hannover, Institute of Computer Science for Business Administration,
Hannover, Germany
{koukal,breitner}@iwi.uni-hannover.de

Abstract. This paper provides a decision support tool (DST) to analyze and evaluate the project value of offshore wind energy projects within the framework of project finance. The DST is based on a discounted cash-flow model in combination with a Monte Carlo simulation (MCS) to measure project risks and manage these risks. To consider the special requirements of debt capital providers in this context, key figures like the debt service cover ratio (DSCR) are calculated. The DST is realized in Excel/VBA with the Excel Add-In Oracle Crystal Ball. An offshore wind park example in the German North Sea is simulated to validate the underlying simulation model and the DST.

Keywords: Decision support tool (DST), Monte Carlo simulation, offshore wind energy, risk management

1 Introduction

Offshore wind energy has been developed rapidly in the last twenty years. The technical aspects have been in the foreground for most of the time so that a continuous enlargement of wind turbines could be observed. In contrast to this development there has been no comparable expansion of the amount of constructed offshore wind energy plants. This is because the costs for such investments go up to two billion euros per wind park with a typical nominal power output of 400 MW [1]. Another reason is the large number of significant risk factors. Some are based on the low technical experience with these kind of plants. Others are inherent in this type of projects as the plants are difficult to access due to the great distance to the coastline. Thus also small failures can lead to long downtimes.

To further international efforts of climate protection the European Commission set their own goals regarding the reduction of greenhouse gas emissions which are widely known as the 20-20-20 targets. Furthermore, the current Renewables Directive defined that 20 % of the total EU energy consumption shall come from renewable energy sources [2]. To meet these goals the member states have set up National Renewable Energy Action Plans. In this context, different incentive systems and feed-in tariffs for supporting the extension of the installed offshore wind capacity have been

established. Currently, it can be observed that more and more offshore wind projects are being realized [3].

A majority of the few realized projects was carried out in the framework of corporate finance by large energy supplying companies [1]. At present, most of the projects are conducted in the context of project finance which means taking into account the different concerns of all involved participants. Here, the lenders are of major importance as projects depend to a great part on the debt capital. The latter is only provided against the background of expected future cash-flows of a project.

It is particularly important to consider the numerous risk factors of an offshore wind project associated with these different requirements within the risk management in order to ensure the economic success of such projects. In the context of this paper, risk is defined as an uncertainty with regard to expected realizations of parameters in combination with possible negative effects on various aspects [4]. Additionally, also positive deviations from the expected values are included.

The objective of this paper is to introduce a tool which gives an answer to the question whether a specific offshore wind project provides adequate returns for investors as well as sufficient debt service coverage from the perspective of lenders. For this purpose, the offered solution delivers aggregated financial data and ratios of central aspects that support an economic evaluation that aims at deciding whether a project can be realized. Regarding the fact that a majority of future offshore projects will be realized in the framework of project finance, several key figures related to the demands of the lenders are also determined. A simultaneous consideration of potential risks is not only necessary but also required for a reasonable analysis due to important technological and economic challenges as a product of limited experiences with offshore wind parks. The latter is realized by the estimation and determination of appropriate risk indicators.

2 Related Work and State of the Art

In this section, publications that conduct economic evaluations of offshore wind parks are presented. The aim of this section is to identify interesting approaches, whose fragments can be used to define a model and build a tool that takes all aspects mentioned above into consideration.

Hirschhausen and Jeske [5] use a discounted cash-flow model to evaluate the net present value of three fictitious offshore wind parks in the German North and Baltic Sea. They consider only pure corporate finance and do not take project risks into account. The research of Madlener et al. [6] provides a cash-flow model on which a MCS is applied. Their quantitative model makes use of the weighted average cost of capital method to discount the future project cash-flows of an offshore wind park in order to calculate the present value of the project. Additionally, a comprehensive analysis of various risk factors and their influences on different parts of a project is performed. On the result of the subsequent MCS which results in a distribution of the discounted project value, the value-at-risk principle is applied. A consideration of further key figures or aspects of project finance does not take place. Levitt et al. [7]

perform an analysis of the breakeven price of electricity for offshore projects in various countries with a cash-flow model. Different financing concepts are considered as well as variances in the investment or operating costs within a sensitivity analysis. An examination of the influence of specific risk factors is not performed, neither are financial key figures calculated. KPMG [1] provides multiple results of scenario analyses and takes project finance in combination with related key figures into consideration. Neither is a detailed model presented nor is a comprehensive risk analysis performed. Prässler and Schaechtele [3] conduct a comprehensive assessment of the financial attractiveness of offshore wind power markets in Europe. In this context a DCF model is set up, key figures like the internal rate of return (IRR) are calculated and multiple scenarios are scrutinized. Aspects of project finance or the consideration of risk factors are not examined. Schillings et al. [8] provide a DST to identify the potential of offshore wind energy and its costs in the North Sea. Their research provides a detailed evaluation of various aspects but does not aim at individual projects so that their findings do not support decisions of an entitlement group which is relevant in this paper.

The result of the literature research indicates that no publication exists that addresses the demands of investors as well as lenders under simultaneous consideration of project risks. Therefore, an approach that includes the determination and provision of financial data which are decision-relevant for all stakeholders is introduced in this paper. According to the categorization of Wilde and Hess [9] a formal-deductive analysis in combination with a simulation is seen as an appropriate research method. This results in the creation of a mathematical model in the shape of a discounted cash-flow model to illustrate and analyze the economic connections of individual aspects on which a MCS is applied to evaluate the effects of risk.

Due to the closeness regarding the approach and content with the research of Madlener et al. [6] this paper is based partly on their model.

3 Decision Support Tool

A DST is a computer-based support for management decision makers who are dealing with semi-structured problems [10]. In the context of project finance in the offshore wind energy sector such a tool has to consider the previously mentioned financial aspects and influences of risk. As a basis of this tool a cash-flow model is set up. For the implementation of this cash-flow model any table calculation software is suitable. Microsoft Excel 2010 is chosen due to its wide use and expandability with Add-Ins. In order to perform a MCS the Add-In Oracle Crystal Ball 11.1.2 is used. It extends Excel with simulation, forecasting and evaluation capabilities. Alternative Add-Ins such as Palisade @Risk have almost the same relevant functionality and would also be applicable. The simulation output is processed with Visual Basic for Applications (VBA) macros in order to implement an automated visual output in the form of diagrams.

3.1 Cash-flow Model

The basis of the DST is a cash-flow model. Figure 1 illustrates the main components and their relationships. The investment cash-flow consists of several individual cost factors referring to the planning process and the construction period of the wind park.

The earnings within the operating cash-flow result directly from selling the produced energy. The latter depends on the theoretical full load hours multiplied with the nominal power output of the entire wind park. In addition to all costs resulting from the operation of the wind park possible costs of deconstruction are also considered in the expenditures within the operating cash-flow.

Furthermore, the cash-flow model makes use of some parameters with an influence on multiple factors in several years. An increase of all operating cost components is conducted by the inflation rate in every year. For the increase of the electricity market price an independent key figure is used to supply a better possibility of setting. This parameter, the construction period and the net full load hours affect only the earnings within the operating cash-flow.

Operating cash-flow, depreciation of the fixed assets and tax rate are used for the calculation of the taxes. The sum of investment cash-flow, operating cash-flow and taxes results in the free cash-flow.

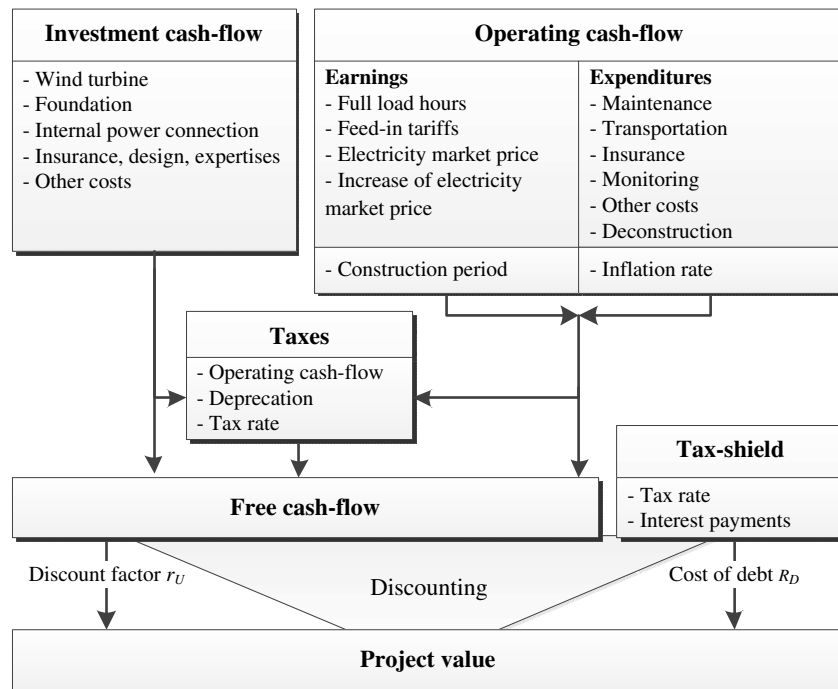


Fig. 1. Project value calculation in the cash-flow model

3.2 Discounted Cash-flow (DCF) Method

The calculation of the project value as well as financial key figures is performed by discounting the project cash-flows. There are several approaches to apply this method. The use of the adjusted present value (APV) approach is a good choice in case the debt-equity ratio is not constant in the course of time [11]. This holds for project finance. The first part of the net present value or project value calculation in equation 1 represents the present value of all free cash-flows (FCF) and the second part the present value of the tax shield of the project [12]. The latter defines an increase in the company's value as a result of the tax saving obtained by the payment of interest [13].

$$Project\ Value = \sum_{t=1}^T \frac{FCF_t}{(1+r_U)^t} + \sum_{t=1}^T \frac{\tau * [r_{D,t} * D_{t-1}]}{(1+r_{D,t})^t} \quad (1)$$

The present value of the tax shield is based on the multiplication of the tax rate r and the interest expense $[r_{D,t} * D_{t-1}]$ discounted with the cost of debt of the respective period. The discounting of the FCF is done by applying the discount factor r_U which is the average of the return on equity r_E and the cost of debt r_D weighted with the share of equity E and debt D on the company value V (equation 2).

$$r_U = r_E * \frac{E}{V} + r_D * \frac{D}{V} \quad (2)$$

The costs of debt are determined by loan agreements while the return on equity has to be determined with the capital asset pricing model (CAPM) in equation 3:

$$r_E = r_f + (r_M - r_f) * \beta \quad (3)$$

It is based on the risk free interest rate r_f , the market risk premium $(r_M - r_f)$ which includes the market interest rate r_M and the beta factor. The latter involves the systematic risk of the investment compared to risks on general markets [14].

3.3 Financial Key Figures

For project developers not only the calculation of the project value is important but also the determination of the internal rate of return (IRR). The IRR for an investment is the discount factor that will make the present value of the project zero [15]. It indicates the interest yield an investor can reach with an investment.

The relationship between project developers and lenders is characterized by the supply of debt capital and the regular repayment of the loan. For the lender it is important to have key figures that aim at the valuation of the possibility of debt service coverage. The basis for the calculation of every subsequent key figure is the cash-flow available for debt service (CFADS). It is the FCF on which the investment cash-flow is added or, in other words, the operating cash-flow after taxes.

The most important key figure for lenders is the debt service cover ratio (DSCR) [16]. It is the quotient of the CFADS and the debt service (DS) and represents the coverage of debt service for every single period of a project (equation 4).

$$DSCR_t = \frac{CFADS_t}{DS_t} \quad (4)$$

Additional key figures are the loan life cover ratio (LLCR) and the project life cover ratio (PLCR) which are only useful in combination with the DSCR. They are calculated as quotient of future CFADS discounted by the cost of debt r_D and the amount of outstanding debt of one period D_{t-1} . The LLCR refers to the CFADS of the remaining loan life, whereas the PLCR refers to all outstanding CFADS of a project.

$$LLCR_t = \frac{\sum_{l=t}^L \frac{CFADS_l}{(1+r_{D,l})^l}}{D_{t-1}} \quad (5)$$

$$PLCR_t = \frac{\sum_{p=t}^P \frac{CFADS_p}{(1+r_{D,p})^p}}{D_{t-1}} \quad (6)$$

3.4 Monte Carlo Simulation

As a pure contemplation of the expected values does not provide a sufficient basis for investment decisions due to an inadequate consideration of possible risks, the model is extended by a MCS. This aims at the inclusion of a wide variety of different risk factors which have an effect of an offshore project.

In order to take the effects of these risks into consideration individual probability distributions are set up for every factor within the investment and the operating cash-flow as well as for the parameters with an influence on multiple factors (see 3.1). Due to an insufficient database on risks in the offshore wind sector, BetaPERT distributions are used because they only need a minimum, a maximum and a most likely value in order to be completely described [17]. The distribution shows a similar shape as a normal distribution if the defined minimum and maximum values have the same absolute deviation from the expected value. The simulation results in a variety of different forms of every target variable (e.g. project value, DSCR).

One key figure which considers the project risks and expresses them in one ratio is the value-at-risk (VaR). It originally indicates the maximum loss in monetary units that is not exceeded within a specific time frame and a specific confidence level [18]. In combination with the MCS this principle can be applied on every target variable of the cash-flow model.

In relation to the project value the VaR expresses the minimum value of the project that is not undercut by a certain probability (confidence level). For the DSCR, LLCR and PLCR it expresses the minimum value of the key figure for a defined probability analogous to the project value.

4 Simulation and Results

The theoretical part with its presented model and methods is now used to perform a simulation of a fictitious offshore wind park in the German North Sea. The assumptions about the characteristics of the wind park are presented in Table 1. They are based on German projects which are currently being planned and supposed to be realized in the near future [1], [6], [19]. The fictitious wind park therefore represents a typical offshore wind project in Germany.

Table 1. Assumptions about the fictitious offshore wind park

Key parameters			
Parameter	Value	Parameter	Value
Wind energy plants	80	Distance to coast	90 km
Nominal output	5000 kW	Depth of water	40 m
Expected annual energy output	1540 GWh	Tax rate	35 %

Investment costs			
Component	Costs per kW	Costs (total)	Discount / Surcharge
Wind turbine	1706 €	682.4 M€	- 5 % / + 5 %
Foundation	852 €	340.6 M€	- 10% / + 10 %
Internal power connection	595 €	238.0 M€	- 5 % / + 5 %
Design / Insurance / Expertise	169 €	67.5 M€	- 10 % / + 15 %
Other costs	279 €	111.5 M€	- 25 % / + 25 %
Total investment costs	3600 €	1440.0 M€	

Operating costs (per year)			
Component		Costs (total)	Discount / Surcharge
Maintenance		22.5 M€	- 25 % / + 25 %
Insurance		18.8 M€	- 5 % / + 25 %
Transportation		3.0 M€	- 25 % / + 25 %
Monitoring		1.0 M€	- 5 % / + 5 %
Other costs		0.9 M€	- 5 % / + 5 %
Total operating costs		46.2 M€	

Parameters with an influence on multiple factors		
Parameter	Value	Discount / Surcharge
Annual inflation rate	2 %	- 25 % / + 50 %
Annual increase of electricity market price (5 ct/ kWh today)	2 %	- 20 % / + 20 %
Net full load hours	3850	- 20 % / +10 %
Construction period	30 months	- 6.67 % / + 20 %

The park consists of 80 wind energy plants with a nominal power output of 5000 kW per engine. Against the background that, in theory, 4000 full load hours can be achieved if internal shadowing effects are considered [20], an assumption of 3850 full load hours after technical unavailability should be suitable. The anticipated annual energy output is calculated as follows: 80 wind energy plants · 5000 kW nominal output 3850 full load hours = 1540 GWh [3]. The project lifetime is set to 20 years beginning at the time when the power generation starts.

Investment costs are set at 3600 € per kW nominal power output. This value is based on different findings in the literature [1], [3], [21]. For the entire OWP with a 400 MW power output the costs amount to 1440 million €. These costs are divided into multiple cost components. The breakdown of the total costs to individual components is based on analyses of the recent past [3], [19], [21].

According to different reports the operating costs are set at 30 € per MWh power output in the year of the commissioning of the wind park in the middle of 2015 [1], [19]. For an expected energy output of 1540 GWh the yearly costs are 1 540 000 MWh · 30 €/MWh = 46.2 M€ and initially 23.1 M€ in 2015. The breakdown to cost components is based on the research of Madlener et al. [6]. For all components of the operating costs an annual increase is assumed due to the annual inflation rate. Besides the investment and operating costs, deconstruction costs of 80 M€ are assumed at the end of the project lifetime [6].

Conforming to the actual market situation for project finance an equity ratio of 40 % is assumed so that the total amount of debt is 864 M€. Two different debt service providers offer the debt for a duration of 12 years. Firstly, there is a banking consortium with a debt amount of 564 M€ and an interest rate of 7 % p.a. Secondly, there is the European Investment Bank (EIB) with a debt amount of 300 M€ and an interest rate of 6 % p.a. [1], [19].

A further important aspect is the compensation of the produced energy. For the first 9 years the initial compensation is 19 cents/kWh due to the use of the compression model offered by the German Renewable Energy Sources Act 2012. After that a location-based compensation of 15 cents/kWh is paid due to the water depth of 40 m and the distance to the coast of 90 km. After the initial and the location-based compensations have expired, the generated power is sold on the general market with a price which is currently around 5 cents / kWh [22].

4.1 Determination of the Discount Rates

In order to apply the APV method for the discounting of the project cash-flows, the discount rates must be determined first. The return on equity has to be determined as shown in equation 3. For this, a risk free interest rate of 2.0 % is used [23]. The market risk premium is approximated to a value of 10 % which implies a market interest rate of 8 %. A beta value of 1.8 is assumed which is based on the following considerations: The unlevered beta for companies which operates in the wind energy sector has an average value of 1.2 [24]. The risk for a single project is likely to be higher than that average, especially in the relatively new offshore wind sector which allows to apply an additional risk premium of 0.3. Due to the use of the compression model for

the initial compensation the financial risk is increased even further [19] which provides a repeated rise of the beta of 0.3. These assumptions result in a return on equity of $r_E = 2 \% + 8 \% \cdot 1.8 = 16.4 \%$. This is similar to the result of a survey about the expectations of the return on equity of project developers. Without the consideration of a higher risk due to the compression model they require a return on equity of 14-15 % [19].

The costs of debt are determined as 6.66 % by the average weighted interest rate of debt over the project lifetime. The insertion of the return on equity and the costs of debt with their percentage of shares of the total capital into equation 2 results in a discount factor of $r_U = 16.4 \% \cdot 40 \% + 6.66 \% \cdot 60 \% = 10.56 \%$.

4.2 Definition of Probability Distributions

Due to the use of BetaPERT probability distributions for performing a MCS it is necessary to specify a minimum, a maximum and a most likely value for every risky parameter. While all expected values of these parameters are used as most likely points of the probability distributions, the minimum and maximum points are calculated with percentage discounts from and surcharges on top of the expected values. These discounts and surcharges are presented in Table 1 next to their respective parameter and are mostly based on the research of Madlener et al. [6]. As an example of this procedure the minimum and maximum points of the probability distribution of the annual inflation rate are calculated. The annual inflation rate in Germany has been between 1 % and 3 % in the last years [25]. The variances of the expected value at a height of 2 % are assumed to be close to the statistic values with -25 % and +50 %. Consequently, the minimum is $75 \% \cdot 2 \% = 1.5 \%$ and the maximum is $150 \% \cdot 2 \% = 3.0 \%$.

4.3 Results

All previously mentioned values of the different parameters are inserted into the cash-flow model. A part of all resulting project cash-flows is presented in Table 2. The use of the APV method can be followed by the presentation of the discounted values of the free cash-flow and the tax-shield and the subsequent addition of these values. The result allows different statements about the OWP:

1. The project value of 72.1 M€ is positive and consequently the investment opportunity offers a return on equity which is bigger than the required 16.4 %.
2. On the basis of this project value the IRR is 12.05 %. The return on equity which results in this value under otherwise equal conditions and assumptions is 20.14 %, an increase of 22.8 % compared to the required return on equity.
3. The cumulative project value provides an answer to the question at which time the project is expected to turn into a positive investment. For the first time, this is the case after 14 years in the year 2027.

Table 2. Calculation of the expected project value (millions of €)

	Σ	2013	2014	2015	2016	2017	...	2035
1. Investment cash-flow		-145.0	-575.0	-720.0			...	
2. Operating cash-flow				123.2	245.5	244.5	...	-89.1
2.1 Earnings				146.3	292.6	292.6	...	59.5
2.2 Expenditures				-23.1	-47.1	-48.1	...	-148.7
3. Taxes						-21,2	...	
Free cash-flow		-145.0	-575.0	-596.8	245.5	223.3	...	-89.1
Tax-shield			6.7	20.6	21.1	19.8	...	
Discounted free cash-flow	-49.8	-131.2	-470.4	-441.7	164.3	135.2	...	-8.9
Discounted tax shield	121.9		6.3	18.1	17.4	15.3	...	
Project value	72.1	-131.2	-464.2	-423.6	181.7	150.5	...	-8.9
Cumulative project value		-131.2	-595.3	-1018.9	-837.2	-686.7	...	72.1

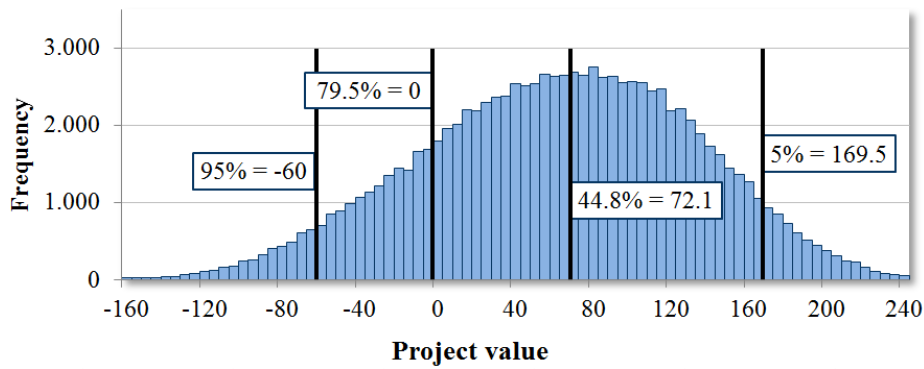


Fig. 2. Distribution of the project value (millions of €, 100000 simulations)

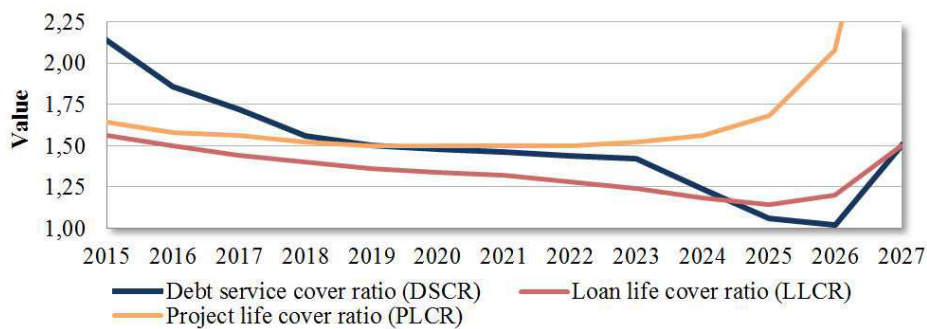


Fig. 3. Key figures at a specified confidence level of 95%

On the cash-flow model the MCS is applied. It is performed with 100 000 simulation runs. It needs 428 seconds on an Intel Core i7-2640M CPU with 2.80 GHz, 8GB Ram

and Microsoft Windows 7 64bit as operating system. The effect on the project value is shown in Figure 2. Some statements can be derived from the distribution:

1. At a confidence level of 95 % the project value is -60 M€. Therefore, with a certainty of 95 %, the value of the project is at this or a higher amount.
2. The calculation of the IRR for the project value of -60 M€ can be made analogous to point 2 of the considerations about the expected project value. In this case the IRR is at least 9.26 % with a certainty of 95 %.
3. With a probability of 79.5% the minimum project value is 0 and investors get an interest yield that is at least at the same level as other investments with similar risk.
4. The probability to reach a project value of at least 72.1 M€ which corresponds with the expected project value is only 44.8 %.

The consequences of the MCS on the financial key figures in combination with the VaR principle are presented in Figure 3. Lenders normally demand a value of the DSCR between 1.35 and 1.45 [1]. Most of the time, the DSCR has an uncritical value greater than this range. In the years between 2023 and 2027, in which the compensation for the produced energy guaranteed by law will have ended, the DSCR is only over a value of 1. Due to the consideration of the project risks at a 95 % confidence level and values of the LLCR and especially the PLCR which are significantly higher and partly rapidly increasing in the previously mentioned critical periods, the project cash-flows offer adequate debt service coverage for the entire loan life.

5 Discussion

The presented results clearly indicate that an investment into an offshore project in the German North Sea is in the average profitable for project developers and lenders. It is obvious that the DST provides an aggregated representation of important financial key figures and gives an answer about the economic efficiency of offshore wind projects which are constructed and operated within the context of project finance.

5.1 Evaluation

This section addresses the question whether the DST with its underlying model meets the criteria of the scientific rigor and the practical relevance.

Applicability to a Class of Problems. The presented model is focused on offshore wind energy projects in Germany which are constructed and operated in the context of project finance. Although this focus exists, the DST can be used for projects with other framework conditions with some restrictions. The main restriction is based on the modeling of the compensation for the produced energy. It is tailored to suit the requirements of the feed-in tariffs guaranteed by law in Germany with its initial and location-based compensation before the produced energy is sold on the general market.

With little adjustments of the parameters for the compensation structure it is possible to fulfill the economic conditions for the sale of energy in other countries. Three conditions must be met for the transfer: the guaranteed feed-in tariff in the respective country (1) has to consist of at most two different amounts (2) which are respectively paid for a predefined time period (3) without any other subsidies regarding the electricity. It is possible that the compensation is based on the provided feed-in tariff for the entire project life or until a market-based compensation starts. With an adjustment of the parameters for the compensation structure it is possible to use the DST at least for projects in Ireland and France [1], [3]. An evaluation of onshore projects is also possible if the same conditions are fulfilled. Additionally, the model can be applied to projects within the context of corporate finance. In this case the special key figures play an insignificant role. In total, the DST can be used for a variety of different projects in at least three countries.

Innovative Contribution to the Published Level of Knowledge. The results of the literature research in section 2 have shown that no contribution exists which provides aggregated financial data and ratios of central aspects under simultaneous consideration of project risks. Those have also to support decisions of project developers and meet the special demand of lenders.

The DST with its underlying cash-flow model focuses on this gap to provide a contribution which extends the current level of knowledge. Aiming at project developers, it provides a detailed presentation of all project cash-flows and a calculation of the project value and the IRR. Lenders focus on the debt coverage by the project cash-flows. An analysis of the debt service coverage is possible with the provided financial key figures. While the DSCR only considers the debt service coverage of one single period, the LLCR and PLCR offer the possibility to extend this consideration and include additional data. Risk factors are taken into account due to the assignment of probability distributions for all parameter within the investment and operating cash-flow and the performing of a MCS. The results show the influence of the risks on the project value and all other mentioned key figures.

Reasonable Reproducibility and Validation. Setting up a cash-flow model with the aim to calculate a project or corporate value is a standard practice. The structure of the cash-flow model and the simulation described in section 3 are based on the published knowledge and reasoning based on that. For an expert in the field of business computer science and finance it should be intersubjectively comprehensible.

Future Benefits for Stakeholders. This paper has both theoretical implications for the entitlement group of scientists as well as practical implications for project developing companies and lenders.

Theoretical Implications. Regarding the discussion of the consideration of project risks, the performed simulation points out the effects of risk factors on different financial key figures. There are no findings in the literature about these effects on the indicators which are particularly important for the lenders. This indicates that the

consideration of risk factors within the relationship between project developers and lenders for offshore wind energy projects has not yet been sufficiently researched.

Against the background of ambitious expansion targets for the offshore wind energy in various countries there is an increasing examination of regions with greater distances to the coastline and higher water depths with regard to their suitability for offshore projects. The economic attractiveness of these regions tends to be lower [3] and therefore a detailed analysis of all relevant aspects has to be performed in order to be able to reduce the effect of individual risks in the future.

Furthermore, the presented model can be modified and expanded easily to take future changes in the general conditions or new insights into consideration.

Practical Implications. The tool can support decision makers to evaluate the economics of specific offshore wind projects. For project developers the interest rate for the invested equity is of major importance and the provided determination of this value and of the IRR is of great significance. Lenders require a covered debt service. Through the calculation of relevant key figures like the DSCR an evaluation of a project's ability to cover the debt is possible. Additionally, the model helps to understand the effects of changes in the general conditions like feed-in tariffs, a changed cost situation of individual cost factors, a different amount of annual full load hours or alternative discount factors.

Furthermore, the influence of risks on the success of a project becomes clear. Against this background the importance of risk management is emphasized. The examination of individual risk factors offers a possibility to detect which risk factors are the greatest threats for the success of a project and at which point of time in the project planning or operating process it is most important to establish and apply risk management methods.

5.2 Limitations and Further Research

The model uses one single tax rate. In particular with regards to the complex German tax system the cash-flow model provides only an approximation. Deviations of the real situation depend strongly on the individual case of the project. However, the key findings of the model retain their validity.

The technical availability is not an independent parameter in the model but a flat adjustment of the theoretical achievable full load hours. A further development of the model could consider the full load hours and the technical availability independently which would also add the possibility to take improvements of the technical availability in the long term [7] into consideration.

The parameter of the construction period only affects at which time the compensation for the produced energy starts. Especially within the framework of the MCS with its different generated values it is neither intuitive nor realistic that a longer construction period does not result in increased construction costs.

All conclusions derived from the results of the MCS are based on BetaPERT probability distributions and an adequate description of possible distributions of risks. This is only a rough approximation. A better consideration of individual risk factors can be

realized when the knowledge about these risks could be increased by more and longer experiences and better scientific investigations of planning, construction and operation of offshore wind parks. In this case the BetaPERT probability distributions can be replaced by more realistic ones. However, in the near future, no improvement of the data situation can be expected because involved companies classify the majority of this data as secret information.

The aspects mentioned in the discussion of the applicability to other countries in section 5.1 should also be included in a further development of the model. An expansion in order to take different feed-in tariffs of various countries with their respective compensation systems into consideration would provide a larger scope of possible applications and scientific investigations.

6 Conclusion

In this paper a decision support tool is presented which helps to evaluate the economic potential of offshore wind projects. The present value of such projects is calculated with a discounted cash-flow model. Due to the different requirements of project developers and lenders additional key figures like the IRR, DSCR, LLCR and PLCR are calculated to meet the requirements of all stakeholders. A consideration of risk factors is done by the assignment of probability distributions to specific components of the cash-flow model. This is followed by performing a Monte Carlo simulation and applying the value-at-risk principle on the distribution of each target key figure. Overall, the tool allows to evaluate the impact of single input parameters or risk factors so that critical aspects can be identified to give a special consideration of these aspects within the risk management process. Generated results can serve as a useful guidance for project decision makers and scholars alike.

References

1. KPMG: Offshore Wind in Europe. 2010 Market Report (2010)
2. European Commission: Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the Promotion of the use of Energy From Renewable Sources and Amending and Subsequently Repealing Directives 2001/77/EC and 2003/30/EC (2009)
3. Prässler, T., Schaechtele, J.: Comparison of the financial attractiveness among prospective offshore wind parks in selected European countries. *Energy Policy* 45, 86–101 (2012)
4. Kaplan, S., Garrick, B.J.: On The Quantitative Definition of Risk. *Risk Analysis* 1, 11–27 (1981)
5. von Hirschhausen, C., Jeske, T.: Offshore Windenergie. Studie zur Rentabilität von Offshore-Windparks in der Deutschen Nord- und Ostsee. *Zeitschrift für Energiewirtschaft* 29, 55–63 (2005)
6. Madlener, R., Siegers, L., Bendig, S.: Risikomanagement und -controlling bei Offshore-Windenergieanlagen. *Zeitschrift für Energiewirtschaft* 33, 135–146 (2009)
7. Levitt, A.C., Kempton, W., Smith, A.P., Musial, W., Firestone, J.: Pricing offshore wind power. *Energy Policy* 39, 6408–6421 (2011)

8. Schillings, C., Wanderer, T., Cameron, L., van der Wal, J.T., Jacquemin, J., Veum, K.: A decision support system for assessing offshore wind energy potential in the North Sea. *Energy Policy* (2012)
9. Wilde, T., Hess, T.: Forschungsmethoden der Wirtschaftsinformatik. *Wirtschaftsinformatik* 49, 280–287 (2007)
10. Keen, P.G., Scott Morton, M.S.: *Decision Support Systems. An organizational perspective*. Addison-Wesley, Reading, Mass (1978)
11. Luehrman, R.: Using APV: A better tool for valuing operations. *Harvard Business Review* 75, 145–154 (1997)
12. Ruback, R.S.: Capital Cash Flows: A Simple Approach to Valuing Risky Cash Flows. *Financial Management* 31, 85–103 (2002)
13. Fernandez, P.: The value of tax shields is NOT equal to the present value of tax shields. *Journal of Financial Economics* 73, 145–165 (2004)
14. Schwall, B.: *Die Bewertung junger, innovativer Unternehmen auf Basis des Discounted cash flow*. Lang, Frankfurt (2001)
15. Brealey, R.A., Myers, S.C., Allen, F.: *Principles of corporate finance*. McGraw-Hill/Irwin, New York (2011)
16. Böttcher, J.: *Finanzierung von Erneuerbare-Energien-Vorhaben*. Oldenbourg, München (2009)
17. Pleguezuelo, R.H., Pérez, J.G., Rambaud, S.C.: A note on the reasonableness of PERT hypotheses. *Operations Research Letters* 31, 60–62 (2003)
18. J.P. Morgan (ed.): *Introduction to RiskMetrics*. New York (1995)
19. BMU: *Vorbereitung und Begleitung der Erstellung des Erfahrungsberichtes 2011 gemäß §65 EEG. Vorhaben IIe Windenergie* (2011)
20. Jarass, L., Obermair, G.M., Voigt, W.: *Windenergie. Wind als stochastische Energiequelle*. Springer, Berlin Heidelberg (2009)
21. EWEA: *The Economics of Wind Energy* (2009)
22. EEX: *EEX Home Page*, www.eex.com
23. *Boerse-Stuttgart: Markt und Kurse. Bundesanleihen*, <https://www.boerse-stuttgart.de/de/marktundkurse/anleihen/bundesanleihen/bundesanleihen.html>
24. VALNES: *Gutachterliche Stellungnahme zum Unternehmenswert der REpower Systems SE, Hamburg, und zur angemessenen Barabfindung gemäß §§327a ff. AktG* (2011)
25. eurostat: *HVPI - Gesamtindex - Inflationsrate des Jahresdurchschnitts*, <http://epp.eurostat.ec.europa.eu/>

Simplified Bid Languages – A Remedy to Efficiency Losses in Large Spectrum Auctions

Stefan Mayer and Pasha Shabalin

Technische Universität München, Department of Informatics, Garching by Munich, Germany
{stefan.karl.mayer, shabalin}@in.tum.de

Abstract. Combinatorial auctions have been suggested as a mean to raise efficiency in multi-item negotiations with complementarities among goods as they can be found in procurement, energy markets, transportation, and the sale of spectrum auctions. Since 2008 the Combinatorial Clock Auction (CCA), a two-stage auction format has been used in many countries. [8] tested CCA in the lab and found out that the efficiency of CCA was relatively low, since bidders tended to submit too few bids. To reduce bidders' complexity concerning evaluating lots of bundles, we simplified the bidding language without losing efficiency. Hereby, we used the knowledge of super-additivity and the fixed descending complementarity type of our value model. In lab experiments, we tested the two phases of the CCA auction, namely the Combinatorial Clock + (CC+) auction and a sealed bid version, with the simplification separately. Both formats yielded in higher efficiency and revenue than the CCA.

Keywords: Electronic markets and auctions, laboratory experiment, combinatorial auction

1 Introduction

Since 1994 more than 70 spectrum auctions were run using the simultaneous multi-round auction (SMRA), an auction format which is based on [18]. In the SMRA auction several items are sold in a single auction. Bids on combinations of items are not allowed, whereas synergies in bidders' valuations cannot be expressed.

Therefore, there are many strategic problems for bidders (see also [12]), like the exposure risk. Hence and because of several other reasons, combinatorial auctions (CA) were used for spectrum auctions.

Generally, CAs are IT-based economic mechanisms, where bidders can define their own combinations of items called "packages" or "bundles" and submit bids for them. Bids on individual items are not allowed to place. So, bidders can express better their valuation, which increases economic efficiency, especially in the presence of super-additivities respectively economies of scope. Often CAs are iterative auctions, in which an auctioneer computes allocations and asks prices in each round. This would not be possible without IT-based auction platforms which solve hard computational problems in each auction round and calculate new ask prices. This is also a reason

why CAs have been a topic in much recent Information Systems (IS) research. Examples can be found within [16], [6], and [23]. The IS literature proposed also a lot for bidder decision support, designs for new application domains, and the analysis of bidder behavior in CAs (see also [1], [3-4]) An overview about current research in IS can be found within [7].

The design of CA, however, led to a number of fundamental design problems, and many contributions during the past few years [10-11].

An important CA, the Combinatorial Clock Auction (CCA) based on [17], is a two-stage auction format with consists of primary bid rounds for price discovery and a sealed bid round The CCA was used the U.K., the Netherlands, Denmark, and Austria for the recent sale of the 2.6 GHz band.

The frequencies of the 2.6 GHz band are available for mobile services in all regions of Europe. The 2.6 GHz spectrum band includes 190 MHz which are divided into blocks of 5 MHz. It can be used to deliver wireless broadband services or mobile TV. In particular, there are two standards which will likely be used in the 2.6 GHz band, LTE and WiMAX. LTE uses paired spectrum (units of 2 blocks), while WiMAX uses unpaired spectrum (units of 1 block).

[8] compared CCA and the SMRA auction in the lab using two different value models. These models were designed similar to the practice.

In each value model, 4 bidders participated and 24 lots were sold. The small model consisted of two bands with 14 respectively 10 blocks, the large one of 4 blocks with six bands each. In total, in the small value model, the complexity for bidders was quite lower, because they needed only to calculate ca. 50 possible bundles. Contrarily, in the large value model around 2,400 value models needed to be evaluated. Because of this high complexity, the efficiency of CCA in the lab was considerably lower than that of SMRA.

In our paper, we tested the two phases of the CCA, namely the Combinatorial Clock + (CC+) auction and a sealed bid version, for the large value model separately.

We addressed the recent criticism of [12], that bidder could not submit enough bids in a value model consisting of 2,400 possible bundles by reducing bidders' complexity. We simplified the bidding language using the knowledge of super-additivity of our value model and the fixed descending complementarity type. As a consequence, the experiments that we conducted did not suffer from "too few bids" because the bidders had also a bidding tool to evaluate all the possible bundles quickly. This made it easy for them to submit their bids on all the relevant packages. We show that higher efficiency is achieved by our modifications.

This paper is organized as follows. The value model, the simplified bidding language, the payment rule and the four competing auctions are introduced in Section 2. In Section 3, the experimental design is described and the results are presented in Section 4. Finally, Section 5 concludes with a discussion.

2 Auction Design and Theory

We use in our experiments three different CA formats, CCA, CC+ and a sealed bid version. As a matter of completeness we also present the result of the SMRA, because we compare our result to those of [8], who tested SMRA and CCA.

In SMRA, all items were sold at the same time whereas each item had its own price. Bidders could not bid on bundles. If there was overdemand for an item, i.e. the demand is larger than the supply, the auctioneer increased the price for this item by a fixed increment. The auction terminated if there was no bid within one round. SMRA is a generalization of the English auction for several items, where bidders have to pay what they bid.

The CCA auction consisted of two phases, the clock phase and the sealed bid part. All items within one band had the same price, so there were 4 different (band-) prices in our value model for all blocks. The auctioneer announced the new ask price for each band in each round of the clock phase. Afterwards, bidders decided to bid on which amount of items in each band. The primary bids phase ended after there was no overdemand in any bands any more.

The sealed bid phase consisted of only one round with as many sealed bids as desired by the bidders. They were able to bid on any combination of items regardless of the bids of the first phase. A detailed description of SMRA and CCA can be found within [8].

In the following, we describe the implementation of the different formats that we actually tested in our experiments. But first, we explain the value model, the simplified bidding language and the payment rule we applied.

2.1 The Value Model

For our value model we used an economic setting which was inspired by spectrum sales as introduced at the beginning of this paper. [8] found out in lab experiments that the CCA performed poorly in this value model, since bidders had to evaluate 2,400 different possible bundles.

The model had 24 items with four bands, whereas each consisted of six items. Band A was of high value to all bidders and bands B, C, and D were less valuable. Every bidder received a base valuation for an item in each band. Base valuations are uniformly distributed: v_A was in the range of [100; 300], while v_B , v_C and v_D were in the range of [50; 200]. Furthermore, bidders had complementary valuations for bundles of items. Within a band, each block has the same value for bidders. For example A01 has the same value as A02.

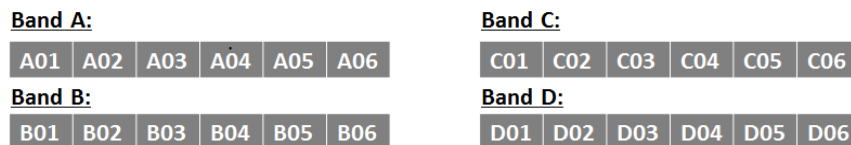


Fig. 1. The value model

In all bands, bundles of two items resulted in a bonus of 60% on top of the base valuations, bundles of three items in a bonus of 50%. More items did not add any extra bonus for this band.

The structure of the value model and the distribution of the item valuations of all bands were known by all bidders. Bidders used an artificial currency called Franc. Although the value models resemble the characteristics of spectrum sales in Europe, this was not known to the subjects in the lab (neutral framing). In figure 2 you find an example concerning the value model with the draw 200 for block A and 100 for the blocks B, C, D. The valuation per unit is the highest at 2 goods. Then, it is decreasing. However, bids could only be placed for the whole amount and not per unit.

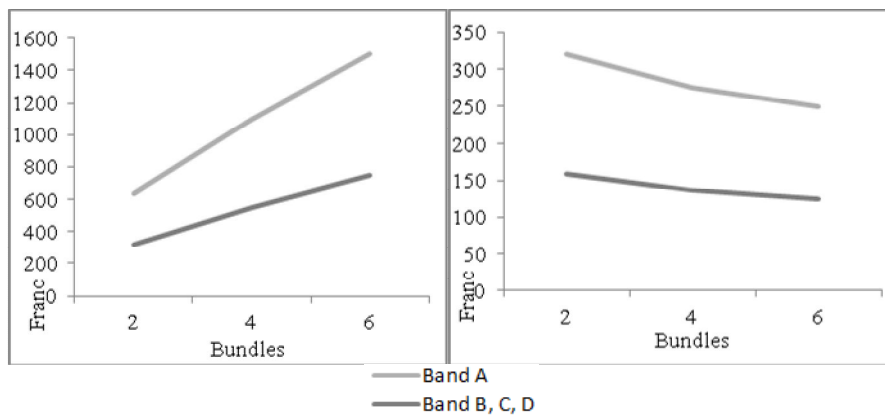


Fig. 2. The valuation structure

2.2 Simplified Bidding Language

The literature often assumes that the auctioneer has no information on synergies in the bidder valuations. This is often not the case. In spectrum auctions, bidders and auctioneers know about the main synergies in the valuations. For example, 2 blocks in 800 have high synergies and 4 blocks in 2.6 as well. Using this knowledge, we constrained the language knowing that the valuation between bands are purely additive, because there are no cross synergies respectively super-additivities between bands. However, we have to remark, that in other situation, when the value model offers e.g. cross synergies or has other properties, the bidding language should be adapted in another way.

Bidders could enter bids for 2, 4, 6 blocks (i.e., their maximum willingness to pay for a desired number of A, B, C and D) for each band, whereas each area has its own price per good which is equal for all goods. That means, altogether 12 bids can be submitted in the sealed bid auction respectively in the CC+ auction. This is a large simplification in comparison to the 2,400 bundles bidders had to evaluate in the CCA. We use an OR (additive-OR) bidding language between the bands and XOR (exclusive-OR) within a band. OR-Bid means that each bidder is willing to win any number

of these disjoint bids, whereas XOR only allows winning at most one of the submitted bids. Thus, a bidder can win a maximum of 4 bids (one bid per band).

Besides, the outcome closure property is fulfilled, because simplification is not constraining. We used the ideas of [19] to restrict the space of messages available to the players. Such a simplification has the advantage that it can eliminate undesirable equilibria. However, tight simplifications, like our modification to the bidding language have to be used. Tightness removes certain equilibria that were created by eliminating profitable deviations. Some notes on this topic area given by [20].

We show, that designing the right bid language can lead to significant efficiency gains over simple SMRA or a full XOR CA, since for CAs to perform well, it is important for bidders to submit all relevant bids.

In all the auction formats, we focus on designs with linear ask prices, where each item in the auction is assigned an individual ask price, and the price of a package of items is simply the sum of the single-item prices. Such prices are easy to understand for bidders in comparison to other pricing rules like the non-linear ask prices.

2.3 The Payment Rule

The bidder-optimal core-selecting payments used within CC+ and the sealed bid auction were calculated using a quadratic program. [14] suggested a procedure to calculate bidder-Pareto-optimal payments from sealed bids right away. Over all total-payment minimizing core points, that one are selected that minimize the sum of square deviations from the VCG payments (minimal Euclidean distance). The objective is to minimize the incentive to misreport one's valuations.

Therefore, the core for package allocation problems has a competitive auction interpretation: an individually rational allocation is in the core if there is no group of bidders who could all do better for themselves and for the seller by raising some of their losing bids.

2.4 The CC+ Auction

[9] described the single-phase CC+, which resembles the first phase in CCA. We tested it in the lab using the simplified bidding language for our value model. With this change we supposed to achieve more robust outcomes in terms of efficiency and revenue than the two-phase CCA. CC+ can be interpreted as a practical implementation of the fictitious "Walrasian auctioneer".

The auction consists of any number of rounds. Bidders enter (bundle) bids for 2, 4, 6 blocks in each round, whereas all bids remain active throughout the auction. OR bids allow the auctioneer to allocate one bundle to every bidder in each of the 4 different bands. In contrast, by XOR, maximum one bundle within a band can be assigned to bidders.

At the beginning of each round, the new prices of goods are announced. Based on these new prices each bidder will report the quantity how much A, B, C and D bundles they want to buy. If there is excess demand (i.e. if the combined demand of all bidders within one band exceeded the number of blocks) in at least one band, a new

round starts with higher prices for the bands with excess demand. Start prices in the first round were set to 100 Franc for items in the A band and 50 Franc in the B, C, and D band. The price update is done by clock ticks for each block within a band. One clock tick in the A range is equivalent to 20 francs, in the B -, C and D range at 15 francs. Besides an activity rule is installed. According to that a bidder must submit at least one bid in each current round, to may bid again for each possible bundle in the next round. If no block is overdemanded, the auctioneer solves the winner determination problem considering all bids submitted during the auction runtime. If the computed allocation does not displace any active last iteration bidder, the auction terminates. The auctioneer selects these bids that maximize the total revenue of the auction.

[9] have shown that it is an ex-post equilibrium strategy for the bidders to submit their power set bids, which means to submit bids on all packages with a positive valuation in each round.

2.5 The Sealed Bid Auction

In the sealed bid auction, bidders enter bids for 2, 4, 6 blocks in the single round, whereas a maximum of 12 bids is possible in the whole auction. The auctioneer collects the bids and determines the revenue maximizing allocation and the prices. The same bidding language is used as within CC+, namely XOR within a band and OR between bands. During the auction, bidders do not get information about activities (bids) of other bidders.

3 Experimental Design

Our experiments are based on the MarketDesigner software framework, which we extended by an implementation of the simplified bidding language and of the two tested auction formats.

We consider 2 treatment factors, auction format and bidding language. The value model was for all formats, SMRA, CCA, CC+ and the sealed bid auction the same. For treatments 1 and 2, where the bidding language was fully expressive we used the results of [8]. For the treatments 3 and 4, which we actually tested in the lab, the simplified bidding language was used. Overall, we get four different treatments:

Table 1. Treatment structure

Treatment no.	Auction format	Bidding language	Auctions
1	SMRA	Fully expressive	16
2	CCA	Fully expressive	16
3	CC+	Simplified	16
4	Sealed bid	Simplified	16

Our Experiments

Our experiments were conducted from June to July 2012 with students in computer science, mathematics, physics, and mechanical engineering at the TU München. The

subjects were recruited via e-mail lists and conducted experiments in a computer lab at the TU München. Each subject participated in a single auction format in one session, but never in different ones. One session comprised all four auctions of one auction format of one wave. For the value model we drew valuations for four waves (A-D) randomly. All auctions of waves A, B, C and D were tested with the four auction formats. CC+ took on average four hours and the sealed bid auction between two and a half hours.

Subjects did not have to prepare for the experiment. All required information was given to the participants. Before each session, the environment, the auction rules and all the relevant information were explained to the participants. All the instructions were read aloud and participants had to participate in a test about the economic environment and the auction rules.

An additional tool to analyze bundle valuations and payoffs was introduced to all subjects. This tool showed a simple list of all available bundles which could be sorted by bundle size, bidder individual valuation or the payoff based on current prices. As a matter of course, last mentioned option was not available for the sealed bid format.

At the beginning of each auction all subjects received the individual draw of valuations, the distribution of valuations, and the information about the complementarity of items. With this information, subjects were asked to reason about the implications of the draw on their bidding in the upcoming auction. Each round was scheduled with 3 minutes. The time given to the subjects in each sealed bid auction was 20 minutes. The subjects could ask for more time if required.

After each session subjects were compensated financially. The total compensation resulted from a 10 Euro show up fee and the auction reward. The auction reward was calculated by a 3 Euro participation reward plus the payoff of all auction payoffs converted from Franc into Euro by a 12:1 ratio. Negative payoffs were deducted from the participation reward. Negative payoffs higher than the participation reward were ignored. Due to the different duration and auction formats, payoffs were different. Therefore, we leveled the expected payoff per participant. In CC+, we compensated two out of four and in sealed bid one out of four auctions. On average, each subject received in CC+ 70.94 EUR and in the sealed bid auction 37.69 EUR.

4 Results

First, we present efficiency and revenue of the different auction formats on an aggregate level.

4.1 Efficiency and Revenue Share

We use allocative efficiency (or simply efficiency) as a primary measure.

Given an allocation X and price set P_{pay} , let $\pi_i(X, P_{\text{pay}})$ denote the payoff of the bidder i for the allocation X and $\pi_{\text{all}}(X, P_{\text{pay}}) := \sum_{i \in I} \pi_i(X, P_{\text{pay}})$ denote the total payoff of all bidders for an allocation at the prices P_{pay} . Further, let $\pi(X, P_{\text{pay}})$ denote the auctioneer revenue. We measure efficiency as the ratio of the total valuation of the resulting allocation X to the total valuation of an efficient allocation X^* .

$$E(X) := \frac{\pi(X, P_{\text{pay}}) + \pi_{\text{all}}(X, P_{\text{pay}})}{\pi(X^*, P_{\text{pay}}) + \pi_{\text{all}}(X^*, P_{\text{pay}})} \in [0, 1] \quad (1)$$

Because of the simplification of the bidding language 100% efficiency cannot always be achieved. To consider this fact, also the relative efficiency $E(X_{\text{rel}}^*)$ is computed. Therefore, we define X_{rel}^* as the best allocation that can be achieved with the simplified bidding language.

$E(X_{\text{rel}}^*)$ is calculated as

$$E(X_{\text{rel}}^*) := \frac{\pi(X, P_{\text{pay}}) + \pi_{\text{all}}(X, P_{\text{pay}})}{\pi(X_{\text{rel}}^*, P_{\text{pay}}) + \pi_{\text{all}}(X_{\text{rel}}^*, P_{\text{pay}})} \in [0, 1] \quad (2)$$

We also report the revenue distribution, which shows how the overall economic value is distributed between the auctioneer and bidders. The auctioneer revenue is measured as the ratio of the auctioneer's revenue to the total sum of valuations of an efficient allocation X^* :

$$R(X) := \frac{\pi(X, P_{\text{pay}})}{\pi(X^*, P_{\text{pay}}) + \pi_{\text{all}}(X^*, P_{\text{pay}})} \in [0, E(X)] \subset [0, 1] \quad (3)$$

Additionally, the average number of unsold items is evaluated.

Table 2. Aggregate measures of auction performance

Auction format	$E(X)$	$E(X_{\text{rel}}^*)$	$R(X)$	Unsold items
CC+	97,26%	97,60%	78,96%	0
Sealed bid	97,21%	97,55%	77,28%	0
SMRA	98,51%	98,85%	81,96%	0
CCA	89,33%	89,64%	37,41%	1.25 (5.2%)

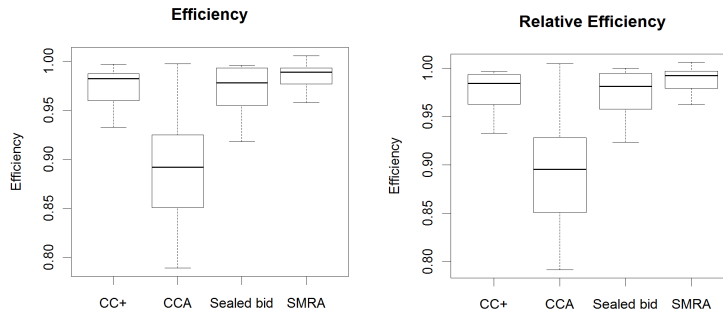


Fig. 3. Efficiency and relative efficiency

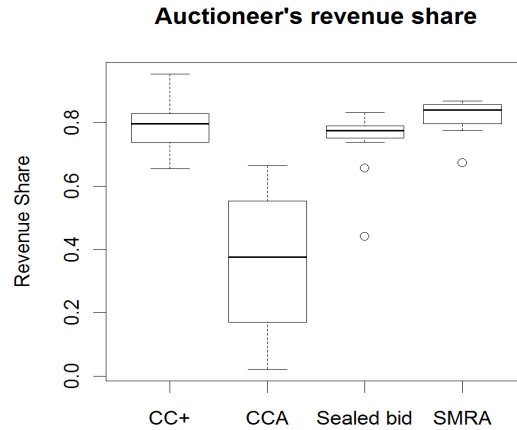


Fig. 4. Auctioneer's revenue share

Table 3. Significance tests (Wilcoxon rank sum tests) for a difference on all pairs of auction formats

Comparison	$E(\mathbf{X})$	$E(\mathbf{X}_{rel}^*)$	$R(\mathbf{X})$
CC+ vs. CCA	W=219 (p=0.000)	W=218 (p=0.000)	W=255 (p=0.000)
CC+ vs. Sealed bid	W=123 (p=0.867)	W=120 (p=0.777)	W=146.5 (p=0.498)
CC+ vs. SMRA	W=74 (p=0.044)	W=84.5 (p=0.105)	W=88 (p=0.138)
Sealed bid vs. CCA	W=255 (p=0.000)	W=218 (p=0.000)	W=219 (p=0.000)
Sealed bid vs. SMRA	W=74 (p=0.044)	W=84.5 (p=0.105)	W=88 (p=0.138)
CCA vs. SMRA	W=30 (p=0.000)	W=9 (p=0.000)	W=0 (p=0.000)

Support for the first three results is presented in figure 3, table 2 and 3.

Result 1: *The efficiency of SMRA, CC+ and the sealed bid format was not significantly different.*

In all the three auction formats we achieved high efficiency, on average higher than 97%. All items were sold and we did not lose social welfare by not assigning some items.

Result 2: *The efficiency of the CCA was significantly lower than that of the other auction formats.*

Bidders' choice set in the value model consisted of 2,400 different bundles, because bidders had to combine the number of items across all bands in each bundle bid. It is very difficult for human bidders to find the most valuable packages in this scale.

Result 3: *By using the simplified bidding language we do not lose efficiency.*

By restricting our bidding language we do not lose efficiency in comparison to the fully expressivity. We see, that $E(\mathbf{X})$ and $E(\mathbf{X}_{rel}^*)$ are only different by ca. 0.3%.

One reason is the complementarity of the value model. Besides, the most efficient allocations consist of combinations of two, four and rarely six sized bundles. As a consequence, we simplify the language from 2,400 possible bids to at most 12 bids significantly without losing efficiency.

Result 4: *The auctioneer revenue of SMRA, CC+ and the sealed bid auction was significantly higher than that of the CCA.*

Support for result 4 and 5 can be found in figure 4 and table 2 and 3. A reason for this is that only within CCA items remained unsold. So, the auctioneer did not earn money for each block and missed revenue.

Result 5: *The auctioneer's revenue of SMRA, CC+ and the sealed bid auction was not significantly different.*

This result is remarkable, because in SMRA a different payment rule was applied as within CC+ and the sealed bid auction. Nevertheless, the revenue is in the same range.

Now, we describe bidder behavior in CC+ and the sealed bid auction at first.

4.2 Bidding Behavior in the Sealed Bid Auction

To analyze bidding behavior, we looked at bid shading. Shading is computed by the difference of bidders' valuation and the bid price for a certain bundle. Percental shading is shading divided by the valuation. In the following we categorize the bidders according to their strength.

Bidders with the highest valuation within a band are called very strong bidders, with the second highest valuation medium strong, with the third highest valuation medium weak and finally, bidders that drew the lowest valuation are named very weak.

Table 4. Average bid shading

Bidders' type	Absolute shading	Percental shading
Very strong	68,75	5,40%
Medium strong	26,09	2,40%
Medium weak	31,66	2,86%
Very weak	2,77	0,47%

Result 6: *In the sealed bid auction both the absolute and the percental bid shading of the very strong bidders is significantly higher than that of the others bidders.*

The higher the valuation was the bidder has drawn within a block; the more he shaded the bid. Bidders did not want to lose when they evaluated their situation as strong. Therefore, they placed bids above their valuation to assure their win.

Table 5. Wilcoxon tests for bid shading

	Very strong	Medium strong	Medium weak	Very weak
Very strong	-	W=29462 (p=0.046)	W=28483 (p=0.114)	W=30348 (p=0.001)
Medium strong	-	-	W=22113 (p=0.709)	W=23740 (p=0.223)
Medium weak	-	-	-	W=23778 (p=0.1303)
Very weak	-	-	-	-

Result 7: *In the sealed bid auction both the absolute and the percental bid shading of the medium strong and medium weak is not significantly different.*

The valuations, the medium strong and medium weak bidders drew, were sometimes not far apart. Bidders could not exactly determine their strength compared to their competitors in such situations. As a consequence, the behavior of these two kinds of bidders did not differ.

Result 8: *The weak bidder did not shade his bids.*

The weak bidder knew that he might be the one with the lowest valuation. Thus, he did not apply bid shading in order to avoid the risk to make a loss.

Support for Result 6, 7 and 8 is given in table 4 and 5. In Figure 5 we made linear interpolations of bidders' shading to get an overview of their actual activities. If a bid is not connected, bidders updated it during the auction and it was not yet active for winner determination. We see that the bid shading of different bidders is quite different.

The literature on explaining this is huge and beyond what we can describe in this context. Overall, risk aversion, regret, spite, and wrong expectations about the bids of others are the most common conjectures.

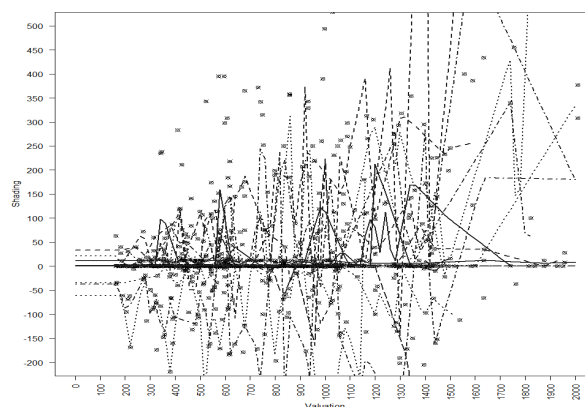


Fig. 5. Shading in the sealed bid auction

4.3 Bidding Behavior in the CC+ Auction

For the following analysis it is of interest if bidders applied their equilibrium bidding strategy for the CC+. Besides, a remark is given how bidders submitted their bids.

Result 9: *In the CC+ auction, bidders followed a powerset strategy.*

Bidders placed always their powerset bids in the CC+ auction at least in the first round. This means 48 bids, since 4 bidders participated in each auction who could submit bids on all the 12 possible bundles.

On average, CC+ consisted of 18.75 rounds and 367.75 bids have been submitted in each auction until the weakest bidder had reached his valuation and dropped out of the auction. The number of winning bidders was on average 3.8 (in 13 out of 16 auctions all four bidders won at least one bid). Support can also be found in figure 6.

Result 10: *Bidders placed their bids on their best bundles.*

Bidders focused on their best bundle, because there was the highest valuation per unit. So, 39.51% of all bids have been for bids on 2 sized bundles, 32.27% for bids on 4 sized and finally, 28.22% on 6 sized bundles. Support can be found on figure 7. To get an impression of the scale of the submitted bids we talk, in figure 7 and 8 the numbers and not percental figures are used.

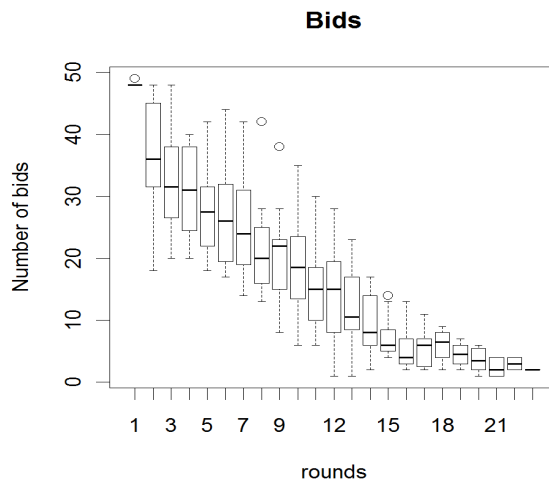


Fig. 6. Number of bids in each round

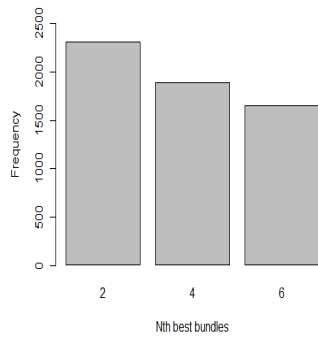


Fig. 7. Distribution of bids

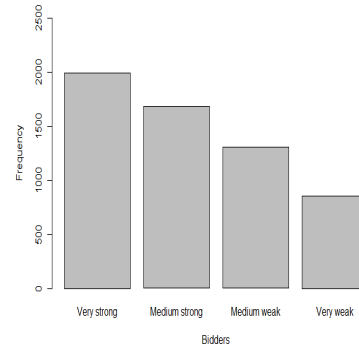


Fig. 8. Bids according to bidders' strength

Result 11: *The number of bids depends on bidders' strength.*

In percentage, 34.17% of all bids submitted the very strong bidder, 28.82% the medium strong bidder, 22.36% the medium weak bidder and 14.65% the weakest bidder. If prices are too high, weak bidders cannot participate in the auction. Later in the auction, even prices for the medium weak or sometimes the medium strong bidder increased too much.

5 Conclusion

Combinatorial auctions have led to a substantial amount of research and found a number of applications in high-stakes auctions for industrial procurement, logistics, energy trading, and the sale of spectrum licenses. Thereby, communication complexity is a problem in all CAs. The question is important, how many bids must be submitted to the auctioneer to calculate an efficient allocation. Hence, we simplified the bidding language using our knowledge of the value model. Bidders only had to submit 12 bids in each round respectively in the whole sealed bid auction in comparison to evaluate all possible 2,400 bundles in the CCA. Besides, we used for CC+ and the sealed bid auction an OR bidding language between bands, which leads to the fact that bidders need to submit much less bids as with a pure XOR bidding language in CCA. Efficiency is not significantly decreased by this modification.

Single-phase auction formats where bidders can submit more than a single bid per round can have advantages and elicit more "relevant" bundle bids throughout the process. So, we tested the two phases of the CCA, the CC+ and the sealed bid auction in lab experiments, separately. We found out, that these single-phase CAs achieved more robust outcomes and realized significantly higher efficiency and revenue than the CCA. However, there are almost no differences concerning the efficiency and revenue between the sealed bid and CC+ auction. But, the SMRA remained the most efficient auction format in our experimental setting. Nevertheless, there are many disadvantages to SMRA like the exposure problem, which says that in case of com-

plementarities, bidders run the risk of winning only a part of a complementary collection of items in an auction without package bids. Besides, in order to maintain eligible bidders temporarily bid for packages they are not interested in which can provide less efficient outcomes. Bidders can also use signaling such as jump bidding to cooperate in SMRA. These reasons are only exemplarily, why it is not proper to use it SMRA in practice.

We argue that in addition to theoretical research, more empirical work is needed, in order to understand real-world bidder behavior. Additionally, the CCA could also be tested with the simplified bidding language and compared with the results obtained by our tested single phase auctions.

References

1. Adomavicius, D., Gupta, A.: Towards comprehensive real-time bidder support in iterative combinatorial auctions. *Information Systems Research (ISR)* (2005)
2. Ausubel, L., Crampton, P., Milgrom, P.: The clock-proxy auction: A practical combinatorial auction design. In: Cramton, P., Shoham, Y., Steinberg, R. (eds.): *Combinatorial Auctions*. MIT Press, Cambridge, MA (2006)
3. Ausubel, L., Milgrom, P.: The lovely but lonely vickrey auction. In: Cramton, P., Shoham, Y., Steinberg, R. (eds.): *Combinatorial Auctions*. MIT Press, Cambridge, MA (2006)
4. Bapna, R., Goes, P., Gupta, A.: A theoretical and empirical investigation of multi-item online auctions. *Information Technology and Management* 1 (1-2) (2000)
5. Bapna, R., Goes, P., Gupta, A.: Insights and analyses of online auctions. *Communications of the ACM* 44 (11), 42-50 (2001)
6. Bichler, M., Shabalin, P., Pikhovskiy, A.: A computational analysis of linear-price iterative combinatorial auctions. *Information Systems Research* 20 (1) (2009)
7. Bichler, M., Gupta, A., Ketter W.: Designing smart markets. *Information Systems Research* 21 (4), 688-699 (2010)
8. Bichler, M., Shabalin, P., Wolf, J.: Efficiency, auctioneer revenue, and bidding behavior in the combinatorial clock auction. In: *The Second Conference on Auctions, Market Mechanisms and Their Applications*, New York. ACM (2011)
9. Bichler, M., Shabalin, P., Ziegler, G.: Efficiency with linear prices? A theoretical and experimental analysis of the combinatorial clock auction. *INFORMS Information Systems Research* (2012)
10. Cramton, P., Shoham, Y., Steinberg, R.: *Introduction to combinatorial auctions* (2006)
11. Cramton, P., Shoham, Y., Steinberg, R.: *Combinatorial Auctions*. MIT Press, Cambridge, MA (2006)
12. Cramton, P.: Spectrum auction design. *Papers of Peter Cramton 09sad*, University of Maryland, Department of Economics - Peter Cramton. <http://ideas.repec.org/p/CCAc/CCAcumd/09sad.html>. (2009)
13. Day, R., Milgrom, P.: Core-selecting package auctions. *International Journal of Game Theory* 36, 393-407 (2007)
14. Day, R., Raghavan, S.: Fair payments for efficient allocations in public sector combinatorial auctions. *Management Science* 53, 1389 -1406 (2007)
15. Day, R., Cramton, P.: The quadratic core-selecting payment rule for combinatorial auctions (2008)

16. Guo, Z., Koehler, G.J., Whinston, A.B.: A market-based optimization algorithm for distributed systems. *Management Science* 53, 1345-1358 (2007)
17. Maldoom, D.: Winner determination and second pricing algorithms for combinatorial clock auctions. Discussion paper 07/01, dotEcon (2007)
18. Milgrom, P.: Putting auction theory to work: The simultaneous ascending auction. *Journal of Political Economy* 108 (21), 245-272 (2000)
19. Perez-Richet, E.: A Note on the Tight Simplification of Mechanisms. *Economics Letters* 110 (1), 15-17 (2011)
20. Porter, D., Rassenti, S., Roopnarine, A., Smith, V.: Combinatorial auction design. In: Proceedings of the National Academy of Sciences of the United States of America (PNAS) 100, pp. 11153-11157 (2003)
21. Porter, D., Smith, V.: Fcc license auction design: A 12-year experiment. *Journal of Law Economics and Policy* 3 (2006)
22. Scheffel, T., Ziegler, G., Bichler, M.: On the impact of cognitive limits in combinatorial auctions: An experimental study in the context of spectrum auction design. In: Jahrestagung der Gesellschaft für experimentelle Wirtschaftsforschung e.V., Luxembourg (2010)

Student Consortium

Jan vom Brocke

Universität Liechtenstein, Vaduz, Liechtenstein

Ziel des Consortiums ist es, Studierenden die Möglichkeit zu eröffnen, sich aktiv in die Internationale Tagung Wirtschaftsinformatik einzubringen. Angesprochen werden Studierende auf Master-Ebene, die sich entweder für eine IT-bezogene Position in der Wirtschaft oder für eine Promotion in der Wirtschaftsinformatik interessieren. Das Consortium soll eine Plattform bieten, über die sich Studierende, Wissenschaftler und Praktiker begegnen und im fachlichen Diskurs intensiv kennenlernen.

Studierende aller Standorte und Fachrichtungen der Wirtschaftsinformatik wurden aufgerufen, qualitativ hochwertige Arbeiten einzureichen. Die Beiträge wurden von mindestens zwei unabhängigen Mitgliedern des internationalen Programmkomitees begutachtet. Eingereicht wurden sowohl Work-in-Progress-Paper als auch kürzlich abgeschlossene Qualifikationsarbeiten. Eine Mitwirkung der Betreuer als Co-Autoren wurde dabei begrüßt. Ferner wurden Gruppenarbeiten von mehreren Studierenden berücksichtigt. Weitere vielversprechende Arbeiten sind als Poster-Präsentationen akzeptiert worden.

Mein besonderer Dank gilt den zahlreichen Kollegen, die im Programmkomitee mitgewirkt haben. Mein herzlicher Dank gilt auch meinem Mitarbeiter Bernd Schenk für die sehr professionelle Organisation des Consortiums.

Die Hilti AG hat einen Best Paper Award ausgelobt, für den vier Arbeiten nominiert worden sind. Dies sind in alphabetischer Reihenfolge der Erstautoren die Arbeiten von Daniel Braunnagel und Florian Johannsen zu „Coupling Metrics for EPC Models“ von der Universität Regensburg, von Christian Reuter und Michael Ritzkatis zur „Unterstützung mobiler Geo-Kollaboration zur Lagebeurteilung von Feuerwehr und Polizei“ von der Universität Siegen, von Jörg Schneider und Markus Reiter zu „Green IT-Service-Management: Eine empirische Voruntersuchung der konzeptionellen Grundlagen“ von der Universität des Saarlandes und von Sanja Tumbas und Theresa Schmiedel zu „Developing an Organizational Culture Supportive of Business Process Management“ von den Universitäten Münster und Liechtenstein. Die Vergabe des Preises wird durch ein Komitee auf der Konferenz erfolgen.

Ich möchte allen Studierenden, die ihre Arbeiten auf dem Student Consortium präsentieren werden, noch einmal herzlich gratulieren. Der Begutachtungsprozess war sehr kompetitiv. Zugleich empfehle ich alle Beiträge des Consortiums sehr zur Lektüre. Sie enthalten viele spannende Anregungen für zukünftige Forschungsarbeiten.

Programmkomitee

Ulrike Baumöl, FernUniversität Hagen
Jörg Becker, Universität Münster
Hans Ulrich Buhl, Universität Augsburg
Werner Esswein, Technische Universität Dresden
Torsten Eymann, Universität Bayreuth
Ulrich Frank, Universität Duisburg-Essen
Norbert Gronau, Universität Potsdam
Thomas Hess, Ludwig Maximilians Universität München
Roland Holten, Universität Frankfurt
Reinhard Jung, Universität St. Gallen
Dimitris Karagiannis, Universität Wien
Ulrike Lechner, Universität der Bundeswehr München
Christine Legner, Université de Lausanne
Susanne Leist, Universität Regensburg
Claudia Loebbecke, Universität zu Köln
Peter Loos, Universität des Saarlandes
Ronald Maier, Universität Innsbruck
Florian Matthes, Technische Universität München
Jan Mendling, Wirtschaftsuniversität Wien
Kathrin Möslein, Universität Erlangen-Nürnberg
Markus Nüttgens, Universität Hamburg
René Riedl, Universität Linz
Detlef Schoder, Universität Köln
Alfred Taudes, Wirtschaftsuniversität Wien
Christof Weinhardt, Technische Universität Karlsruhe

Erfolgsfaktoren von On-Demand-Enterprise-Systemen aus der Sicht des Anbieters – eine explorative Studie

Alexander Wieneke¹, Sebastian Walther¹, Rüdiger Eichin², und Torsten Eymann¹

¹ University of Bayreuth, Chair of Information Systems Management, Bayreuth, Germany
{alexander.wieneke,s.walther,torsteneymann}@uni-bayreuth.de

² SAP AG, Product Management, Walldorf, Germany
ruediger.eichin@sap.com

Abstract. Diese Arbeit thematisiert erfolgsrelevante Aspekte für On-Demand-Enterprise-Systeme aus Sicht eines Serviceanbieters. Auf Basis der Forschungsfrage: „Inwiefern bieten in Abgrenzung zu konventionellen On-Premise-Lösungen On-Demand-Enterprise-Systeme Vorteile für Unternehmen“ reduziert die Studie die Wissenslücke bezüglich der Erfolgsfaktoren dieses speziellen Serviceangebotes. Anhand 12 explorativer Interviews mit Experten des führenden deutschen Softwareherstellers werden 17 Erfolgsfaktoren identifiziert und detailliert erläutert. Zur Kategorisierung der Erfolgsfaktoren werden diese dem „IS-Success Model“ von DeLone und McLean zugeordnet. Dabei wird verdeutlicht, dass das Potenzial von On-Demand-Enterprise-Systemen sich nicht nur auf reine Kostensenkungsmaßnahmen reduzieren lässt. Wesentliche Erfolgsfaktoren finden sich vielmehr in dem Wirkungsfeld von höherer Flexibilität und Einfachheit. Dem Kunden respektive den Unternehmungen können sich somit durch den Einsatz von Cloud-Applikationen neue strategische Optionen bieten.

Keywords: Cloud Computing, Software-as-a-Service, SaaS, Enterprise System

1 Einleitung

Cloud Computing als ein neues Modell der IT-Dienstleistung erfährt aktuell in der IT-Branche eine außerordentliche Aufmerksamkeit. Die Bereitstellung von Hard- und Softwarenutzungsangeboten über das Internet wird in der Fachwelt bereits als Paradigmenwechsel der IT bezeichnet [1]. Entsprechende Dienstleistungen weisen dabei ein hohes wirtschaftliches Potenzial auf. Laut Gartner wurden mit „Software-as-a-Service“ (SaaS) im Jahr 2011 mit einer Vorjahressteigerung von rund 21 Prozent Erlöse in Höhe von 12 Mrd. Dollar erwirtschaftet. Bis 2015 wird ein Wachstum des Cloud-Marktes auf 21 Mrd. Dollar prognostiziert [2].

Der ausschlaggebende Faktor für die hohe Akzeptanz des Cloud Computing in verschiedenen Wirtschaftsbereichen ist nach Ansicht einiger Wissenschaftler und Praktiker die rein betriebswirtschaftliche Auffassung, wonach unternehmensbezogene IT-Lösungen in erster Linie instrumentell zur Steigerung der Kosten- und Prozesseffektivität

vität beitragen sollen [3]. Die Ergebnisse dieser Studie zeigen jedoch, dass Cloud-basierte-Dienstleistungen für die Unternehmen nicht nur eine Kostensenkungsmaßnahme darstellen, sondern eine Bandbreite von Systemmerkmalen und strategischen Gesichtspunkten positiv beeinflussen. Im Gegensatz zu einer Vielzahl wissenschaftlicher Untersuchungen zu Erfolgsfaktoren im Segment konventioneller Enterprise-Systeme (z. B. [4]), finden sich im Bereich des Cloud-Computing zum Thema Erfolgsfaktoren lediglich Studien zu nicht-spezifizierten SaaS-Lösungen (s. z. B. [5]). Wissenschaftlich verifizierte Ergebnisse über die Erfolgstreiber von On-Demand-Enterprise-Lösungen¹ liegen nicht vor.

Inhaltspunkte für die Erfolgsfaktoren entsprechender Dienstleistungen stellen lediglich die von Cloud-Anbietern herausgegebenen „Marketingbroschüren“ dar, die jedoch eine stark verzerrte Sichtweise liefern (z. B. [6]). Diese Wissenslücke gilt es im Rahmen der vorliegenden Studie anhand einer qualitativ-empirischen Untersuchung durch die Beantwortung folgender Forschungsfrage zu schließen:

- *„Inwiefern bieten in Abgrenzung zu den konventionellen On-Premise-Lösungen On-Demand-Enterprise-Systeme Vorteile für Unternehmen?“*

In Anlehnung an die Aussage von Vargo und Lusch: „The Enterprise can only make Value Propositions“ [7] werden durch explorative Interviews mit Experten eines weltweit führenden deutschen Softwareherstellers jene Erfolgsfaktoren identifiziert, die nach Auffassung eines Cloud-Anbieters der Kunde respektive Nutzer des On-Demand-Enterprise-Systems für eine Systementscheidung als ausschlaggebend ansieht.

Die vorliegende Arbeit ist wie folgt aufgebaut: Zunächst werden in Abschnitt 2 die theoretischen Grundlagen der Studie dargestellt. Dabei wird insbesondere auf die Erfolgsmessung von konventionellen IT-Systemen sowie moderner SaaS-Lösungen näher eingegangen. Im dritten Abschnitt wird die Methodik des Forschungsprozesses erläutert. Kapitel 4 behandelt die Studienergebnisse. Es werden 17 Erfolgsfaktoren identifiziert und gemäß den Dimensionen des „IS-Success-Models“ von DeLone und McLean kategorisiert. Abschließend werden die Ergebnisse zusammengefasst und kritisch betrachtet.

Mit Bezug auf das „IS-Success-Model“ von DeLone und McLean zeigt die Studie, dass die Erfolgsfaktoren von On-Demand-Enterprise-Lösung sich hauptsächlich in der Ausgestaltung der Systemqualitäten, sowie in den organisatorischen Vorteilen (Net-Benefits) von klassischen On-Premise Lösungen unterscheiden. Neben den Vorteilen wie Kostenreduktion und Systemflexibilität werden cloud-spezifische Faktoren wie Ubiquität und Flexibilität als Erfolgstreiber identifiziert.

Die Arbeit kann weiterführend dazu verwendet werden, die vom IS-Success Model vorgeschlagenen Erfolgskategorien formativ im Kontext von On-Demand-Enterprise

¹ On-Demand Enterprise Systeme sind eine spezielle Form von SaaS. Dabei wird das Enterprise System (die Software) über eine Cloud Infrastruktur angeboten. On-Demand Enterprise Systeme haben spezifische Eigenschaften, die sich weder durch wissenschaftliche Arbeiten im Bereich der Enterprise Systeme, noch im Bereich von SaaS komplett abdecken lassen.

Systems zu operationalisieren, um damit eine empirische Verifikation der Ergebnisse zu ermöglichen.

2 Theoretischer Hintergrund

2.1 Definition grundlegender technischer Begriffe

Der Begriff Cloud Computing ist ein weitreichender Begriff, dessen Gebrauch häufig Verwirrung stiftet und auch von Praktikern oftmals falsch verstanden wird [11-12]. „Die Klärung der Cloud Definition“ [12] ist deshalb für das wissenschaftliche Vorgehen unerlässlich.

Laut Definition des National Institute of Standards and Technology bietet Cloud Computing dem Nutzer über ein On-Demand-Netzwerk einen Zugang zu einem gemeinsamen Pool konfigurierbaren Computer-Ressourcen [13]. Damit verbinden sich die drei verschiedenen Servicemodelle Plattform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS) und Software-as-a-Service (SaaS) [13]. Die für diese Arbeit relevante Bereitstellung von On-Demand-Enterprise-Anwendungen wird unter dem Begriff „Software-as-a-Service“ (SaaS) subsumiert. Unter diesem Begriff sind Softwaredienstleistungen zu verstehen, die es Kunden ermöglichen, die von dem Anbieter bereitgestellten Applikationen über eine Cloud-Infrastruktur zu nutzen und auf diese über verschiedenen Client-Devices zuzugreifen [13].

Der Terminus „On Demand“ ist eine signifikantes Merkmal des Cloud Computings. Er unterstreicht das Cloud-Spezifikum, IT-Applikationen, IT-Dienstleistungen oder weitere Computer- bzw. Netzleistungen dem Kunden bedarfsgerecht durch den Serviceanbieter umgehend bereitzustellen [14].

Konträr zu On Demand steht der Begriff On Premise. In einem On-Premise-Modell erwirbt der Kunde die Applikation und betreibt und verantwortet diese eigenständig mit sämtlichen betriebstechnischen Implikationen (Pflege und Wartung der Anwendungen, Systeme und Infrastrukturen) [15].

2.2 Erfolgsmessung von Informationssystemen

Die Diskussion der Erfolgsfaktoren von Informationssystemen beginnt bereits mit der Einführung von IT-Systemen in Unternehmen. Um die hohen Investitionen in moderne Informationssysteme zu rechtfertigen, muss der Erfolg der neuen Technologie bzw. der positive Einfluss auf die Unternehmung messbar sein [4].

Eine weitverbreitete Methode zur IT-Erfolgsmessung besteht in der Erstellung mehrdimensionaler Modelle, die die Evaluationsperspektiven verschiedener Stakeholder abbilden. Von besonderer Relevanz ist dabei die Frage, welchen zukünftigen Nutzen und welche Vorteile eine Unternehmung auf Grund verschiedener Qualitätsaspekte des Informationssystems erwarten kann (zukunftsbezogene Betrachtung) [4].

Diese Methodik und Fragestellungen bilden ebenfalls die Grundlage für das in der Fachwelt am meisten zitierte und breit akzeptierte [8] „IS-Success-Model“ von

DeLone und McLean ([9-10]): Deren Ansätze sollen deshalb als wissenschaftliche Grundlage auch für diese Untersuchung herangezogen und nachfolgend in aller Kürze skizziert werden.

Das überarbeitete Modell beschreibt die Messung von „IS-Success“ anhand sechs verschiedener Dimensionen [10].

- „System Quality“: Erwünschte Charakteristika und Features des Informationssystems [9]. (Bspw. System Efficiency)
- „Information Quality“: Erwünschte Charakteristika und Features des Produktes, also der Information [9]. (Bspw. Sufficiency)
- „Service Quality“: Erwünschte Qualität der Unterstützung, die dem Kunden vom IS-Dienstleister angeboten wird [9]. (Bspw.: Responsiveness)
- „System Use“: Erfüllung des erwünschten Verwendungszwecks eines Informationssystems [9]. (Bspw.: Number of Inquiries)
- „User Satisfaction“: Level der Zufriedenheit als Konsequenz der Interaktion des Kunden mit dem System [9]. (Bspw. Overall Satisfaction)
- „Net Benefits“: Ausmaß (positiv und negativ), in dem ein System zur Verbesserung einer Sache beiträgt [9]. (Bspw. Cost Reduction)

Darüber hinaus beschreiben DeLone und McLean die Entstehung von „IS-Success“ als einen Prozess resp. ein Verfahren, da die Dimensionen untereinander verschiedene Kausalzusammenhänge aufzeigen [10]. Die nachfolgende Abbildung zeigt die Dimensionen, eingeordnet in den Prozess der Erfolgseinstellung, den zeitlichen Ablauf eines Informationsflusses sowie dessen Einwirkungen nach angeordnet [10].

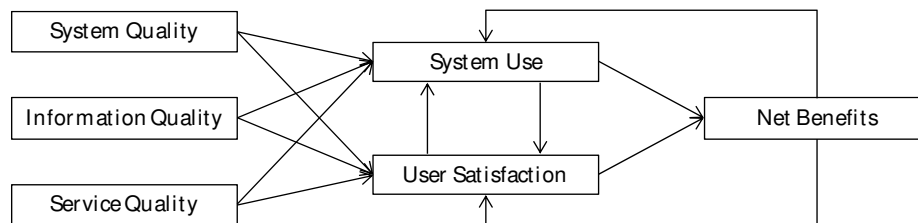


Abb. 1. IS-Success-Modell von DeLone und McLean [10]

2.3 Definition kritischer Erfolgsfaktoren und Nutzenversprechen

Für das weitere Vorgehen der Studie ist die Definition sowie die Zusammenführung der Termini „Erfolgsfaktor“ („Success Factor“) und „Nutzenversprechen“ („Value Proposition“) wesentlich.

Ein Erfolgsfaktor ist der kritisch zu betrachtende Aspekt einer Unternehmung oder eines Produktes respektive Dienstleistung, der bei entsprechender Entwicklung und Ausführung einen Wettbewerbsvorteil generiert und somit den Erfolg des Unternehmens steigert. Value Propositions hingegen sind jene Faktoren einer Dienstleistung, auf deren Basis der Kunde entscheidet, ob ein Serviceangebot seine Bedürfnisse befriedigt und für ihn einen Mehrwert darstellt [7].

Da ein Erfolgsfaktor bei korrekter Ausführung den Kunden einen Mehrwert verschafft, kann dieser in dem Fall auch als Nutzenversprechen angesehen werden. Da gemäß des „Service-Centered View of Marketing“ dem Kunden die richtigen Nutzenversprechen zu unterbreiten sind, um sich als Unternehmung am Markt zu etablieren und Wettbewerbsvorteile zu erlangen [7], ist der Begriff Nutzenversprechen ebenfalls vergleichbar mit dem Begriff Erfolgsfaktor. Es bleibt somit zu konstatieren, dass sowohl Erfolgsfaktoren als auch Nutzenversprechen unerlässliche Prädiktoren des Erfolgs von On-Demand-Enterprise-Systemen sind, so dass im Rahmen dieser Arbeit keine Differenzierung vorgenommen wird, somit unter dem Begriff „Erfolgsfaktor“ beide zuvor genannten Termini zusammengefasst werden.

3 Zur Untersuchungsmethodik

Um die Frage: „What makes new services successful and what causes failure“ [16], für On-Demand-Enterprise-Systeme zu beantworten, versucht die Studie jene Erfolgsfaktoren zu identifizieren, die in der Einschätzung von Akteuren eines On-Demand-Enterprise-Anbieters für eine erfolgreiche Marktablierung des Angebotes unerlässlich sind und die nach deren Einschätzung einen Mehrwert beim Kunden generieren. Letztendlich stellt sich die Frage, welchen Einfluss ein On-Demand-Enterprise-System auf das Unternehmen hat bzw. auf Grund welcher Qualitätsaspekte haben könnte. Ein On-Demand-Enterprise-Anbieter sollte demnach folgende Fragen beantworten können:

- Warum entscheidet sich der Kunde für ein On-Demand-Enterprise-Angebot und nicht für eine konventionelle On-Premise-Lösung?
- Welche positiven Einflüsse hat ein On-Demand-Enterprise-System auf die Unternehmung?
- Welche Qualitätsaspekte sieht der Kunde als entscheidend für die Mehrwertgenerierung an?
- Wie lässt der Kunde sich bei der Entscheidungsfindung beeinflussen?

Zur Rekonstruktion des Wissens respektive der Sichten der untersuchten Akteure wird auf die qualitative Methode des explorativen Experteninterviews zurückgegriffen und bei Konzeption und Umsetzung der Befragung auf die von Gläsel und Laudel vorgeschlagene Verfahrensweise Bezug genommen [17].

Zur Datenerhebung wurden insgesamt 12 Interviews mit Experten eines führenden Softwareherstellers² von On-Demand-Enterprise-Lösungen in Deutschland geführt. Mit der Absicht, ein möglichst breites Wissensspektrum zu erschließen, wurden Experten aus unterschiedlichen Bereichen der Unternehmung und mit unterschiedlichen Aufgaben im On-Demand-Enterprise-Umfeld ausgewählt. Die Probanden kamen unter anderem aus den Bereichen Entwicklung, Strategie, Produktmanagement sowie Vertrieb und gehörten unterschiedlichen hierarchischen Ebenen an. (vgl. Tabelle 1)

² Das Unternehmen verbuchte im Jahr 2011 einen Umsatz i. H. v. 14 Mrd. Euro und führte rund 60 tausend Mitarbeiter weltweit.

Tabelle 1. demografische Variablen

Interview	Abteilung	Alter	Geschlecht
Interview 1	Presale	40-50	Männlich
Interview 2	Vertriebsorganisation	30-40	Weiblich
Interview 3	Entwicklung	30-40	Männlich
Interview 4	Supply Chain Management	50-60	Männlich
Interview 5	Product Management	20-30	Männlich
Interview 6	Extensibility Team	30-40	Männlich
Interview 7	Solution & Knowledge Packaging	40-50	Männlich
Interview 8	Strategy OnDemand	20-30	Männlich
Interview 9	Strategy OnDemand	30-40	Männlich
Interview 10	Strategy OnDemand	40-50	Männlich
Interview 11	Cloud Portfolio Management	50-60	Männlich
Interview 12	Risk Management	30-40	Weiblich

Als anzuwendende Interviewtechnik wurde die halbstandardisierte Form gewählt, um den explorativen Charakter dieser Studie zu wahren und trotz freier Form den Rückgriff auf gewisse Vorgaben zu ermöglichen. Diese Vorgaben sind in einem Interviewleitfaden festgehalten. Als Orientierungsrahmen dienten zum einen die in Abschnitt 2.1 dargestellten Fragestellung und einschlägige Studien und Abhandlungen, in denen die Erfolgsfaktoren für eine nachhaltige Marktdurchdringung von Cloud Computing thematisiert werden [18].

Zu Beginn des Interviews wurde mit Blick auf die geforderte Offenheit und Explorativität dem Befragten eine allgemeine, offene Frage gestellt, welche den eigentlichen inhaltlichen Aspekt zunächst außer Acht lässt. Diese dienten im weiteren Verlauf dazu, weitere „zentrale Fragen im geeigneten Moment zur Diskussion zu stellen“ [19], um den Gesprächsverlauf anzuregen. Um das Expertenwissen mit einem möglichst hohen Breitenstandard zu erschließen, wurde die Form des Einzelinterviews gewählt [19].

Wie bei explorativen Vorgehen üblich, bezieht sich der Forschungsprozess nicht auf vorab formulierte Theoriekonzepte in Form von vorgefertigten Hypothesen. Als Analysetechnik wurde deshalb die induktive Kategorienbildung auf Basis der zusammenfassenden Inhaltsanalyse von Mayring [20] angewendet, um offen und ohne Verzerrungen von theoretischen Vorüberlegungen das erhobene Datenmaterial zu bearbeiten. Im weiteren Analyseverlauf wurden abstrakte Aussagen getroffen, anhand derer sich das Material paraphrasieren lässt und die letztendlich zu Kategorien zu subsumieren sind [21].

Die generalisierten Kategorien sollen gemäß den vorab festgelegten Selektionskriterien [20] akkumulierte Aussagen über Erfolgsfaktoren des On-Demand-Enterprise-Angebots und deren Einfluss auf den Kunden darstellen.

4 Ergebnisse

Als Ergebnis der Auswertung des erhobenen Datenmaterials anhand der induktiven Kategorienbildung konnten 17 Klassen identifiziert werden.

Tabelle 2 gibt einen Überblick über die Klassen und die Interviews, in welchen erklärende, der Klasse entsprechende Aussagen zu finden sind.

Tabelle 2. Darstellung der Ergebnisse der Datenauswertung

Interview Kategorie	1	2	3	4	5	6	7	8	9	10	11	12	Summe
flexible Erweiterbarkeit	x	x	x	x	x	x	x	x	x	x	x	x	12
Kostenreduktion	x	x	x	x	x	x	x		x	x	x	x	11
höhere Verfügbarkeit	x	x	x	x	x	x	x	x		x	x	x	11
Qualität betriebswirtschaftlicher Funktionen	x	x	x	x	x	x	x	x		x	x	x	11
Konzentration auf Kernkompetenzen		x	x	x		x	x	x	x	x	x	x	10
geringer Implementierungsaufwand	x		x	x	x	x		x	x	x	x	x	10
flexible Anpassbarkeit		x	x	x				x	x		x	x	9
Updateability	x		x	x		x	x	x	x	x	x	x	8
höhere Sicherheit	x	x	x	x	x	x	x	x	x	x	x	x	8
Ubiquität		x	x		x	x			x	x	x		7
Usability	x	x		x				x		x	x	x	7
Riskoverlagerung			x				x	x		x	x	x	7
Elastizität			x	x	x	x		x		x			6
Flexibilität bei Businessentscheidungen		x	x			x	x			x			5
geringes Investitionsrisiko			x		x	x		x		x			5
Analytics		x			x			x		x	x		5
Planbare Kosten									x	x		x	3

Bei den in Tabelle 2 dargestellten Klassen handelt es sich um Aspekte, die aus der Sicht eines On-Demand-Enterprise-Anbieters für eine Mehrwertgenerierung beim Kunden ausschlaggebend sind.

Diese Qualitätsaspekte lassen sich nun durchaus als Faktoren bezeichnen, welche gemeinsam den Erfolg einer innovativen Lösung ausmachen – die Erfolgsfaktoren. Da die Klassenbildung aus der Generalisierung der Darstellung des Einzelnen stammt, werden die Definitionen aus den Erklärungen der Experten abgeleitet, so dass die Klassen und deren Einfluss, wie in Tabelle 3 dargestellt, definiert werden können:

Tabelle 3. Beschreibung der Erfolgsfaktoren (Teil 1)

Einflussfaktoren	Beschreibung
Kostenreduktion	Diese Kostenaspekte, die zuvor noch eine „Hürde für den Einsatz von traditionellen ERP-Lösungen“ darstellten, weisen nun deutliche Vorteile einer On-Demand-Nutzung auf. Kostenreduktionen ergeben sich durch den Wegfall oder die Verringerung der „Upfront Investments“ laufender und unvorhersehbarer IT-Kosten. Der Aspekt der Kostenreduktion wirkt dabei auf alle weiteren Qualitätsaspekte ein.
Geringer Implementierungsaufwand	Der grundlegende Vorteil liegt bei den Implementierungskonditionen. Die Anschaffung eigener Hardware sowie die Umsetzung personalbelastender, aufwendiger Beratungsprojekte sind nicht erforderlich, da der Implementierungsprozess deutlich an Komplexität verloren hat. Neben positiven Kosteneffekten wirkt sich dies auch auf die Flexibilität von Entscheidungsprozessen aus.
Elastizität	Der Nutzer kann das System der tatsächlichen Auslastung adäquat zur Nachfrage anpassen. Die Auslastung bezieht sich auf die variierende, anforderungsgerechte Bereitstellung von technischen Systemressourcen sowie auf die Anzahl der Lizenzen. Mit dem elastischen, nutzungsgerechten Systembereitstellungs- und Preismodell gewinnt der Kunde in seinen unternehmerischen Entscheidungen und seinem Handeln an Flexibilität. Ein entscheidender Aspekt ist, den bisher als unsicher geltenden, kapitalbindenden Fix-Kostenfaktor IT durch einen konstanten und planbaren Mietzins zu ersetzen.
Ubiquität	Unter Ubiquität ist der Aspekt des allgegenwärtigen Systems zu verstehen. Dieser Aspekt baut auf den Grundgedanken von Cloud Computing auf, Software wie Strom aus der Steckdose beziehen zu können. Auf Grund der nicht „backendorientierten Programmierweise“ kommt das System ohne Middleware aus, weshalb es für die Kunden und das System nicht relevant ist, von welchem Endgerät aus sie zugreifen. Der Vorteil für den Kunden ist also der flexible, einfache und ortsunabhängige Zugriff auf die Funktionen und Informationen des IT-Systems.
Planbare Kosten	Die Finanzierung des On-Demand-Angebots durch den Kunden erfolgt auf Grundlage periodisch unterschiedlicher Mietmodelle. Verbunden mit den dargestellten geringen Implementierungskosten für eine Softwarenutzung und einem elastischen, bedarfsbezogenen Preismodell fallen im Gegensatz zu On-Premise-Systemen für die IT keine unvorhersehbaren Kosten an.

Tabelle 3. Beschreibung der Erfolgsfaktoren (Teil 2)

Einflussfaktoren	Beschreibung
Flexibilität bei „Business Entscheidungen“	Die Flexibilität bei „Business Entscheidungen“ resultiert zum einen aus dem Aspekt der Ubiquität und zum anderen aus der schnellen Aktivierung bzw. Implementierung der On-Demand-Anwendung in die Unternehmensumwelt. Essenzielle Entscheidungen zum Unternehmenswachstum wie internationale Standorterweiterung, Einbindung oder der Zukauf von „Subsidiaries“, verlieren an Komplexität, da sie nun unabhängig von IT-Implementierungsaspekten getroffen und zeitnah realisiert werden können. Die Prozesse einer gesamten Organisation können somit nahezu ohne Unterbrechung an allen Standorten reibungslos und umfassend abgedeckt werden.
Geringes Investitionsrisiko	Da bei der Implementierung sehr geringe Anfangsinvestitionen anfallen und der Kunde sich lediglich über einen Mietvertrag an den Anbieter bindet, ist die Abhängigkeit des Kunden von dem Anbieter und von der Weiternutzung des jeweiligen On-Demand-Angebots sehr gering. Somit lassen sich Entscheidungen über die Anschaffung resp. Nutzung einer Software einfacher und schneller treffen und ebenso flexibel revidieren. Eine weiterer positiver Effekt der deutlich geringeren Investitionsrisiken sind die geringen Aufwände für die üblicherweise für eine Systemauswahl notwendigen, vorlaufenden Testaktivitäten.
Qualität betriebswirtschaftlicher Prozesse	Auf Grund der Kostenreduktion kann der Kunde einen Funktionsstandard für die Abdeckung betriebswirtschaftlicher Prozesse nutzen, den er sich zuvor nicht leisten konnte. Damit steigt auch das betriebswirtschaftliche Niveau, da mit dem erweiterten Funktionsset neue Möglichkeiten in der Prozessbearbeitung zur Verfügung stehen. Allerdings wird eine On-Demand-Lösung nicht den tiefen und variationsreichen funktionalen Umfang bieten wie On-Premise-Lösungen. Auf Basis der standardisierten Prozesse steigt jedoch die Effizienz, Brüche innerhalb von Prozessen werden entfernt, wodurch die Betriebsabläufe saubere Strukturen aufzeigen.
Analytics	Auf Grund der in den On-Demand-Angeboten zusammengeführten OLTP- und OLAP-Systemen wird die Untergliederung in transaktionale und analytische Seite aufgehoben. Dem Kunden bietet sich für eine bessere Entscheidungsunterstützung nun Nutzungsmöglichkeiten relevanter Analysetools in sämtlichen Funktionen. Damit entsteht eine Transparenz der den Unternehmen zugrunde liegenden betriebswirtschaftliche Prozesse und der wesentlichen Kennzahlen.

Tabelle 3. Beschreibung der Erfolgsfaktoren (Teil 3)

Einflussfaktoren	Beschreibung
Flexible system- übergreifende Erweiterbarkeit	Entscheidend für Systemerweiterungen ist der Vergleich von zeitlichem und monetärem Aufwand. Sind für die Erweiterungen von On-Premise-Systemen Spezialisten und eine Vielzahl an „Beratertagen“ zu buchen, bietet ein On-Demand-Angebot die Option des einfachen Zukaufs von Add-On Lösungen über das Internet und die Möglichkeit, diese selbstständig in das System zu implementieren.
Flexible lösungs- interne Anpass- barkeit	Es handelt sich bei den On-Demand-Angeboten um ein Low Touch Modell, das darauf abzielt, einfach, adaptierbar und änderbar zu sein. Nicht die Möglichkeit von Systemanpassungen ist ein Spezifikum, sondern deren Einfachheit und ihr funktionaler Umfang. Anpassungen können vom Kunden eigenständig ohne Unterstützung durch externe Berater oder die eigene IT-Abteilung vorgenommen werden. Individuelle Gestaltung von Screens, Formularen und Reports wird ermöglicht.
Usability	Die einfache Bedienbarkeit äußert sich in erster Linie in der Nutzbarkeit der Funktionen ohne notwendige und zusätzliche Schulungen der Mitarbeiter. Das System ist so gestaltet, dass der Kunde durch Explorieren die Nutzung selbstständig erlernen kann.
Sicherheit	Trotz diverser Bedenken zur Datensicherheit ist diese - in Relation zu den Kosten und in Bezug auf die technologischen Möglichkeiten - deutlich höher als bei einer On-Premise-Lösung. Der Kunde übergibt die mit der Sicherheit verbundenen Aufgaben. Diesem stehen neben Expertenwissen, Erfahrungen und technischen Standards eine Vielzahl sicherheitsgewährleistender Einrichtungen und Verfahren zur Verfügung, die z. B. ein mittelständisches Unternehmen nur mit unverhältnismäßig hohen Kosten erreichen kann.
Updateability	Die Updateaktivitäten des Systems liegen im Zuständigkeitsbereich des Anbieters. Dieser entscheidet über die Notwendigkeit von Updates und somit über den Zeitabstand und Umfang. Für den Kunden fällt kein Aufwand an, da dieser Service in dem Mietkostenmodell inkludiert ist. Dem Vorteil eines ständig aktualisierten Systems steht allerdings der Nachteil gegenüber, sich in kurzen Zyklen auf Veränderungen einstellen zu müssen. Insoweit bleibt es die Herausforderung an den Anbieter, Updates für den Kunden möglichst „schmerzfrei“ einzuführen.
Höhere Verfüg- barkeit	Für die Gewährleistung einer hohen Datenverfügbarkeit stehen großen Data-Center-Betreibern im Vergleich zum Kunden wesentlich weitreichendere Möglichkeiten zu. Allerdings wird zugestanden, dass wegen der für Wartungs- und Pflegeaktivitäten erforderlichen "Downtimephasen" sowie möglicher Instabilitäten des Datennetzes keine hundertprozentige Verfügbarkeit gewährleistet werden kann.

Tabelle 3. Beschreibung der Erfolgsfaktoren (Teil 4)

Einflussfaktoren	Beschreibung
Risikoverlagerung	Da Sicherheits- bzw. Verfügbarkeitsaspekte im Verantwortungsbereich des Anbieters liegen, kann der Kunde im Falle unvorhergesehener Ausfallzeiten oder im Falle von Datenklau den Anbieter der Vertragsgestaltung entsprechend in Regress nehmen und Schadensersatzforderungen beanspruchen. Somit wird das Risiko von Sicherheitslücken und Ausfallzeiten des Systems vom Kunden zum Anbieter verlagert.
Konzentration auf Kernkompetenzen	Mit der Verlagerung der Anforderungen an die Bereitstellung einer anforderungsgerechten Software und an eine sichere und verlässliche IT-Betriebsführung auf einen Dienstleister, mit der Reduktion der Implementierungsaufwände und mit der Flexibilität bei organisationsbedingten Anpassungs- und Erweiterungsprozessen werden beim Kunden Ressourcen freigesetzt, die bislang für unterstützende Sekundärprozesse gebunden waren und nun für die eigentlichen Kernprozesse aufgewendet werden können.

Um im Weiteren zum einen die Forschungsfrage, und zum anderen die Fragestellungen zur Messung von „IS-Success“ vollständig zu beantworten, werden die identifizierten Erfolgsfaktoren gemäß ihres Einflusses nach kategorisiert und den in der Fachdiskussion als valide identifizierten Determinanten von „IS-Success“ zugeordnet. Es gilt dabei zu hinterfragen, ob sich die Vorteile für den Kunden auf die gesamte Gestaltung des Informationssystems, dessen Nutzungsrahmen oder die Konsequenzen der Nutzung (Net Benefits) beziehen [10].

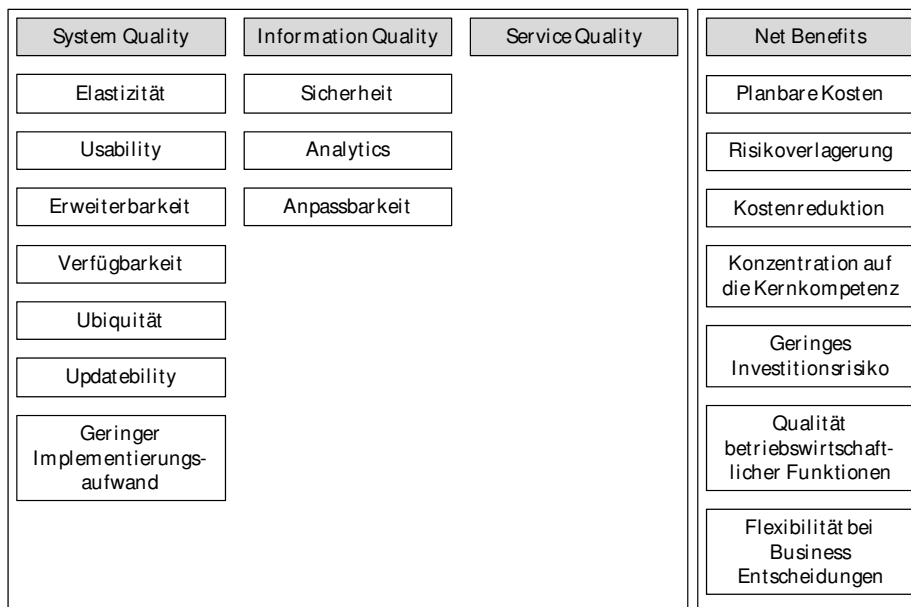


Abb. 2. Untergliederung der identifizierten Erfolgsfaktoren nach DeLone und McLean

An der Zuordnung aus Abbildung 2 wird deutlich, dass sich die Erfolgsfaktoren in den direkten Einfluss auf die Unternehmung (Net Benefits) und die auf Grund der Qualitätsaspekte zu erwartenden Vorteilen untergliedern lassen (System Quality, Informations Quality).

Der Aspekt der steigenden Service Qualität wurde in den Interviews von keinem Probanden erwähnt. Helpdeskfunktionen und Service Qualitäten sind bereits bei On-Premise-Systemen üblich und deshalb nicht als On-Demand-Enterprise typischer und relevanter Erfolgsfaktor zu qualifizieren. Erfolgsfaktoren bzgl. des Nutzungsrahmen bzw. der Kundenzufriedenheit eines On-Demand-Enterprise-Angebots fanden ebenfalls keine Beachtung, da der eigentlicher Verwendungszweck von On-Demand-Enterprise-Lösungen identisch ist mit der Zielsetzung des Einsatzes von On-Premise-Systemen: *„Wir haben ja eine komplette On-Premise-Lösung nachgebaut, aber die ist natürlich letztlich schlanker als das, was normalerweise im Angebot ist, und legt viel mehr Wert auf die Flexibilität und Upgradefähigkeit. (...) Ich würde sagen: Same same, but different.“* (Quelle, z. B. Interview 6)

Betrachtet man die der „**System Quality**“ zugeordneten Erfolgsfaktoren, generieren diese in erster Linie eine höhere Flexibilität und Einfachheit. Ohne eigenen Aufwand und ohne spezifisches Know-how kann der Kunde sein System einfach implementieren, es schnell und ohne großen Aufwand erlernen und davon ausgehen, dass es automatisch aktualisiert wird, ohne sich selbst um diese Updateprozesse kümmern zu müssen. *„Bei On Demand Lösungen sollte es sich um ein Low Touch Modell handeln, das darauf abzielt, einfach, adaptierbar und änderbar zu sein.“* (Quelle, z. B. Interview 8) Die Gefahr der Überlastung von Systemressourcen wird vom Anbieter gebannt und des Weiteren muss der Kunde die Verfügbarkeit des Systems nicht weiter verantworten. Er kann flexibel und ubiquitär auf das System zugreifen und schnell und einfach Änderungen der betriebswirtschaftlichen Prozesse mit speziellen Lösungen abdecken.

Ein ähnliches Schema zeigt sich bei der „**Information Quality**“. Berichte lassen sich schnell und einfach den Kundenwünschen anpassen. Zudem können die bislang nur komplexen On-Premise-Lösungen vorbehaltenen und aufwendig zu betreibenden, systemintegrierten Business-Warehouse-Analysefunktionen als fester Bestandteil kostengünstig in das Dienstleistungsangebot implementiert werden, so dass konkrete Informationen aus komplexen Datenmengen (bspw. Liquidationsforecasts) einfach und bedarfsgerecht bezogen werden können. *„Dadurch, dass wir die traditionelle Trennung zwischen OLTP und OLAP-System, also zwischen der transaktionalen Seite und der analytischen Seite aufgehoben haben, hat der Kunde bei einer On Demand Lösung immer die relevanten Entscheidungsunterstützungswerkzeuge zur Verfügung. (...) Natürlich versuchen wir das auch in der On-Premise-Welt hinzubekommen, aber so wie bei On-Demand-Systemen, weder von der Qualität noch von dem Preis, werden wir es nicht hinbekommen.“* (Quelle, z. B. Interview 11)

Mit diesen Qualitäten der innovativen Systemgestaltung zeigen sich für den Kunden verschiedene Vorteile resp. „**Net Benefits**“ auf:

- Kalkulierbare System-/Dienstleistungskosten und Vermeidung unvorhersehbare Aufwendungen.

- Einsparung von Ressourcen für den Systembetrieb bei Konzentration auf das eigentliche Kerngeschäft.
- Verlagerung des Betriebsrisikos auf den Dienstleistungsanbieter.
- Vermeidung von Investitionsrisiken.
- Die einfache und schnelle Bereitstellung eines kompletten On-Demand-Enterprise-Systems.
- Hohe Flexibilität bei eigenen Businessentscheidungen wegen geringerer Abhängigkeiten von IT-Implikationen.

Die Betrachtung der einzelnen Erfolgsfaktoren verdeutlicht, dass die meisten Aspekte in dem Wirkungsfeld von höherer Flexibilität und Einfachheit stehen und den Kunden resp. den Unternehmungen durch den Einsatz der innovativen IT-Services neue strategische Optionen liefern. *„Du brauchst keine Mannschaft mehr, die selber IT im Detail versteht, sondern du brauchst eigentlich hauptsächlich eine Mannschaft, die hauptsächlich betriebswirtschaftliche Prozesse versteht. Du kannst dich im Prinzip auf dein Core Business konzentrieren und musst dich nicht mehr mit Technologie groß rumärgern.“* (Quelle, z. B. Interview 7)

Deutlich wird auch, dass die dargestellten Erfolgsfaktoren systematisch in unterschiedlichen Konstellationen aufeinander einwirken und der Erfolg eines On-Demand-Enterprise-Systems von verschiedenen Wirkungsmechanismen abhängt. Dementsprechend ist, wie von DeLone und McLean für konventionelle IT-Systeme beschrieben, die Generierung von „IS-Success“ bei On-Demand-Enterprise-Systemen ebenfalls als eine Art Prozess resp. Verfahren zu verstehen, indem jeder der 17 identifizierten Erfolgsfaktoren zu berücksichtigen ist.

Ein On-Demand-Enterprise-Anbieter sollte sich somit für eine erfolgreiche Marktabtastung seines Angebots vergegenwärtigen: *„Schlussendlich ist es ja, was der Kunde damit machen muss, nach wie vor das Gleiche. Er muss seine Geschäftsprozesse abdecken können. Und wir als Anbieter müssen halt schauen, wie wir es für die Kunden noch einfacher, noch flexibler, noch innovativer (...) gestalten.“* (Quelle, z. B. Interview 2)

5 Zusammenfassung und kritische Würdigung

Ziel der vorgestellten Untersuchung war es, die Erfolgsfaktoren von On-Demand-Enterprise-Lösungen aus der Sicht eines On-Demand-Enterprise-Anbieters zu identifizieren. Es wurde versucht, die aufgezeigte Wissenslücke durch eine qualitative Studie zu schließen, in der 12 explorative Interviews mit Experten eines weltweit führenden deutschen Softwareherstellers durchgeführt wurden. Durchführung und Auswertung der Experteninterviews erfolgten unter Verwendung wissenschaftlich anerkannter Methoden der qualitativen explorativen Untersuchung und Inhaltsanalyse.

Aus dem speziellen Betrachtungswinkel eines Anbieters gibt die Arbeit einen Überblick über die Erfolgstreiber von On-Demand-Enterprise-Lösungen, der in dieser Form in der Fachliteratur noch nicht zu finden ist.

Die 17 wissenschaftlich valide identifizierten Erfolgsfaktoren können durchaus als Kriterien zur Bewertung von On-Demand-Enterprise-Systemen dienen.

Durch die Zuordnung der identifizierten Erfolgsfaktoren zum IS-Success-Modell von DeLone und McLean konnte aufgezeigt werden, dass die Nutzungsbegründung von On-Demand-Enterprise-Systemen zum einen auf die Reduktion von unternehmenbezogener IT-Kosten ausgerichtet ist. Diese wird zum anderen ergänzt durch Erfolgsfaktoren, die auf höhere Flexibilität des IT-Einsatzes zur Unterstützung betriebswirtschaftlichen Prozesse sowie auf die Einfachheit der IT-Nutzung zielen und den Unternehmen neue strategische Möglichkeiten eröffnen.

Es wird darüber hinaus ersichtlich, dass der Erfolg eines On-Demand-Enterprise-Systems in erster Linie von der Ausgestaltung der Systemqualitäten und den Netto-Nutzen des Informationssystems abhängt.

Anhand der identifizierten Erfolgsfaktoren zeigt sich der Problematik der Überbeanspruchung und missbräuchlichen Verwendung des Cloudbegriffs, dem sogenannten „Cloud Washing“ [22]. Gewisse Faktoren können nicht als Spezifikum der On-Demand-Enterprise-Systeme tituliert werden, da diese Nutzenversprechen ebenfalls mit konventionellen On-Premise-Systemen zu realisieren sind. So wird beispielsweise die intuitive Nutzeroberfläche oft als Merkmal von Cloud-Lösungen bezeichnet, obwohl diese auch bei On-Premise Lösungen realisiert werden könnte.

Abschließend bleibt darauf hinzuweisen, dass im Rahmen dieser Untersuchung lediglich die Sichten eines On-Demand-Enterprise-Anbieters untersucht wurden. Interessant und wichtig für die Identifikation und Bewertung der Erfolgsfaktoren wäre es, nun eine methodisch gleiche Untersuchung von Sichten verschiedener Kundengruppen durchzuführen. Darüber hinaus bleibt anzumerken, dass sich im Rahmen dieser qualitativen Studie keine repräsentativen Aussagen über die Bedingungen von Kausalmechanismen treffen lassen. Diese Arbeit konnte nur Ansatzpunkte für die Definition von Erfolgsfaktoren liefern, auf denen dann eine quantitative Studie aufsetzen kann, die die Bedeutung der dargestellten Faktoren unter repräsentativen Aspekten zu bewerten hat.

Literatur

1. Beuchat, A., Ziehler, P.: Cloud Computing ermöglicht Paradigmenwechsel (2011)
2. Gartner: Gartner Says Worldwide Software as a Service Revenue ist Forecast to Grow 21 Percent in 2011, <http://www.gartner.com/it/page.jsp?id=1739214>
3. Roberts, R., Sikes, J.: McKinsey Global Survey results: How IT is managing new demands (2010)
4. Gable, G., Sedera, D., Chan, T.: Re-conceptualizing information system success: the IS-Impact Measurement Model. *Journal of the Association for Information Systems* 9, 377-408 (2008)
5. Walther, S., Plank, A., Eymann, T., Singh, N., Phadke, G.: Success Factors and Value Propositions of Software as a Service Providers – A Literature Review and Classification (2012)
6. Saugatuck Technology: Cloud-basierte IT: Ja oder Nein? Herausforderung im Mittelstand (2010)
7. Vargo, S., Lusch, R.: Evolving to a New Dominant Logic for Marketing. *The Journal of Marketing* 68, 1-17 (2004)

8. Urbach, N., Smolnik, S., Riempp, G.: The State of Research on Information Success – A Review of Existing Multidimensional Approaches. *Business and Information Systems Engineering* 4, 315-325 (2009)
9. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems* 19, 9-30 (2003)
10. DeLone, W.H., McLean, E.R.: Information systems success: The quest for the dependent variable. *Information Systems Research* 3 (1), pp. 60-95 (1992)
11. Forrester Research, Inc.: *Decoding Cloud Computing: How Enterprises Should Segment Cloud Computing Opportunities* (2010)
12. Vehlow, M., Golkowsky, C.: *Cloud Computing – Navigation in der Wolke.* (eds.): Price-waterhouseCoopers AG Wirtschaftsprüfungsgesellschaft (2010)
13. Mell, P., Grace, T.: *The NIST Definition of Cloud Computing.* (eds.): National Institute of Standards and Technology (2009)
14. IT-Wissen: On-Demand, <http://www.itwissen.info/definition/lexikon/on-demand-OD.html>
15. SaaS-Magazin.de: Was ist Software as a Service (SaaS)?, <http://www.saasmagazin.de/lexikon-und-glossar/cisco-webex060809.html>
16. Brentani, U.: Success Factor in Developing New Business Services. *European Journal of Marketing* 2, 33-59 (1991)
17. Gläser, J., Laudel, G.: *Experteninterviews und qualitative Inhaltsanalyse.* VS, Wiesbaden (2009)
18. Faisst, W.: Die nächste Generation der Unternehmens-Software am Beispiel von SAP Business ByDesign. *Wirtschaftsinformatik & Management* 4, pp. 24-31 (2011)
19. Atteslander, P.: *Methoden der empirischen Sozialforschung.* Erich Schmidt, Berlin (2008)
20. Mayring, P.: *Qualitative Inhaltsanalyse: Grundlagen und Techniken.* Beltz, Weinheim und Basel (2008)
21. Lamnek, S.: *Qualitative Sozialforschung Band 2: Methoden und Techniken.* Psychologie Verlags Union, Weinheim (1995)
22. Mines, S., Kisker, H., Matzke, P.: *The Evolution of Cloud Computing Markets* (2010)

Cloud Computing lohnt sich (noch) nicht

Carsten Ingo Berendes¹, Markus Ertel¹, Thomas Röder¹, Thomas Sachs¹,
Thomas Süptitz², und Torsten Eymann²

¹ Universität Bayreuth, Bayreuth, Germany

{s3cabere, s3maerte, s3throed, s3tssach}@stmail.uni-bayreuth.de

² Universität Bayreuth, Lehrstuhl für Wirtschaftsinformatik (BWL VII), Bayreuth, Germany

{thomas.sueptitz, torsten.eymann}@uni-bayreuth.de

Abstract. Das Thema Cloud Computing steht nicht nur zunehmend im Interesse der medialen Berichterstattung, sondern auch vermehrt der wissenschaftlichen Forschung. Letztere widmet sich u. a. der Frage, ob Cloud Computing den Nachfragern die oft versprochenen Kostenvorteile bietet. Mit Hilfe einer agentenbasierten Simulation wurden verschiedene Nachfrageszenarien simuliert, um beurteilen zu können, ob die Nutzung von Cloud-Computing-Lösungen kostengünstiger als die Eigenfertigung, der Betrieb eigener Rechenzentren, ist. Die Ergebnisse zeigen, dass die Nutzung von Cloud-Computing-Lösungen etablierten Unternehmen (noch) keine Kostenvorteile bringt. Die Hauptursache dafür liegt in den relativ hohen Kosten für die Speicherung von Daten. Die Nachfrage nach zusätzlichen Sicherungen verstärkt diesen Effekt. Bei jungen, dynamisch wachsenden Unternehmen (z. B. Start-ups) konnte festgestellt werden, dass in den ersten zwei Jahren Cloud-Computing-Lösungen der Eigenfertigung vorzuziehen sind.

Keywords: cloud computing, make-or-buy, agent-based modeling, simulation, Amazon Web Services

1 Einführung

Auf dem deutschen Cloud-Computing-Markt werden im Jahr 2012 voraussichtlich 5,3 Milliarden Euro umgesetzt. Der Markt wird Schätzungen zufolge weiter mit zweistelligen jährlichen Raten auf 17,1 Milliarden Euro im Jahr 2016 wachsen [1]. Diese Zahlen belegen die steigende Bedeutung des Cloud Computing für Anbieter und Nachfrager. Unternehmen stehen zunehmend vor der Frage, ob IT-Leistungen wie die Speicherung und Verarbeitung von Daten selbst gefertigt oder fremdbezogen werden sollten. Bei dieser Entscheidung gilt es, sowohl quantitative (harte) als auch qualitative (weiche) Faktoren zu berücksichtigen [2-3]. Trotz der aufgezeigten Relevanz nehmen sich nur wenige wissenschaftliche Untersuchungen dieses Themas – aus Kostenperspektive – an.

2 Literaturüberblick

Verschiedene Vorgehensweisen sind denkbar, um die Wirtschaftlichkeit von Computing-Lösungen abzuschätzen. Diese reichen von einer lastspitzenabhängigen Heuristik [4], über eine Break-even-Analyse [5] bis hin zu einem komplexen Entscheidungsmodell, das quantitative und qualitative Faktoren berücksichtigt [6].

Mit der Anwendung der Methoden der Kostenrechnung auf IT-Infrastrukturen beschäftigen sich Thorsteinsson [7] und Brandl [8]. Ersterer identifiziert Kostenfaktoren im Detail und kategorisiert diese, um darauf aufbauend Anhaltspunkte für adäquate Kostenschätzungen abzuleiten. Brandl schlägt ein speziell auf verteilte IT-Infrastrukturen ausgerichtetes Modell vor, das gerade auch im Cloud Computing zur Anwendung kommen kann: Mit seiner Methode lässt sich der zu erwartende Ressourcenbedarf von an verteilte Systeme gerichteten Anfragen ermitteln.

Henneberger et al. nutzen eine multikriterielle Entscheidungsunterstützungsmethodik, um den Einsatz von SaaS- und IaaS-Services für das gesamte Unternehmen oder für dessen einzelne Ebenen zu eruieren. Diese prozessorientierte Analyse zeigt, dass SaaS-Services als Sourcing-Option in Frage kommen, wohingegen potenzielle (Sicherheits-)Risiken gegen den Einsatz von Cloud Sourcing sprechen [9].

Seltener werden Simulationen basierend auf realen Daten zur Evaluierung eingesetzt. Die Studie von Deelman et al. untersucht Kostensenkungspotentiale, die aus der Nutzung von Cloud-Lösungen resultieren [10]. Allerdings können diese Ergebnisse nicht auf gesamte Unternehmen oder verschiedene Branchen extrapoliert werden. Dieser Umstand begründet die zentrale Motivation der vorliegenden Untersuchung.

Hinsichtlich zu erwartender Ergebnisse gibt es Hinweise, dass sich Cloud-Lösungen eher für kleinere und mittlere Unternehmen lohnen, jedoch nicht für große [11]. Obwohl Entscheider vorwiegend Kosten als Beurteilungskriterium nutzen [12], sollte dieses nicht allein den Ausschlag geben [2-3]. Dennoch werden Kosten stets grundlegender Bestandteil einer Wirtschaftlichkeitsanalyse sein.

Auf Basis einer agentenbasierten Simulation wird in der vorliegenden Arbeit ein quantitativer Vergleich zwischen den zwei Alternativen Eigenfertigung und Fremdbezug angestrebt, um theoretische Kosten-/Nutzenüberlegungen zu unterstützen. In diesem Zusammenhang wird für den Fremdbezug exemplarisch die Preisstruktur von Amazon Web Services (AWS) als einem der führenden Cloud-Computing-Anbieter herangezogen, dessen umfassendes Angebot von der Technologieebene ausgeht [13]. Ziel der Untersuchung ist, anhand von systematischen Analysen Handlungsempfehlungen für verschiedene Branchen abzuleiten.

3 Darstellung der Nachfrageszenarien

Das Modell betrachtet für die Analyse vier Nachfrageszenarien, die sich hinsichtlich der Ausgestaltung der Anforderungsparameter, die in Abschnitt 4.2 näher erläutert werden, unterscheiden.

Einem beispielhaften *mittelständischen Unternehmen* wird nachfolgend unterstellt, dass es zu Beginn einen Speicherplatz von 4,8 Terabyte (TB) benötigt [14] und die

Daten nur einmal zusätzlich sichert. Die angesetzte Rechenleistung von $19,1 \text{ TFLOPS} \times h$ orientiert sich an einem typischen Serverangebot für mittelständische Unternehmen, bei einer durchschnittlichen Inanspruchnahme von 40 % in einem Monat mit 720 Stunden [15]. Als Breitbandanschluss setzt das Modell eine Anbindung von 50 Megabit pro Sekunde (Mbit/s) voraus, die einen Datentransfer von 5,5 TB pro Monat zu bewältigen hat. Wachstumsannahmen stützen sich auf die Vorhersage des Deutschen Industrie- und Handelskammertages, der für das Jahr 2012 ein Wachstum von 1,3 % prognostizierte [16].

Das Szenario *Handelskette* bildet ein Unternehmen nach, das einen schnell wachsenden Bedarf an Speicherplatz aufweist: pro Tag fallen bis zu 14 TB an [17]. In diesem Szenario kommt dem Parameter Speicherplatzbedarf große Bedeutung zu, zum einen durch den hohen täglichen Zuwachs an Daten, zum anderen durch eine dreifache Sicherung. Die Rechenleistung weist keine außergewöhnlichen Spitzen auf, sondern ist mit einer konstanten monatlichen Summe von $52,5 \text{ TFLOPS} \times h$ angesetzt. Die angenommene Wachstumsrate von 1,3 % stützt sich auf die Prognose des Handelsverbands HDE und der Gesellschaft für Konsumforschung für das Jahr 2012 [18]. Es steht dem Unternehmen eine Backbone-Anbindung mit einem Durchsatz von 1.200 Mbit/s zur Verfügung.

Die Daten im Szenario *Wetterdienst* basieren auf Informationen des Deutschen Wetterdienstes (DWD). Der DWD besitzt ein Datenarchiv von 4 Petabyte (PB), welches nicht in die Cloud migriert werden soll, sondern nur die täglich neu anfallenden Daten in Höhe von 7 Gigabyte (GB) [19]. Der anfängliche Speicherbedarf beträgt 0,2 TB. Im Modell kommen zwei Sicherungsinstanzen zur Anwendung. Die monatlich benötigte Rechenleistung beträgt $15.700 \text{ TFLOPS} \times h$. Diese Angabe stützt sich auf die mögliche Spitzenrechenleistung des DWD von 109 TFLOPS [20], von der eine durchschnittliche Auslastung von 20 % angenommen wird. Die Rechenleistung wird u. a. für tägliche Wettervorhersagen sowie für spezielle saisonale Berechnungen wie Gewitterrisiken eingesetzt. Der DWD benötigt zwar eine hohe Rechenleistung, zielt dabei aber insbesondere auf die Ergebnisse der Wetter- und Klimasimulationen ab. Dies begründet einen eher geringen Datentransfer von 0,3 TB pro Monat.

Das Szenario *Start-up* bildet Unternehmen ab, die durch ein hohes Wachstum aller Parameter – beispielhaft 10 % pro Jahr – gekennzeichnet sind [21]. Das höhere Wachstum im Vergleich zu den anderen Szenarien verdeutlicht sich in dem schneller wachsenden Bedarf an Speicherplatz, Rechenleistung und Datentransfer. Der anfängliche Speicherplatzbedarf beträgt 0,01 TB, wobei eine Sicherungsinstanz zur Anwendung kommt. Die Rechenleistung zeigt zwar durch das Wachstum eine gewisse Dynamik, weist jedoch keine hohe Grundlast auf; annahmegemäß beträgt sie monatlich $1,9 \text{ TFLOPS} \times h$. Der zu erwartende Datentransfer beträgt zu Beginn 0,3 TB pro Monat. Als Anbindung kommt ein Breitbandanschluss mit 32 Mbit/s in Frage.

Tabelle 1. Zusammenfassende Darstellung der Nachfrageszenarien

	Einheit	Mittelstand	Handelskette	Wetterdienst	Start-up
Speicherbedarf	TB	4,8	400	0,2	0,01
Anzahl der Sicherungen	-	1	3	2	1
Rechenleistung	TFLOPS × h	19,1	52,5	15.700	1,9
Spitzenlast Rechenleistung	GFLOPS	60	120	109.000	12
Datentransfer	TB/Monat	5,5	370	0,3	0,3
Bandbreite	Mbit/s	50	1.200	1.000	32
Wachstum	%/Monat	0,1	0,1	0,02	0,8

4 Forschungsmethodik

4.1 Agentenbasierte Simulation

Unternehmen, die einen Bedarf an Leistungen elektronischer Datenverarbeitung oder -speicherung haben (Nachfrager), besitzen mindestens zwei Entscheidungsalternativen: jene Leistungen (a) in eigenen Rechenzentren zu erstellen (= Eigenfertigung / „Make“) oder (b) als Dienstleistung von externen Anbietern zu beziehen, zu denen die Daten über Internetverbindungen gelangen (= Fremdbezug / „Buy“). In der agentenbasierten Simulation werden beide Marktteilnehmer als Agenten dargestellt. Das Kosten-/Nutzenkalkül der – anhand verschiedener Anforderungsparameter – typisierten Nachfrager führt zur Entscheidung zwischen den Extremen (a) und (b). Unternehmen können derart gewonnene Simulationsdaten als Grundlage für ihre Entscheidung zwischen Eigenfertigung und Fremdbezug von Computing-Leistungen dienen.

Agentenbasierte Simulation ist eine dezentral aus der Sicht eines Individuums oder mehrerer Individuen geprägte Nachbildung komplexer Systeme. Jedes handelnde Individuum (Person, Wettbewerber, Unternehmen etc.) wird dabei zum Agenten. Diese Agenten, auch unterschiedlichen Typs, interagieren gemäß ihrer zuvor definierten Verhaltens- und Entscheidungsmöglichkeiten in der Simulation miteinander. Computergestützte Simulationen haben den erheblichen Vorteil, dass sie nicht real vollzogen werden müssen; damit reduzieren sich sowohl Entwicklungs- und Forschungskosten als auch der Zeitaufwand für die Modellbildung, oder es lassen sich Modelle überhaupt erst entwickeln. Die Autoren führen Modellbildung und Simulation mithilfe der Software Anylogic (Version 6.8.1) durch. Eine zeitliche Einheit (Periode) der Simulation entspricht jeweils einem Monat.¹

¹ Im Modell umfasst ein Monat 30 Tage, an denen IT-Leistungen beansprucht werden. Bei vielen Unternehmen bestehen nur vernachlässigbare Unterschiede zwischen Werktagen und Sonn-/Feiertagen, bei anderen gleichen sich Differenzen im Wochenverlauf aus.

4.2 Parameter auf Nachfrageseite

Zur Nachfrageseite des modellierten Marktes können beliebig viele Agenten gleichzeitig hinzugefügt werden, gekennzeichnet durch sieben unterschiedliche Attribute. Diese Simulationsparameter entsprechen den wesentlichen Anforderungen, die Nachfrager an Computing-Leistungen stellen.

Der Parameter *Start_Speicherplatz* wird in der Einheit TB bemessen und gibt das Speichervolumen an, in dessen Umfang pro Monat neue Daten hinzugefügt werden. Einmal (in Vormonaten) gespeicherte Datenvolumina bleiben erhalten und daher als Grundstock bei der Kostenberechnung stets berücksichtigt, sodass es sich bei dem Parameter um den Monat für Monat tatsächlichen zusätzlichen Speicherbedarf handelt. Der Speicherbedarf wächst außerdem durch die erforderliche Anzahl von Datenspiegelungen (zusätzliche Sicherungskopien der gespeicherten Daten), angegeben durch den Parameter *Anzahl_Sicherungen*.

Eine zweite zentrale Anforderung stellt die Datenverarbeitung dar, deren Umfang der Parameter *Start_Rechenleistung* bemisst. Maß für die Verarbeitungsleistung ist dabei die Anzahl der vom Prozessor ausgeführten Gleitkommaoperationen pro Sekunde – engl. Floating Point Operations Per Second, FLOPS – in Billionen (10^{12} FLOPS = TFLOPS). Durch den zweiten Faktor der Einheit, Stunden (h), wird berücksichtigt, für welche Dauer die angegebene Rechenleistung in Anspruch genommen wird. Ähnlich der Kilowattstunde (kWh) als Energieeinheit ergibt sich als Einheit der Rechenleistung das Maß TFLOPS \times h. Bei Verteilung der Rechenleistung über den Monat ergäbe sich rechnerisch eine durchschnittliche Serverlast, die es zu unterscheiden gilt in Grundlast und temporäre Spitzenlast. Insbesondere letztere, angegeben durch *Spitzenlast_Rechenleistung* (in GFLOPS), ist relevant für die Berechnung notwendiger Kapazitäten: Da der Servercluster jene Leistung im Bedarfsfall vollumfänglich gewährleisten muss, wird der Bedarf an Servern an der Spitzenlast ausgerichtet.

Sowohl die Übermittlung von Daten über Internetverbindungen in das fremde Rechenzentrum (Traffic eingehend) als auch der Abruf der dortigen Daten (Traffic ausgehend) finden Niederschlag in der Größe *Start_Datentransfer*, gemessen in TB pro Monat. Damit die Nutzung von Cloud Computing überhaupt möglich wird, ist zu beachten, dass der monatliche Datentransfer – zumindest im Mittel – die am Standort des Nachfragers verfügbare Bandbreite nicht überschreiten darf. Eine Routine prüft in der Simulation dieses K.-o.-Kriterium und warnt, falls mangelnde Bandbreite den Fremdbezug unmöglich macht. Durch den gleichnamigen Parameter wird die nutzerspezifische Angabe der am Standort jeweils verfügbaren *Bandbreite*, gemessen in Mbit/s, abgebildet: zum Beispiel Werte im ein- und zweistelligen Bereich bis 32 Mbit/s für DSL-, 50 Mbit/s für VDSL- sowie Werte bis zu 1.000 Mbit/s für unmittelbare (Gigabit-) Backbone-Anbindungen.

Für die drei Parameter beginnend mit „Start_“ gilt, dass deren Werte jeweils den Markteintritt des Nachfragers, d. h. den Monat nach Hinzufügen des Agenten, betreffen; in den Folgemonaten wachsen diese drei Startwerte jeweils um den angegebenen Faktor *Wachstum*, das bedeutet:

$$\text{Wert}_{(t+1)} = \text{Wert}_{(t)} \times (1 + \text{Wachstum}). \quad (1)$$

4.3 Parameter auf Angebotsseite

Die Anbieterseite umfasst aus Vereinfachungsgründen einen einzigen Agent, der in der Lage ist, Anforderungen aller Nachfrager vollumfänglich zu befriedigen. Als Referenz wird hierfür AWS als großer Anbieter auf dem Markt für Cloud Computing herangezogen, ähnlich Matros et al. [5].

Drei Parameter und Formeln bilden die Preisstruktur von AWS möglichst gut nach. Da die Anbieterpreise von AWS in US-Dollar (USD) notiert sind, orientiert sich das gesamte Modell an dieser Währung, bezieht sich jedoch auf den deutschen bzw. europäischen Markt. Pro genutztem GB Speicherplatz berechnet AWS bei dem Produkt Simple Storage Service (Amazon S3) 0,11 USD [22], dies entspricht 110 USD pro TB,² dem Standardwert des Parameters *Preis_Speicherplatz*. Ab gewissen Speichervolumina gelten ermäßigte Tarife. Die Preisstaffelung ist für das Modell mittels einer Funktion nachgebildet worden: Indem auf den durchschnittlichen Preis pro TB abgestellt wurde, wurde eine hohe Güte der Anpassung ($R^2 \approx 0,97$) erreicht. Mit S in TB als Speichervolumen ergibt sich der Durchschnittspreis pro TB wie folgt.

$$P_S^\emptyset = f(S) = \begin{cases} 110 \frac{\text{USD}}{\text{TB}} & \text{für } S \leq 50 \text{ [TB]} \\ (-0,0545 \times \ln(S) + 1,206) \times 110 \frac{\text{USD}}{\text{TB}} & \text{für } S > 50 \text{ [TB]} \end{cases} \quad (2)$$

Eine Preisstaffelung tritt ebenfalls in Kraft bei dem Preis des Datentransfers. Im Parameter *Preis_Datentransfer* ist der Grundpreis von 120 USD pro im Monat übertragenen TB festgehalten. Die Preisstaffelung ist mittels einer Funktion nachgebildet worden, wobei auch hier auf den durchschnittlichen Preis pro übertragenem TB abgestellt und eine beliebige Teilbarkeit unterstellt worden ist. Die Funktion weist eine sehr hohe Güte der Anpassung auf ($R^2 > 0,99$). Mit T als Transfervolumen in TB ergibt sich der Durchschnittspreis pro TB wie folgt:

$$P_T^\emptyset = f(T) = \begin{cases} 120 \frac{\text{USD}}{\text{TB}} & \text{für } T \leq 10 \text{ [TB]} \\ (-0,1276 \times \ln(T) + 1,2707) \times 120 \frac{\text{USD}}{\text{TB}} & \text{für } T > 10 \text{ [TB]} \end{cases} \quad (3)$$

Der Preis für die Rechenleistung ist bei dem AWS-Produkt Elastic Compute Cloud (Amazon EC2) fixiert: Der Preis pro Rechenstunde (*Preis_Rechenleistung*) liegt bei 0,085 USD [23]. Dem Modell zugrunde liegt die Standard-Instanz (Small), die nach Bedarf gebucht werden kann. AWS misst die entsprechende Rechenleistung in der eigenen Einheit EC2 Compute Unit (ECU), wobei die Standard-Instanz genau eine ECU bietet [24]. AWS setzt diese mit der Rechenkapazität eines 1,0–1,2 GHz Opteron- oder Xeon-Prozessors aus dem Jahr 2007 gleich. Ein Vergleichsmaßstab baut auf der Herleitung von Ostermann et al. auf: „a 1.1 GHz 2007 Opteron can perform 4 flops per cycle at full pipeline, which means at peak performance one ECU equals 4.4 gigaflops per second (GFLOPS)“ [25]. Beliebige Teilbarkeit unterstellt, entsprechen 0,085 USD pro ECU-Rechenstunde ($4,4 \text{ GFLOPS} \times h$) einem Preis von 19,3182 USD für $1 \text{ TFLOPS} \times h$. Hierbei tritt keine Degression in Kraft.

² AWS wendet den Umrechnungsfaktor 10^3 an, also $1 \text{ TB} = 1.000 \text{ GB}$.

Wie oben dargestellt, umfasst die Angebotsseite genau einen Agenten. Dies trägt auch der Tatsache Rechnung, dass Cloud Computing ohne Sonderleistungen als Commodity aufgefasst werden kann.

Um die Wettbewerbssituation auf dem Markt für Cloud-Computing-Leistungen ansatzweise ins Modell zu integrieren, sind dem Anbieter auch Parameter bezüglich seiner Wettbewerbsstrategie zugeordnet worden:

- *Sicherheitsaktivitäten*: Der Anbieter investiert in die Sicherung seiner Systeme gegenüber Datenverlust, unbefugten Zugriffen und Angriffen von Dritten.
- *Service*: Der Anbieter bietet seinen Kunden unterstützende Dienstleistungen wie eine Hotline, Autokonfiguration, Hilfsfunktionen, außergewöhnliche Usability etc.
- *Verfügbarkeit*: Der Anbieter ergreift Maßnahmen, um Ausfallsicherheit und eine hohe Verfügbarkeit zu gewährleisten, schafft beispielsweise Redundanz.
- *Hardwareleistung*: Der Anbieter legt bei der Beschaffung Wert auf hochwertige und leistungsfähige Hardware.

Diese Parameter messen jeweils die relative Position im Branchenvergleich: Anbieter können eine im Branchenvergleich führende und entsprechend teure Position anstreben (Werte tendieren zu 1) oder eine zurückhaltende, kostensparende Strategie verfolgen (Werte tendieren zu -1). Mit Werten um 0 wird jeweils eine durchschnittliche Position angegeben. Durch die Ausprägung der vier Parameter wird die Abbildung konträrer Wettbewerbsstrategien möglich.

4.4 Fremdbezug: Zusammenfügen von Nachfrage und Angebot

Auf der Nachfrageseite existieren nach dem Hinzufügen beliebig viele Agenten mit gleichen oder verschiedenen Attributen. Diese Zahl der Nachfrager besitzt einen Bedarf an IT-Leistungen, den die individuell verschiedenen Attribute definieren. Um zu ermitteln, zu welchem Preis der Anbieter diese befriedigen kann, findet ein Austausch der dem Bedarf entsprechenden Preisinformationen statt.

Die Nachfrage bezüglich Speicherplatz, Rechenleistung und Datentransfer wird zugleich aggregiert und vom Anbieter gespeichert. Er ist somit in der Lage, seinen Hardware-Bestand dem durch die Nachfrage bestimmten Bedarf anzupassen. Der Anbieter hat die Möglichkeit, seine Grundpreise zu ändern. Da ceteris paribus die Kostenbestandteile für Ersteller von IT-Leistungen identisch bleiben, verändert sich dadurch lediglich die Gewinnspanne des Anbieters. Die oben beschriebenen Wettbewerbsstrategie-Parameter sind so implementiert, dass ein von ihren Werten abhängiger Teil der Gewinnspanne als (monatlicher) Deckungsbeitrag verbleibt. Der Verlauf wird wie folgt charakterisiert:

- Sind alle vier Parameter gleich -1, so verbleibt die gesamte Gewinnspanne als Deckungsbeitrag.
- Sind die vier Parameter gleich 1, so verbleibt kein Deckungsbeitrag.
- Zwischen diesen Extremen verläuft die Funktion in Form einer S-förmigen Kurve.
- Am Mittelpunkt der Kurve („Branchendurchschnitt“), hier beträgt die Summe der Parameterwerte 0, verbleiben ca. 70 % der Gewinnspanne als Deckungsbeitrag, da

ein durchschnittliches Qualitätsniveau annahmegemäß mit verhältnismäßig geringeren Kosten als die führende Position zu erreichen ist. In anderen Worten: Branchentypische Anbieter bemühen sich, bei den Investitionen in Qualität hauszuhalten und Qualitätsziele gegen den erzielbaren Deckungsbeitrag abzuwägen.

- Um ein überdurchschnittliches Qualitätsniveau zu erreichen, sind dementsprechend hohe Investitionen notwendig, die den Deckungsbeitrag beträchtlich schmälern.

4.5 Kosten der Erstellung von Datenverarbeitungsleistungen

Für Agenten auf beiden Marktseiten gelten identische Kostenbestandteile hinsichtlich der Erstellung von Computing-Leistungen, d. h. der Bereitstellung von Speicherplatz, Rechenleistung und Datentransfervolumen. Diese Leistungen an sich (unter Vernachlässigung oben beschriebener Qualitätsmerkmale) können dabei als Commodity aufgefasst werden, sie sind also qualitativ identisch. Diese Annahme ermöglicht den einfachen Vergleich der Vorziehenswürdigkeit von Eigenerstellung oder Fremdbezug anhand der Kosten. Sowohl der Anbieter von Cloud-Computing-Dienstleistungen als auch der Nachfrager von IT-Leistungen, sofern er sich für die Eigenerstellung entscheidet, müssen Investitionen in die Serverinfrastruktur tätigen und die Kosten des laufenden Betriebs tragen.

Bei den *Infrastrukturinvestitionen* handelt es sich meist um fixe Kosten. Sprungfix sind allerdings die Investitionen in Festplatten und Server, da diese in Abhängigkeit der Ausbringungsmenge von Speicherplatz (S in TB) bzw. der Rechenleistung (C in TFLOPS $\times h$) anzuschaffen sind. In Annahme von 1-TB-Einheiten ist die Auswirkung bei Festplatten gering.

Andererseits ist der sprungfixe Charakter der Server-Investitionen stärker: In Abhängigkeit von der benötigten Rechenleistung werden im Modell Server des Typs HP ProLiant DL160 Gen8 angeschafft. Ausgestattet mit dem Prozessor Xeon E5-2603 besitzen diese eine Leistungsspitze von 57,6 GFLOPS. Ein angenommener durchschnittlicher Auslastungsgrad von 75 % ist dabei u. a. durch Virtualisierung erreichbar. Die Server werden in Verbindung mit Netzwerkhardware wie Switches, Firewalls, Load Balancer und Verkabelung in Racks untergebracht. Pro Rack mit 42 U-Einheiten können dabei annahmegemäß 36 Server montiert werden, der restliche Teil der U-Einheiten wird von der Netzwerkhardware belegt.

Folgende Summanden ergeben zusammen die Infrastrukturkosten pro Monat K_I :

1. Lineare Abschreibung der Anschaffungskosten von Servern (angenommene Nutzungsdauer drei Jahre [5]), erhöht um einen prozentualen Zuschlag für Netzwerkhardware (annahmegemäß 20 %)

$$AfA = \frac{\text{Anschaffungskosten}}{\text{Nutzungsdauer}} \times (1 + p_{\text{Netzwerkhardware}}) \quad (4)$$

2. Wiederkehrende Kosten aus Wartungsverträgen als Prozentsatz der Anschaffungskosten von Servern und Netzwerkhardware, annahmegemäß 1 % monatlich

3. Einmalige Kosten für Software zur Ausstattung angeschaffter Server, insbesondere mit einem Betriebssystem, annahmegemäß null dank dem Einsatz einer Open-Source-Distribution von Linux
4. Lineare Abschreibung der Anschaffungskosten von 42 U-Racks (angenommene Nutzungsdauer fünfzehn Jahre [5])
5. Flächenkosten als Produkt der (ggf. kalkulatorischen) Miete für die Standfläche der Racks (annahmegemäß der deutsche Durchschnittsmietpreis für Büroflächen), der Zahl im Bestand befindlicher Racks und dem Platzbedarf pro Rack inklusive Kühlung, Laufwegen etc.
6. Lineare Abschreibung der Anschaffungskosten für Festplatten in 1-TB-Einheiten, in Abhängigkeit von der benötigten Menge an Speicherplatz S
7. Abzug eines einmaligen Rabatts vom Preis angeschaffter Hardware, um Einsparungen zu berücksichtigen, die insbesondere bei großen Einkaufsvolumina möglich werden (Fixkostendegression):

$$\text{Rabattsatz} = -0,01 \ln (\text{Wert gekaufter Hardware}) \quad (5)$$

8. Kalkulatorische Kosten des in der Infrastruktur gebundenen Kapitals, als Produkt des gewichteten Kapitalkostensatzes und der fortgeführten Anschaffungskosten (nach Abschreibungen) aller im Bestand befindlicher Hardware

Bei den *Betriebskosten* handelt es sich hingegen um variable Kosten. Hierbei kommt auch die Abhängigkeit von der Datentransfermenge T zum Tragen. Folgende Summanden ergeben zusammen die Betriebskosten K_O :

1. Elektrizitätskosten unter Heranziehung des deutschen Strompreises für Gewerbekunden; der Strombedarf berechnet sich dabei als Produkt der Leistung eines Netzteils zur Versorgung eines Servers bei Spitzenlast (nach Herstellerangaben 500 W), des Umwandlungsfaktors (verwendeter Anteil der Stromversorgung bei typischer Belastung, berechneter Wert 40 % basierend auf [5]), des Multiplikators für den Gesamtbedarf eines Rechenclusters einschließlich Kühlung, Netzwerk etc. (bemessen am Strombedarf eines einzelnen Servers, annahmegemäß 200 % analog [26]) und der Betriebsstunden im Monat (durchgängiger Betrieb, d. h. 720 h):

$$\begin{aligned} \text{Strombedarf} &= \text{Netzteilleistung} \times \text{Umwandlungsfaktor} \\ &\times \text{Multiplikator Gesamtstrombedarf} \times \text{Betriebsstunden} \end{aligned} \quad (6)$$

2. Kosten des Datentransfers, d. h. Kosten für Bandbreite und die Nutzung der Datenübertragung ins Internet, multipliziert mit dem Datentransfer (T in TB/Monat, als Monatsmittel umgerechnet in Mbit/s) – annahmegemäß entfallen diese Kosten für Nachfrager, die eine eigene IT-Infrastruktur vor Ort aufbauen
3. Kosten der technischen Betreuung des Rechenzentrums, berechnet als Produkt des Monatsgehalts eines IT-Administrators ohne Personalverantwortung in Deutschland (Jahresgehalt inkl. variablen Bestandteilen: 44.400 Euro [27]) und der Zahl im Bestand befindlicher Server, die mit einer Vollzeitstelle betreut werden können

Alle genannten Kostenbestandteile fließen im Modell in eine Kalkulation ein, wofür die Nachfrage der Agenten der Nachfrageseite aggregiert wird. Auf Grundlage der Gesamtwerte berechnen sich die dem Anbieter entstehenden Kosten für die Bereitstellung der Cloud-Computing-Leistung. Auf Nachfrageseite ergeben sich die Kosten der Eigenerstellung für jeden Agenten individuell.

5 Darstellung der Simulationsergebnisse

5.1 Hinweise zur Interpretation der gewonnenen Daten

Die beschriebenen Szenarien wurden mit folgenden Grundeinstellungen simuliert:

- Anzahl der Anbieter und Nachfrager: jeweils einer
- Simulationszeitraum: 60 Monate (= fünf Jahre)

Die Kostendifferenz zwischen „Buy“ und „Make“ berechnet sich wie folgt:

$$\text{Kostendifferenz} = \text{Gesamtkosten Buy} - \text{Gesamtkosten Make} \quad (7)$$

Ein positiver Betrag bedeutet, dass eine Eigenfertigung gegenüber der Inanspruchnahme von Cloud-Computing-Lösungen (Fremdbezug) vorzuziehen ist und vice versa. Diese Differenz gilt nachfolgend als ausschlaggebendes Kriterium bei den entsprechenden Darstellungen, Würdigungen und Handlungsempfehlungen.

5.2 Szenario: Mittelständisches Unternehmen

Unter den gesetzten Parameterwerten kann in dem Szenario „Mittelständisches Unternehmen“ keine Empfehlung für den Bezug von Cloud-Computing-Lösungen ausgesprochen werden. Dies zeigt sich daran, dass im Zeitablauf die Kostendifferenz zwischen Buy und Make zunimmt; dies verdeutlicht Abbildung 1.

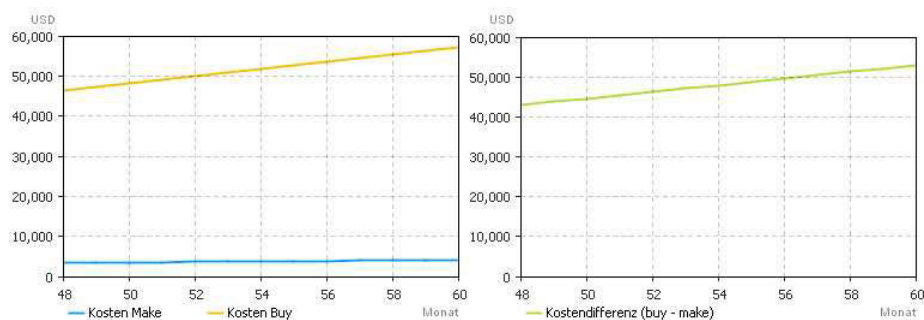


Abb. 1. Kosten beim Szenario „Mittelständisches Unternehmen“

Eine künftige Umkehr der Entscheidung ist auszuschließen, da die Kurven „Kosten Make“ und „Kosten Buy“ nicht zueinander streben. Hauptursache für diese Entwick-

lung ist der zunehmende Bedarf an Speicherplatz. Die Rechenleistung und der Datentransfer sind hingegen keine entscheidenden Kostentreiber.

5.3 Szenario: Handelskette

Bei den gewählten Parametern in dem Szenario „Handelskette“ kann ebenso keine Empfehlung zu Gunsten des Cloud Computing ausgesprochen werden. Auch hier nimmt die Kostendifferenz zwischen Make und Buy im Zeitablauf zu. Eine Umkehr der gegebenen Handlungsempfehlung in künftigen Monaten ist nicht zu erwarten, da ein Divergieren der Kostenkurven erkennbar ist.

Wie bei dem Szenario „Mittelständisches Unternehmen“ kann der Bedarf an Speicherplatz als Kostentreiber und somit ausschlaggebender Faktor für die Handlungsempfehlung identifiziert werden.

5.4 Szenario: Wetterdienst

Das Szenario „Wetterdienst“ hebt sich von den beiden zuvor genannten Szenarien dadurch ab, dass es einer erhöhten Rechenleistung bedarf, dafür aber weniger Speicherplatz und Datentransfer. Trotzdem ist auch hier die Eigenfertigung für das Unternehmen günstiger, wie Abbildung 2 zeigt.

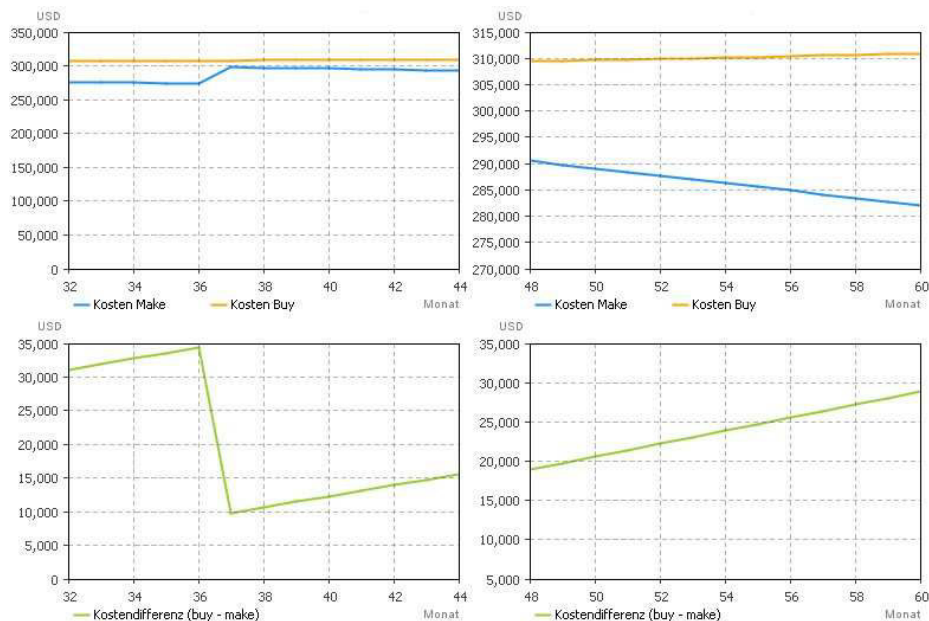


Abb. 2. Kosten beim Szenario „Wetterdienst“

Ursächlich für den Kostennachteil der Cloud-Lösung sind in diesem Szenario die hohen Ausgaben für Rechenleistung. Obwohl sich dies hierfür im Falle des Eigenbe-

etriebs anfallenden Investitionen sprunghaft verhalten, liegt die zugehörige Kostenkurve durchgängig unter der des Fremdbezugs. Im Zeitverlauf sinken aufgrund des linearen Abschreibungsverlaufs die Kosten gebundenen Kapitals, wodurch die Kurve der Eigenherstellung fallend verläuft. Ein Knick in der Kostenkurve entsteht immer dann, wenn Ersatzinvestitionen getätigt werden. Hingegen steigen die Kosten des Fremdbezugs kontinuierlich. Durch das langfristige Auseinanderstreben der beiden Kurven ist auch hier keine Umkehr der getroffenen Entscheidung zu erwarten.

5.5 Szenario: Start-up

Cloud Computing wird oft im Zusammenhang mit Start-ups genannt, die auf Grund ihres unvorhersehbaren Bedarfs und limitierter Kapitalausstattung auf solche Lösungen zurückgreifen. Die Simulation ergab, dass dies durchaus als sinnvoll betrachtet werden kann, denn bis zum 28. Monat ist der Fremdbezug von IT-Leistungen für das Start-up günstiger. Erst zu Beginn des 29. Monats kehrt sich der Vorteil um; dies verdeutlichen der Schnittpunkt der Kostendifferenzkurve mit der x-Achse, ab dem sie im positiven Bereich verläuft, bzw. der Schnittpunkt der Kostenkurven Make und Buy. Am Szenario Start-up zeigt sich, dass der schnell wachsende Bedarf an Speicherplatz auf lange Sicht die Entscheidung in Richtung Eigenfertigung beeinflusst.

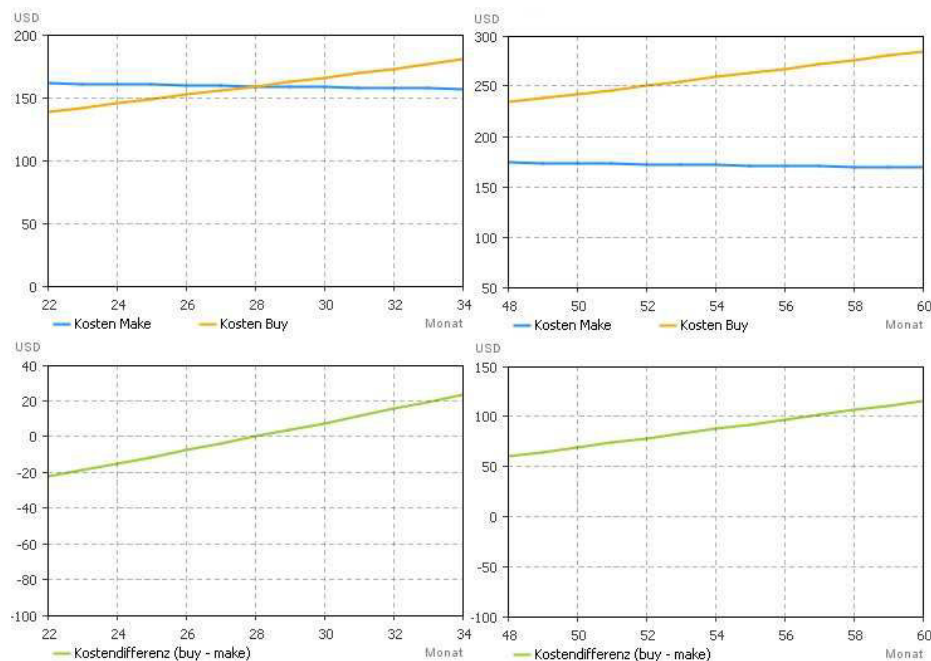


Abb. 3. Kosten beim Szenario „Start-up“

6 Würdigung der Ergebnisse

6.1 Einschätzung im Kontext bisheriger Forschung

Matros et al. [5] und Lamberth/Weisbecker [6] gehen davon aus, dass sich Cloud Computing nur bei rechenintensiven Unternehmen lohnt. Wie die Ergebnisse der Simulation zeigen, ist Cloud Computing noch keine lohnenswerte Alternative für etablierte Unternehmen mit hohem Bedarf an IT-Leistungen. Junge, dynamisch wachsende Unternehmen wie Start-ups können hingegen von Cloud-Lösungen profitieren. Als wesentlicher Kosten- und Entscheidungsfaktor wurde die Nachfrage nach und der Preis von Speicherplatz identifiziert.

Die Studie von Harms/Yamartino [28] prognostiziert, dass die Preise für Cloud-Lösungen künftig fallen werden. Sollte dies zutreffen, könnten auch größere Unternehmen mit einem differenzierteren Anforderungsprofil auf Cloud Computing setzen.

6.2 Modellrestriktionen

Das aufgestellte Modell bildet einen vereinfachten Ausschnitt der Wirklichkeit ab und wird u. a. durch folgende zu berücksichtigende Faktoren beschränkt:

- Das Modell geht davon aus, dass es im Markt nur einen Anbieter gibt, der jegliche Nachfrage befriedigen kann. Das Modell bildet die tatsächliche Wettbewerbssituation somit nur ansatzweise ab.
- Die im Zuge der Erstellung von Leistungen anzuschaffende IT-Ausstattung wird dem Erwerber bei Bedarf ohne Zeitverzögerung bereitgestellt. Wartezeiten für Lieferung, Installation und Konfiguration bleiben somit außer Betracht.
- Gemäß der Annahme des Modells können Nachfrager Angebote für Speichervolumen, Rechenleistung und Datentransfer in beliebiger Höhe wahrnehmen. Die Angebote gelten damit als teilbar. Pauschal bezahlte Inklusiv-Kontingente oder Grundgebühren bleiben außer Betracht.

7 Fazit

Nachfrager von IT-Leistungen, die vor der Wahl zwischen Eigenerstellung und Fremdbezug stehen, befinden sich in einer komplexen Entscheidungssituation, in der sie quantitative wie auch qualitative Faktoren zu berücksichtigen haben. Die Simulation konzentrierte sich auf den Kostenaspekt als elementares Entscheidungskriterium und nutzte verschiedene Szenarien, um einen ausgesuchten Branchenquerschnitt praxisnah nachzubilden.

Die Ergebnisse zeigen, dass Cloud-Lösungen bislang nur für kleine und dynamisch wachsende Unternehmen in der Gründungs- und Anlaufphase sinnvoll sind. Größere Unternehmen mit einem differenzierteren Anforderungsprofil, z. B. Anzahl von Sicherungen und Speicherplatz, können mit Cloud Computing noch keine Kostensenkungspotentiale realisieren. Als Grund dafür wurde der relativ hohe Preis für Spei-

cherplatz (und somit auch für Sicherungen) identifiziert. Dabei sollte beachtet werden, dass sich durch den zunehmenden Wettbewerb zwischen Cloud-Anbietern Preis-senkungen ergeben können [29].

In der Praxis finden sich auch hybride Infrastrukturlösungen, die Cloud Computing mit der eigenen IT-Infrastruktur von Unternehmen kombinieren. Diese wurden in der Arbeit nicht thematisiert, stellen jedoch nach Ansicht der Verfasser ein interessantes und weiter zu untersuchendes Forschungsfeld dar.

Literatur

1. BITKOM: Umsatz mit Cloud Computing steigt über 5 Milliarden Euro, http://www.bitkom.org/de/markt_statistik/64086_71376.aspx
2. Khajeh-Hosseini, A., Greenwood, D., Smith, J.W., Sommerville, I.: The Cloud Adoption Toolkit: Supporting Cloud Adoption Decisions in the Enterprise. *Software: Practice and Experience* 42 (4), 447–465 (2012)
3. Khajeh-Hosseini, A., Greenwood, D., Sommerville, I.: Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS, <http://arxiv.org/pdf/1002.3492>
4. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A View of Cloud Computing. *Communications of the ACM* 53 (4), 50–58 (2010)
5. Matros, R., Stute, P., von Zuydtwyck, N.H., Eymann, T.: Make-or-Buy im Cloud-Computing – Ein entscheidungsorientiertes Modell für den Bezug von Amazon Web Services. *Bayreuther Arbeitspapiere zur Wirtschaftsinformatik*, Bayreuth (2009)
6. Lamberth, S., Weisbecker, A.: Wirtschaftlichkeitsbetrachtungen beim Einsatz von Cloud Computing. In: Pietsch, W., Krams, B. (eds.): *Vom Projekt zum Produkt*. LNI, Vol. 178, pp. 123–136. GI, Bonn (2010)
7. Thorsteinsson, P.B.: *On the classification and estimation of costs in information technology*. University of Iceland, Reykjavik (2010)
8. Brandl, R.: *Cost Accounting for Shared IT Infrastructures*. Gabler, Wiesbaden (2008)
9. Henneberger, M., Strebler, J., Garzotto, F.: Ein Entscheidungsmodell für den Einsatz von Cloud Computing in Unternehmen. *HMD – Praxis der Wirtschaftsinformatik* 47 (275), 76–84 (2010)
10. Deelman, E., Singh, G., Livny, M., Berriman, B., Good, J.: The Cost of Doing Science on the Cloud: The Montage Example. In: *Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, article No. 50. IEEE Press, Piscataway (2008)
11. McKinsey & Company: Clearing the air on cloud computing, http://www.isaca.org/Groups/Professional-English/cloud-computing/GroupDocuments/McKinsey_Cloud%20matters.pdf
12. Durkee, D.: Why Cloud Computing Will Never Be Free. *Communications of the ACM* 53, 62–69 (2010)
13. Repschläger, J., Zarnekow, R.: *Cloud Computing in der IKT-Branche*. *Research Papers in Information Systems Management*, Vol. 2. TU Berlin, Berlin (2011)
14. Onpulson: Weltweite Backup-Studie: Deutscher Mittelstand nimmt Spitzenposition ein, <http://www.onpulson.de/themen/3315/weltweite-backup-studie-deutscher-mittelstand-nimmt-spitzenposition-ein>
15. Fritz & Macziol: AIX 5L- und Linux-Server, http://www.fum.de/webfm/fmhome.nsf/pages/ibm_system_p

16. Focus Online: Deutsche Firmen sind trotz Krise zuversichtlich, http://www.focus.de/finanzen/news/dihk-sagt-wirtschaftswachstum-voraus-deutsche-firmen-sind-trotz-krise-zuversichtlich_aid_760113.html
17. Hüske, F.: Entwurf und Evaluation eines Algorithmus zur Verteilung von materialisierten Sichten unter Berücksichtigung der Eigenschaften von RFID-Daten. Universität Ulm, Ulm (2008)
18. Der Handel: Einzelhandel darf auch 2012 auf Wachstum hoffen, <http://www.derhandel.de/news/finanzen/pages/Konjunktur-Einzelhandel-darf-auch-2012-auf-Wachstum-hoffen-8252.html>
19. Deutscher Wetterdienst: Datenmanagement – Einleitung, http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&T1400657811144045287327gsbDocumentPath=Content/Oeffentlichkeit/TI/TI1/Informationstechnik/Datenmanagement/German/Datenmanagement__teaser__dt.html
20. Deutscher Wetterdienst: Neuer Superrechner für den Deutschen Wetterdienst, www.dwd.de/bvbw/generator/DWDWWW/Content/Presse/Pressekonferenzen/2009/PK__17__03__09/20090317__Rede__Hoffmann,templateId=raw,property=publicationFile.pd
21. IHK-Startup: Studie zur Wachstumsfinanzierung im aktuellen Marktumfeld, <http://www.ihk-startup.de/themen-gruender/news/studie38.html?type=123>
22. Amazon Web Services: Amazon S3 – Preise, <http://aws.amazon.com/de/s3/pricing/>
23. Amazon Web Services: Amazon EC2 – Preise, <http://aws.amazon.com/de/ec2/pricing/>
24. Amazon Web Services: Amazon EC2-Instancetypen, <http://aws.amazon.com/de/ec2/instance-types/>
25. Ostermann, S., Iosup, A., Yigitbasi, N., Prodan, R., Fahringer, T., Epema, D.: A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing. In: D.R. Avresky et al. (eds.): Cloud Computing. LNICST, Vol. 34, pp. 115–131. Springer, Berlin (2010)
26. Smith, V., Roberts, S., Berman, D.: The Green Grid Energy Policy Research for Data Centres. The Green Grid Association Whitepapers, Beaverton (2009)
27. StepStone: Der StepStone Gehaltsreport 2012. StepStone Deutschland, Düsseldorf (2012)
28. Harms, R., Yamartino, M.: The Economics of the Cloud. Microsoft, Redmond (2010)
29. Etro, F.: The Economics of Cloud Computing, <http://www.intertic.org/Policy%20Papers/Report.pdf>

Foundations for the Integration of Enterprise Wikis and Specialized Tools for Enterprise Architecture Management

Max Fiedler, Matheus Hauder, and Alexander W. Schneider

Technische Universität München (TUM),
Chair for Informatics 19 (sebis),
Boltzmannstr. 3, 85748 Garching by Munich, Germany
{max.fiedler,matheus.hauder,alexander.schneider}@tum.de

Abstract. Organizations are challenged with rapidly changing business requirements and an ever-increasing volume respectively variety of information. Enterprise Architecture (EA) and its respective management function are considered as means to overcome these challenges. Appropriate tool support to this end is an elementary success factor to guide the EA management (EAM) initiative. Nevertheless, practitioners perceive currently available tools specialized for EAM as not sufficient in their organizations. Major reasons are inflexible data models as well as missing integration with processes and their focus on expert users. Regarding these limitations Enterprise Wikis provide practice proven solutions already exploited by organizations. These Enterprise Wikis are able to extend the capabilities of existing EA tools to cope with unstructured information and leverage a better utilization of structured EA information. In this paper we present the foundations for an integration of specialized EAM tools and Enterprise Wikis. We elaborate scenarios for both tool species using a practitioner survey and differentiate four integration cases.

Keywords: Enterprise Architecture Management, Enterprise Wiki, Tool Support, Collaboration, Knowledge Management

1 Introduction

Organizations are challenged by rapidly changing business requirements and external drivers that compel them to adapt their information technology (IT) management practices [1-2]. EA and its respective management functions are increasingly applied in industry to overcome these challenges and realize a better alignment of business and IT [3-4]. According to the ISO Standard 42010, an EA is the “fundamental organization of a system (enterprise) embodied in its components, their relationships to each other, and to the environment, and the principles guiding its design and evolution” [5]. Therefore, the EA covers business as well as IT aspects to provide a holistic view of an organization and support decision makers with appropriately represented structured information thereof. In current practices this information is typically cap-

tured in an EA model that covers the information demands of decision makers and presented in stakeholder specific views [6].

Although organizations already apply notations, models, methods, and EA tools to analyze this structured information content, many of these initiatives are not successful in practice [7]. Aier et al. [7] revealed long-term success factors for EA that were identified within a study of six companies over a timespan of eight years. Besides formal structure and processes the long-term success factors include aspects on the training, communication, intensified EA representation in projects, and tool support. In current practices specialized tools for EAM are developed to support architects with the formal structure and processes, rather than with collaboration among stakeholders in the organization and unstructured information content. We argue that many initiatives do not solely rely on this structured EA information and the long-term success factors for EAM require, e.g., deployment of collaboration tools. In this sense our goal is not to replace existing tools that are specialized for EAM, but rather extend their capabilities with practice proven solutions that are already widely used in organizations for related purposes.

Among these practice proven solutions are Enterprise Wikis that are used in organizations in various operational areas, e.g., requirements management, corporate platforms, project communication and documentation, and knowledge management in IT departments [8]. In comparison to traditional Wikis they often have a larger functionality. In general, their purpose is to improve knowledge sharing and collaboration between employees in organizations. They allow users to easily manipulate content and contribute to the knowledge base. Typically these tools allow the management of unstructured information using, e.g., Wiki pages, file attachments, and multimedia content as well as the interlinking of items. In previous work the application of Enterprise Wikis has been proposed to support the start of an EA endeavor [9]. As a result particular Wiki functionalities were used in this context but not integrated with existing EA tools.

At the same time specialized tools for EAM provide sophisticated capabilities for the modeling of the EA and the generation of visualizations thereof. These visualizations are common means to support decision makers with stakeholder specific representations of relevant information. While previous efforts integrated Wikis and modeling tools for software architectures [10], related efforts in the field of EAM are not available to the best of the authors' knowledge. Main goal of our research is to enhance existing structured information, present in the form of models on the EA level, with unstructured information provided by Enterprise Wikis, in order to support long-term success factors. We expect future research to investigate the integration of structured as well as unstructured information in the context of EAM.

Therefore, we provide a research fundament on the integration of Enterprise Wikis and specialized tools for EAM using a practitioner survey as empirical evidence for our approach. In order to lay the ground for future research we elaborate the following research questions:

- Can the integration of Enterprise Wikis and specialized EA tools provide added value for an organization?

- Which EAM scenarios are more suitable for which tool?
- What are challenges that have to be solved for this approach?

The main contribution of this paper is the discussion of the first two research questions. To the best of our knowledge, it presents the first study on the integration of these tool species for EAM. The findings presented in this paper are the foundation for our future research goal to integrate unstructured information in EA tools and enable better support for collaboration in this context.

2 Related Work

There is a variety of research on EAM concerning methods, implementation, tools, etc. The research on tools focuses mainly on repository based EAM modeling suits and how they can be used to best support the EAM efforts of an organization.

Keuntje et al. [11] differentiate the main tasks of EAM and how they are supported by different software solutions. They compare the performance of different EAM software packages available on the market on each of these categories regarding specific needs of different companies. For this purpose, they classify them into meta-model centered, method centered or process centered, EAM solution or EAM platform, EAM warehouse or modeling tool reflecting how much configurability vs. off-the-shelf functionality they offer as well as how far they integrate information of other systems.

Braun et al. [12] describe the basic requirements an EAM tool should meet (metamodel-based repository with visualization capability) and strongly advocate the use of such a professional tool for any organization. They find that “Enterprise Architecture is evolving into an active concept that supports business systems design, i.e. to contribute to the analysis and proactively support the optimization of business strategies, organizational structures, business processes, information flows, application structures as well as the underlying information systems” [12]. An adequate EAM tool would thus have to “support the development, storage, communication respectively presentation and enhancement of all relevant enterprise architecture artifacts” [12].

EAM is knowledge intensive and managing the related knowledge is crucial for its long-term success. Buckl et al. [13] evaluate different EAM approaches from a knowledge management (KM) perspective based on the KM cycle of Probst et al. [14]. Strengths of the evaluated approaches lie in the development and use of knowledge, whereas they all show weaknesses in the distribution and preservation of knowledge. The latter is only mentioned as a challenge to be addressed by a tool. Probst et al. [14] state user acceptance and respectively proper maintenance as crucial features for tool support of the distribution and preservation activities. Furthermore, neither of the analyzed approaches specifies how to incorporate knowledge from other sources.

Lucke et al. [15] present a case study of six software projects in six different companies, analyzing how well the chosen architectures support the core KM processes defined by Probst et al. [16] and how much they have been used by different target

audiences. Their observations strongly suggest that the choice of architecture does influence KM and that the chosen architecture can support KM as an interactive index of contents, i.e. as a “knowledge map” [15] (enhanced by hyperlinks between KM artifacts) and by creating a common vocabulary thus facilitating cross-team communication. Proper introduction of the architecture to all required audiences is crucial. Incorporating a well-known architecture and offering different perspectives on each target audience respectively have been proven helpful. Lucke et al. [15] mention the need to keep knowledge alive and accessible for decades as an increasingly important challenge to be addressed by further research.

Farenhorst et al. [17] present a case study creating and using a Wiki for the EA department of a large organization, where none of the involved architects had worked with a Wiki before. They then compare it to existing tools for architecture knowledge management as an alternative to those. In their experiment they used a Confluence Wiki and compared it to existing tools for architecture knowledge management such as ADDSS [18], DGA DDR [19], PAKME [20], and Archium [21]. They find that if properly introduced the strengths of a Wiki solution outweighs its weaknesses when compared to the aforementioned tools, especially when it comes to support for online communication, managing non-architectural knowledge, collaboration support and involving and integrating knowledge and communication needs of other functions within the company. The dedicated tools only prevailed in modeling and retrieving architectural knowledge concepts. The lacking out-of-the-box support for these concepts within a Wiki solution needs to be compensated by specifically developed and introduced plug-ins, templates and enriched meta-data. Other general advantages of the Wiki identified are its support for integration with other tools, its intuitive interface and low learning curve.

Hirsch et al. [10] present and evaluate three prototypes for a Visual Wiki, combining Thinkmap visualization software with three different kinds of Wikis: freebase, Confluence and MediaWiki. Their goal is to enhance the respective Wiki with a visual representation of its knowledge space to be used to navigate and organize the knowledge space (with some additional search algorithms integrated). The three Visual Wikis differ in the way the nodes and edges of the graphical representations can be edited and how they are referenced or “mapped” to the textual Wiki content. They also differ in the consistency between textual contents and graphical representation of the knowledge space (two-sided, one-sided, no consistency). Furthermore, Hirsch et al. [10] describe a conceptual model of a Visual Wiki that can be used to design and analyze different implementations. They discuss the results of a survey with 14 participants answering how much the visual component improved the respective “basic” Wiki with regards to: knowledge-creation, -organization, -distribution and -search. For all four tasks the participants felt the visual components of the prototypes increased the usefulness of the respective Wiki, with the biggest improvements for organization and search. The differences in the ratings of the three prototypes were in part explained by their conceptual differences and in part by their differences in technical maturity and ease of use.

Su et al. [22] describe and discuss a software architecture documentation tool (KaitoroBase) built within the Visual Wiki Thinkbase presented in [10]. The tool

supports non-linear navigation and visualization of Software Architecture Documents (SAD) created using the Attribute-Driven Design. It does not require the use of a specific Architectural Description Language within the SADs. It combines a graph-based interactive visualization of the high-level structure of the topics (nodes) in an SAD and their relations (edges) with detailed documents attached to each of these elements.

3 Research Methodology

In order to substantiate our hypothesis that additional benefits can be achieved if an Enterprise Wiki is integrated with a specialized EAM tool, we started our research with the goal to identify concrete tasks in which such an integration would provide benefits. We analyzed related literature, e.g., [23] and [11], and identified nine potential scenarios. These scenarios serve as a basis to analyze different ways of integrating an Enterprise Wiki and a specialized EAM tool and, additionally, to confirm the relevance of this topic.

To demonstrate the relevance of such a tool mix and to identify scenarios in which an Enterprise Wiki might be better suited than a dedicated EAM tool, we performed an expert survey among 105 practitioners from who 54 describe themselves as Enterprise Architects. Therein, the participants were asked to categorize each of the nine previously identified EAM scenarios according to the tool they prefer for this task. Because the survey results revealed expected as well as unexpected insights, we then analyzed the characteristics of both types of tools on a general level. Consequently, their respective strengths and weaknesses were identified and their preferred usage within the different scenarios could be explained. Since the practitioners' tool preferences for concrete EAM scenarios are mostly homogenous, the scenarios can be divided into three categories: Wiki scenarios, EAM tool scenarios and undecided. Therefore, these categories will form the basis for future research in integrating both tools. As it would be hardly possible to analyze every conceivable EAM scenario, we categorized the identified scenarios into four cases that can be used to understand data flows between an Enterprise Wiki and a specialized EAM tool and respective use cases.

An online questionnaire was sent to over 1.100 domain experts in the field of EA and strategic IT management. Next to these invitations the survey was published in interest groups that are specialized on these topics and distributed during industry workshops and scientific conferences. In total the survey was available for 21 days and published on October 12th, 2012.

The questions investigated in this paper were enclosed within a larger survey that targets to evaluate current practices and challenges in EAM. In the first part of this survey, the dissemination of EAM tools and Enterprise Wikis in organizations were evaluated. These findings are shown in Section 4 of this paper. The second part evaluated the EAM scenarios identified in literature and their support by Enterprise Wikis and specialized EA tools according to the domain experts. These scenarios are shown in Section 5 and form the basis for the identification of four integration cases.

In total, we received 178 answers with organizations from, e.g., Germany (31.43%), USA (17.14%), Netherlands (6.67%), India (4.76%), United Kingdom (4.76%), Switzerland (5.71%), and Sweden (2.86%). During the survey 73 (41.01%) organizations dropped out or were deleted from the empirical basis in case they answered on behalf of the same organization resulting in 105 complete answers. The distribution of the industry sectors of the participating organizations is shown in Table 1. The majority of the organizations are acting in the *IT Consulting* sector followed by the *Finance*, and *IT Products and Services* sectors. Table 2 summarizes the job titles of the participants with the largest group of 54 (51.43%) being *Enterprise Architects*. *IT Architects* (14.29%) and *Consultants* (11.43%) follow this group. The consultants were asked to answer on behalf of a specific customer in the survey.

Table 1. What is the industry sector of your organization?

Industry Sector	n	% of all
IT Consulting	24	22.86%
Finance	18	17.14%
IT Products and Services	9	8.57%
Telecommunications	6	5.71%
Public Service	8	7.62%
Education	5	4.76%
Production and Manufacturing	6	5.71%
Transportation and Logistics	3	2.86%
Health	2	1.90%
Management Consulting	4	3.81%
Other	20	19.05%

Table 2. What is your job title in the organization?

Job Title	n	% of all
Enterprise Architect	54	51.43%
IT Architect	15	14.29%
Consultant	12	11.43%
Business Architect	6	5.71%
IT Operations	3	2.86%
CxO	6	5.71%
Software Engineer	1	0.95%
Solution Architect	1	0.95%
Business Analyst	1	0.95%
EA Analyst	1	0.95%
Other	1	0.95%

In average the participants have 6.7 years of experience in the field of EAM with a standard deviation of 4.88. The organizations apply EAM methods in average for 5.5 years with a standard deviation of 5.61. Among the organizations 48.57% model the current state of the EA. 29.52% model planned states and only a minority of 14.29% models a long-term target state of the EA.

4 Dissemination of Tools for EAM

In this section we provide an overview of the used Enterprise Wikis and specialized tools for EAM in organizations. For this purpose, the participating organizations were asked which tools they use and whether the tools are integrated or used separately. These questions are necessary to justify our assumption that many organizations already have the respective tools in operation. The results from this question are summarized in Table 3 and explained in the following.

Table 3. Which tools does your organization use for EAM?

Tools for EAM	n	% of all
EA Tool	36	34.29%
Wiki & EA Tool separately	37	35.24%
Wiki & EA Tool integrated	6	5.71%
Wiki	9	8.57%
None of these	14	13.33%
No response	3	2.86%

While 36 (34.29%) organizations only use an EA Tool for the EAM initiative, 37 (35.24%) organizations use a Wiki and an EA Tool separately. Only a minority of 9 (8.57%) organizations use only a Wiki to support the EAM initiative. Among the participants are also 6 (5.71%) organizations that already use an integrated solution of an EA Tool and an Enterprise Wiki. With these results, we confirm our assumption that organizations already apply both tool species for EAM but still struggle to integrate both.

Table 4. Do you think a Wiki can be useful (in addition to another tool) for EAM?

Wiki for EAM useful	n	% of all
Yes	98	93.33%
No	7	6.67%
No answer	0	0.00%

Table 5. Do you plan to use a Wiki for EA management in the future?

Wiki for EAM in future	n	% of all
Yes	63	60.00%
No	13	12.38%
No answer	29	27.62%

When asking whether it can be useful for organizations to use an Enterprise Wiki in addition to a specialized EA tool 98 (93.33%) organizations confirmed this presumption (Table 4). Only a small minority of 7 (6.67%) organizations does not think an additional Enterprise Wiki can be useful to extend an existing EA Tool. The relevance of the research conducted in this paper is also approved with the findings shown in Table 5. 63 (12.38%) organizations plan to use an Enterprise Wiki for EAM in the future. Only 13 (12.38%) organizations can exclude that they plan to use an Enterprise Wiki for EAM in the future while 29 (27.62%) organizations were not able to provide an answer to this question.

5 Scenarios for Enterprise Wikis and Specialized EAM Tools

We identified nine scenarios relevant for EAM tools from literature [11], [23] and our practical experience from industry projects. These scenarios are evaluated in this section on the basis of the aforementioned survey among EA practitioners. The findings from this survey are confronted with key characteristics identified in literature for the investigated tool species:

Definition of terms: A glossary or definition of terms is used in an organization to create a common terminological basis. This definition of terms contains a list of rele-

vant terms within the organization and an informal description for every term. Typically these terms are interlinked with each other in order to identify e.g., synonyms, categories, etc. Next to creating such a definition it also has to be made available to a potentially large community.

Providing references to EA relevant documents: Information about the EA is foremost stored in a central model repository. Besides this structured information on the EA model, additional information is distributed over many documents in the organizations, e.g., excel sheets, power point presentations, and word documents. These documents often contain information highly relevant for EAM and connecting them to the EA tool can yield benefit. Examples for relevant documents are strategy documents, executive board presentations, or governmental regulations.

Document guidelines and FAQs: Guidelines and FAQs published by the architecture teams need to be distributed to EA stakeholders throughout the organization. Thereby, EA stakeholders are allowed to search for specific topics concerning their daily work efficiently. This scenario helps to improve the communication with management and operations processes in the organization.

Annotating visualizations: Visualizations are common means in EAM to support decision makers using stakeholder specific views. These visualizations can be annotated with additional detail or background information on the elements of the application landscape. This can be accomplished by extending a visualization that has been generated before. Examples for annotations are for instance a color-coding of application elements or additional symbols.

Discussion of business and EA requirements: Major goal of EAM initiatives is to improve the alignment between business and IT. This means IT capabilities have to be developed according to the business requirements. The discussion about these requirements and their impact on the EA can be preserved with unstructured information that is linked to elements in the EA model. The versioning of these requirements helps to track changes and supports the planning of future decisions on the EA.

Import, edit and validate EA information: Data previously collected with a spreadsheet or available from productive systems in the organization [24] are imported or used to maintain the EA model. Only a limited number of people are usually involved in this task, e.g., as information providers. This scenario also includes capabilities that deal with data transformations and validations of the model consistency.

Analyze EA information: Typically the analysis of EA information is performed by various kinds of visualizations (for demand management, portfolio management, or infrastructure management, etc.) that are preconfigured and can be customized depending on the information demand of different stakeholders. Besides visualizations, the use of metrics that assess certain aspects of the EA can be used to analyze the development of the organization. Stakeholders for this scenario are usually architects or decision makers in the management board.

Simulate impacts of EA decisions: In recent work, models are evaluated with respect to assessment properties, e.g., [25]. A formalized theory can be then used to describe and compare alternative EA decisions. With such evaluation capabilities, impacts of EA decisions can not only be documented and considered but also analyzed.

EAM Activities and Workflows: Workflows and activities can support the architects and stakeholders with EAM related tasks, such as among others the documentation of the EA, development of planned states, or architecture reviews.

Next to these examples, further scenarios were identified in this research, e.g., collaborative work, support for lightweight access, editing of visualizations, but omitted due to space limitations in this paper, since they can be classified analogously within the identified cases (Section 6). Within the scenario *analyze EA information* several typical scenarios for EAM support, e.g., Landscape Management, Demand Management, Infrastructure Management, are summarized since they are typically targeted by specialized tools for EAM.

Table 6. Please decide whether a Wiki or a specialized EA tool is better to fulfill the following scenarios (n=105)?

Scenario	Wiki	Rather Wiki	Neither	Rather EA Tool	EA Tool	No Response
Definition of terms	52	24	5	5	12	7
Providing references to EA relevant documents	40	24	7	14	11	9
Document guidelines and FAQs	54	25	7	5	6	8
Annotating visualizations	10	7	5	32	42	9
Discussion of business and EA requirements	28	22	20	11	12	12
Import, edit, and validate EA information	3	3	4	18	64	13
Analyze EA information	5	1	3	19	69	8
Simulate impacts of EA decisions	2	2	6	18	65	12
EAM activities and workflows	10	8	10	18	50	9

The survey results for the scenarios with the classification according to the appropriate tool species are shown in Table 6. The classification for every scenario ranges from Wiki and rather one of the species to the EA Tool. In the following these scenarios are assigned to one of the tools depending on the assessment of the organizations.

5.1 Enterprise Wikis

The majority of the organizations (49.52%) expect the *definition of terms* to be within the Wiki. Only a small fraction of 12 (11.43%) organizations completely disagree and expect it to be within the EA Tool. The tendency is also supported by 24 (22.86%) organizations that foresee the definition of terms rather in the Wiki. Major reasons for the applicability of Wikis for this purpose can be found in literature. A Wiki allows everyone to view and edit content and create hyperlinks that connect Wiki pages with each other [26-28]. This is especially useful for the definition of terms, since the content needs to be made available for the entire staff in the organization and interlinking of definitions is required.

References to EA relevant documents are better supported by Enterprise Wikis. 40 (38.10%) organizations rate Wikis as more suitable for this purpose. The opposite (EA tool) opinion is only supported by 11 (10.48%) organizations. Furthermore 24 (22.86%) organizations expected the references to EA relevant documents to be rather in the Wiki. The ability of Wikis to create file attachments is the basis for referencing EA documents [28]. The management of these documents is supported by the integrated full-text search [26], [28]. Properly referenced EA documents help architects to find the required information more efficiently. This also includes additional information that is not being captured in the model of the EA Tool.

One of the main goals of the EAM initiative is the establishment of principles and guidelines to manage the evolution of the organization. The *documentation of guidelines and FAQs* is supported well by a Wiki according to 54 (51.43%) organizations. This is confirmed by 25 (23.81%) organizations that mentioned rather the Wiki for this scenario. Only 6 (5.71%) organizations think a specialized EA tool is more suitable. Several of the strengths of a Wiki are useful for this scenario. The possibility to use text formatting and layouting [26], [28], as well as the fast built-in full-text search [26], [28] allows for the creation of appropriate information content by and for EAM stakeholders. In addition, Wikis allow the linking to pages not yet written in order to show a need for information [27-28]. These links are highlighted in the system and help identifying the information need from stakeholders. Another example for this scenario is the technical documentation or manual of the EA tool that benefits of an integration of both tool species.

According to the survey results, the Enterprise Wiki is more suitable to *discuss business and EA requirements*. 28 (26.67%) organizations expect it to be only in the Wiki. In addition, 22 (20.95%) organizations rate rather in the Wiki and 12 (11.43%) organizations in the EA tool only. Especially the possibility to easily edit and view content makes the Enterprise Wiki useful for this scenario [26-28]. Furthermore arbitrary and unstructured content can be placed in Wikis [28]. Integration with the EA tool can be beneficial to relate the discussed requirements in the Wiki to EA elements retrieved from the structured model. At the same time tracing requirements from the EA tool back to unstructured information or documents becomes possible in case they are properly documented.

5.2 Modeling and EA Tools

Among the participants, 42 (40.00%) stated that the *annotation of visualizations* needs to be supported by the EA Tool. With 32 (30.48%) organizations, the second largest group stated that rather the EA Tool is more suitable. Only 10 (9.52%) organizations have the opposite opinion for this scenario and expect it to be in the Wiki. Existing EA tools have already largely adopted visualizations techniques [23]. Therefore, domain experts expect this scenario to be rather in the EA tool.

The scenario *import, edit, and validate EA information* has a strong rating with respect to the EA tool. 64 (60.95%) organizations expect it to be performed in the EA tool. 3 (2.86%) organizations perform this scenario in the Wiki instead. Since the Enterprise Wiki usually does not support the modeling of complex structures, the EA

tool should execute this scenario. Only in case the documentation of the EA does not have enough level of formalization an approach as proposed by Buckl et al. [9] is reasonable. Recent research provided mechanisms to automatically document EA elements from productive system [24]. The proposed solution connects external information systems with the repository of the EA tool to automate the documentation.

Based on the formalized EA elements *analyze EA information* provides visualizations to support, e.g., Demand Management, Portfolio Management, or Infrastructure Management, or computes metrics on the EA for decision makers. 69 (65.71%) organizations classify this scenario to be in the EA tool. Only 5 (4.76%) organizations mentioned it should be within the Enterprise Wiki. These organizations conduct a solely Wiki-based approach in their EA endeavor and therefore analyze EA information only in this tool. Similarly, *simulating impacts of EA decisions* is placed in the EA tool according to 65 (61.90%) organizations. This scenario also requires a structured model to create an assessment theory. The execution of *EAM activities and workflows* should be performed in the EA tool according to 50 (47.62%) organizations. However, 10 (9.52%) organizations shared the opposite opinion and expect this scenario in the Enterprise Wiki.

6 Classification of the Tool Scenarios

The results of the survey clearly show that practitioners would prefer an Enterprise Wiki in comparison to a specialized EAM tool in four of our nine scenarios, whereas they prefer an EAM tool compared to a Wiki in five scenarios. If both types of tools should be used within different EAM scenarios, an integration of both is inevitable. To design an optimal integration strategy, the appropriate type of tool has to be determined for every conceivable EAM scenario. Although we were able to analyze nine concrete scenarios with our survey, future research has to proceed on a more abstract level. Therefore, we derived four general tool usage cases that are described in the following.

Case 1 encompasses scenarios in which highly structured models provided by specialized EAM tools are not needed. For example, *definition of terms* and *references to EA documents* belong to this case. The documentation of EA relevant information is performed in a merely unstructured manner and there is no need to add structural overhead in the future.

Case 2 comprises scenarios in which an EA description is initially done without a preexisting structure (model), but after a while transferred to the structured EAM tool. For example, *evaluation of external requirements* belongs to this case. Reasons for using a Wiki before an EAM tool might be that a) the appropriate documentation structure is not known upfront or b) the need for a structured documentation evolves over time.

Case 3 summarizes scenarios in which structured data is transferred from a specialized EAM tool to an Enterprise Wiki. The only scenario already identified for this case is the removal of unused EA data. In this scenario, data for which restrictive data models are not needed anymore is swapped to a Wiki where it remains accessible.

Case 4 describes scenarios in which highly structured data or the special features of EAM tools are needed. For example, *simulate impacts of EA decisions*, *analyze EA information* and *annotating visualizations* belong to this case. Activities in that kind of scenarios often relate to analysis, simulation or validation tasks.

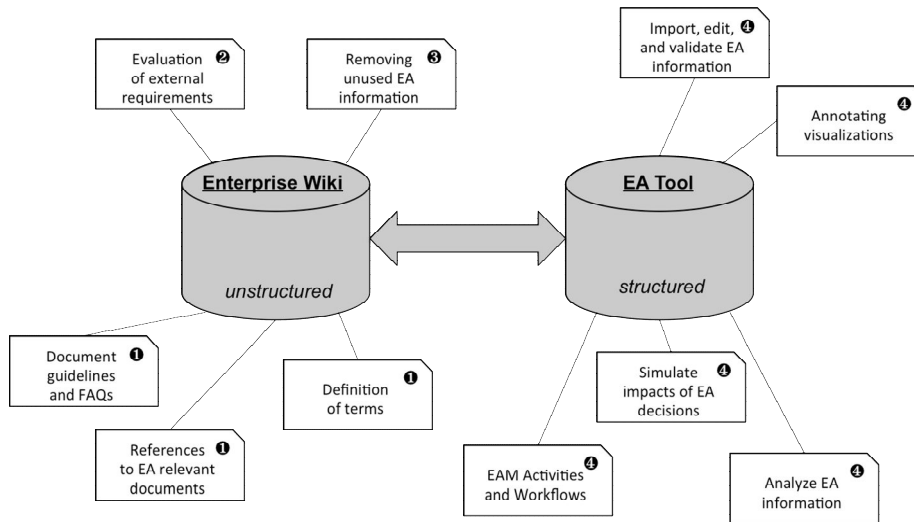


Fig. 1. Scenarios for the Enterprise Wiki and the EA tool

Figure 1 shows the conceptual integration of an Enterprise Wiki and a specialized EAM tool. Therein, each scenario is linked to the tool practitioners prefer to execute it. The small number in the upper-right corner indicates the general case the scenario is part of.

7 Discussion and Future Research

From the survey presented in Sections 4 and 5, we learnt that the parallel usage of an Enterprise Wiki and a specialized EAM tool is attributed to additional support for typical EAM tasks. At least 40% of the organizations actually use both types of tools and 5% decided to integrate them. We assume that such integration is especially useful in certain scenarios, e.g., in *providing references to EA relevant documents*. The Wiki can provide an easier way to create such references, but knowledge about the references themselves might also be beneficial if an enterprise architect is working with a specialized EAM tool. It can thus be concluded, that the integration of Enterprise Wikis and specialized EAM tools provides added value for organizations, which answers our first research question.

By analyzing the survey results, we were able to show, that at least four EAM scenarios would benefit from the usage of an Enterprise Wiki integrated with a specialized EAM tool. These scenarios are: *definition of terms*, *providing references to EA*

relevant documents, document guidelines and FAQs, and discussion of business and EA requirements.

The third research question posed in Section 1 requires the identification of potential challenges of an integration of an Enterprise Wiki and a specialized EAM tool. The existence of challenges is affirmed by the survey results, which show that 60% of the participants plan to use a Wiki in addition to a specialized EAM tool, while currently only 5% were able to implement an integrated approach.

The challenges in the context of an integration of both types of tools might consist of the following. If, for instance, the Enterprise Wiki should be able to provide links to architecture objects or preconfigured visualizations documented or managed by the EAM tool, the EAM tool would have to provide direct links (e.g. URLs) to its entities. If the Enterprise Wiki is to provide the ability to search for architectural entities managed by the EAM tool a respective interface would have to be provided by the EAM tool. While many of the scenarios described in Section 5 require some kind of information exchange between the two types of tools an appropriate technical implementation remains unclear, since many different technologies could be applied, e.g., Web services or HTTP requests.

When two independent tools use partially the same data basis ensuring data consistency is an additional challenge. In a similar context Hirsch et al. [10] distinguish three types of consistency: no consistency, one-way consistency and two-way consistency. Besides the implementation of a consistency assurance mechanism the identification of the best suitable type of consistency remains an open challenge.

Architectural information, e.g. IT strategy, is often subject to restrictive access policies within an EAM tool. If such information is used or referenced within an Enterprise Wiki the respective stakeholders might require the same access policies in the Wiki as the EAM tool implements them. Such a requirement stands directly in contradiction with the suggestion of Farenhorst et al. [17] to enable a maximum of collaboration in a Wiki. Therefore, appropriate ways to address both requirements have to be explored.

Up to now, the four different interaction cases of Enterprise Wikis and EAM tools as described in Section 6 have not been evaluated in practice yet. They are deduced from the nine scenarios but additional research is necessary to confirm their integrity. Therefore, our future work will elaborate detailed interactions with a prototypical implementation with our industry partners.

8 Conclusion

Existing tools for EAM are mainly focused on formal structures and processes neglecting collaboration among stakeholders as well as unstructured information content. For the latter two, Enterprise Wikis successfully provide solutions already widely applied in organizations. In this paper foundations for the integration of Enterprise Wikis and specialized EA tools are presented. After an identification of common scenarios for EAM, a practitioner survey was performed in order to provide empirical evidence for our approach. The results from the survey indicate potential for a better

integration of both tool species. The identified scenarios were analyzed and assigned to a tool according to the findings from the survey and results from current literature. The findings presented in this paper lay the ground for future research on the integration of unstructured information content with EA models.

References

1. Aier, S., Gleichauf, B., Winter, R.: Understanding Enterprise Architecture Management Design - An Empirical Analysis. In: Proceedings of the 10th International Conference on Wirtschaftsinformatik WI 2011 (Zürich) (2011)
2. Ross, J.W.: Creating a Strategic IT Architecture Competency: Learning in Stages. MIT Sloan Working paper 4314-03 (2003)
3. Langenberg, K.: Enterprise architecture: What aspects is current research targeting. Laboratory of Systemic Modeling (2004)
4. Ross, J.W., Weill, P., Robertson, D.: Enterprise architecture as strategy: Creating a foundation for business execution. Harvard Business Press (2006)
5. International Organization of Standardization: ISO/IEC/IEEE 42010:2011 - Systems and software engineering -- Architecture description (2011)
6. Lankes, J., Matthes, F., Wittenburg, A.: Softwarekartographie: Systematische Darstellung von Anwendungslandschaften. In: Proceedings of Wirtschaftsinformatik 2005, pp. 1443–1462 (2005)
7. Aier, S., Schelp, J.: A Reassessment of Enterprise Architecture Implementation. In: ICSSOC/ ServiceWave 2009 Workshops. LNCS, Vol. 6275, pp. 35–47. Springer (2010)
8. Andersen, E.: Using Wikis in a Corporate Context. Handbuch ELearning. 11. Erg. (2005)
9. Buckl, S., Matthes, F., Neubert, C., Schweda, C.M.: A wiki-based approach to enterprise architecture documentation and analysis. In: Newell, S., Whitley, E.A., Pouloudi, N., Wareham, J., Mathiassen, L. (eds.): 17th European Conference of Information Systems ECIS, pp. 1476–1487 (2009)
10. Hirsch, C., Hosking, J., Grundy, J., Chaffe, T., Macdonald, D., Halytskyy, Y.: The Visual Wiki : A New Metaphor for Knowledge Access and Management. In: HICSS' 09, pp. 1–10. IEEE (2009)
11. Keuntje, J.H., Reinhard, B.: Enterprise Architecture Management (EAM) in der Praxis. Symposium Publisher (2012)
12. Braun, C., Winter, R.: A Comprehensive Enterprise Architecture Metamodel and Its Implementation Using a Metamodeling Platform. In: LNI, Vol. P-75, pp. 64-79. GI, Bonn (2005)
13. Buckl, B.S., Matthes, F., Schweda, C.M.: Future Research Topics in Enterprise Architecture Management – A Knowledge Management Perspective. In: LNCS, Vol. 6275, pp. 1-11. Springer (2009)
14. Probst, G.J.B.: Practical Knowledge Management: A Model That Works. Prism 2, 17–29 (1998)
15. Lucke, C., May, M., Kratzke, N., Lechner, U.: How do system and enterprise architectures influence knowledge management in software projects ? – An exploratory study of six software projects. LNI, Vol. P-152, pp. 69–82. GI, Bonn 2009
16. Probst, G., Steffen, R., Kai, R.: Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen. Gabler (2006)
17. Farenhorst, R., Vliet, H.V.: Experiences with a Wiki to Support Architectural Knowledge Sharing (2008)

18. Capilla, R., Nava, F., Pérez, S., Dueñas, J.C.: A web-based tool for managing architectural design decisions. *SIGSOFT Softw Eng Notes* 31, article 4 (2006)
19. Falessi, D., Becker, M., Cantone, G.: Design decision rationale: experiences and steps ahead towards systematic use. *ACM SIGSOFT Software Engineering Notes* 31 (5), article 2 (2006)
20. Babar, M.A., Gorton, I.: A Tool for Managing Software Architecture Knowledge. In: *Second Workshop on Sharing and Reusing Architectural Knowledge Architecture Rationale and Design Intent SHARK/ADI07 ICSE Workshops 2007*, p. 11 (2007)
21. Jansen, A., Ven, J.V.D., Avgeriou, P., Hammer, D.K.: Tool Support for Architectural Decisions. *IEEE*(2007)
22. Su, M.T., Hirsch, C., Hosking, J.: KaitoroBase: Visual Exploration of Software Architecture Documents. In: *2009 IEEE/ACM International Conference on Automated Software Engineering*, pp. 657–659 (2009)
23. Matthes, F., Buckl, S., Leitel, J., Schweda, C.M.: Enterprise Architecture Management Tool Survey 2008. *Information Systems Journal* 19, 1–21 (2008)
24. Buschle, M., Ekstedt, M., Grunow, S., Hauder, M., Matthes, F., Roth, S.: Automated Enterprise Architecture Documentation using an Enterprise Service Bus. In: *18th Americas Conference on Information Systems, Seattle* (2012)
25. Johnson, P., Johansson, E., Sommestad, T., Ullberg, J.: A Tool for Enterprise Architecture Analysis. *IEEE* (2007)
26. Ebersbach, A., Glaser, M., Heigl, R.: *Wiki: Web Collaboration*. Springer, Heidelberg (2005)
27. Bachmann, F., Merson, P.: Experience using the web-based tool wiki for architecture documentation (2005)
28. Cunningham, W., Leuf, B.: *The Wiki way: quick collaboration on the Web*. Addison-Wesley (2001)

Green IT-Service-Management: Eine empirische Voruntersuchung der konzeptionellen Grundlagen

Jörg Schneider und Markus Reiter

¹ Institut für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH und Universität des Saarlandes, Universitätscampus, Gebäude D 3 2, 66123 Saarbrücken

s9jvschn@stud.uni-saarland.de, markus.reiter@iwi.dfki.de

Abstract. Das IT-Service-Management (ITSM) bildet einen organisatorischen Rahmen zur Koordinierung von IT-Services und ist ein geeigneter Ansatzpunkt, um diese nachhaltig auszugestalten. Dieser Forschungsbeitrag leitet vor diesem Hintergrund Vermutungen über ein nachhaltig ausgerichtetes ITSM ab und prüft diese empirisch auf Plausibilität. Ziel ist es, zum Verständnis konzeptioneller Grundlagen und der Entwicklung neuer Forschungsfragen beizutragen. Nach einer Analyse der bestehenden Literaturlbasis erfolgt eine empirische Voruntersuchung unter deutschen IT-Dienstleistern. Das Ergebnis zeigt, dass ein deutliches Interesse an wissenschaftlich entwickelten Erweiterungskonzepten für etablierte ITSM-Standards besteht, an diese aber auch spezifische Anforderungen gerichtet werden. So unterscheiden sich Beeinträchtigungen der nachhaltigen Ausführung von IT-Services hinsichtlich der Dimensionen Priorität, Erkennbarkeit und Lösungskomplexität von herkömmlichen Servicebeeinträchtigungen. Die Befragungsergebnisse stützen darüber hinaus die Vermutung, dass es einen positiven Zusammenhang zwischen dem Reifegrad der ITSM-Rahmenwerkverwendung und einer Green-IT-Adoption gibt.

Keywords: Green IT, ITIL, Nachhaltigkeit, Empirische Voruntersuchung

1 Einleitendes

1.1 Motivation

Die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) geht in ihrem aktuellen „Umweltausblick bis 2050“ davon aus, dass sich ohne einschneidende politische Maßnahmen und unter Fortsetzung der bisherigen sozioökonomischen Trends ein „Schwund des Umweltkapitals“ fortsetzen kann, der „die während zweier Jahrhunderte erzielten Fortschritte bei der Anhebung des Lebensstandards zunichte machen“ könnte ([1], S. 3). In Anbetracht dieser ersten Zukunftsprognose ist der Handlungsbedarf zur Erreichung einer nachhaltigen Entwicklung akut. Diese muss konventionell zur gegenwärtigen Bedürfnisbefriedigung genügen,

ohne aber zukünftige Generationen daran zu hindern, ihren eigenen Bedürfnissen nachkommen zu können [2].

Der zunehmende Ressourcenverbrauch durch Informationstechnik wurde empirisch belegt und als optimierungsbedürftig bewertet [3]. Unter dem Begriff der *Green Information Technologies* (Green IT) wird daher die Entwicklung, Implementierung und Evaluation von Ansätzen zur Steigerung der Nachhaltigkeit von Informationstechnik sowohl theoretisch diskutiert als auch in der Unternehmenspraxis betrieben [4]. Der Schwerpunkt dieses, ab 2008 auch zusehends wissenschaftlich aufgegriffenen Themas [5], liegt in der ressourcenbezogenen Optimierung der IT-Infrastruktur selbst, weshalb es in einer engen Konzeptualisierung der Erreichung von ökologischer Nachhaltigkeit nur teilweise gerecht wird und an seine Grenzen stößt [6]. Die vielversprechendere „IT for Green“-Perspektive [7-8] sieht die Verwendung von IT als „enabler“ vor, um an anderer Stelle eine höhere Nachhaltigkeit zu ermöglichen [9]. Eng damit verwandt sind die Konzepte der *Green Information Systems* (Green IS) [6] und des *Green Business Process Managements* (Green BPM) [10], deren Betrachtungsgegenstände Informationssysteme respektive Geschäftsprozesse sind.

Eine bisher wenig auf Nachhaltigkeitspotentiale durch die Forschung untersuchte Disziplin ist das *IT-Service-Management* (ITSM). Das ITSM findet als Managementkonzept für den Entwurf, die Bereitstellung und die stetige Verbesserung von IT-Services wie u.a. Web- und Anwendungshosting, Websuche sowie Cloud-Computing breite Anwendung [11]. Die große Marktrelevanz von IT-Services, ihre starke Korrespondenz zur Auslastung von Rechenzentren sowie die ITSM-Standardisierung in Rahmenwerken erfordern jedoch eine ausführlichere Auseinandersetzung mit der Nachhaltigkeitsthematik als bisher geschehen.

1.2 Forschungsziele und Methodik

Dieser Beitrag hat zum Ziel, im Rahmen einer empirischen Voruntersuchung, konzeptionelle Grundlagen eines grünen IT-Service-Managements abzuleiten. Ausgehend vom Forschungsstand des Themas werden Vermutungen formuliert, die anschließend auf Plausibilität überprüft werden. Die damit verbundenen Erkenntnisse sollen der konzeptionellen Green ITSM-Forschung als neue Impulse dienen und bilden zugleich Ausgangspunkte für weitere empirische Forschungsvorhaben, die dann mit vermindertem Forschungsrisiko durchzuführen sind.

1.3 Aufbau des Beitrages

Kapitel 2 stellt das Green ITSM vor, das den Forschungsgegenstand dieser Untersuchung bildet. Es vermittelt die ITSM-Grundzüge, den Forschungsstand des Green ITSM und entwickelt eine Definition. Kapitel 3 widmet sich den Vorüberlegungen und der Ableitung von vermuteten Zusammenhängen. In Kapitel 4 werden Forschungsdesign, Durchführung und Ergebnisse der Studie vorgestellt, bevor Kapitel 5 den Beitrag mit einem Fazit abschließt.

2 Forschungsgegenstand: Green IT-Service-Management

2.1 Grundlagen des ITSM und des ITIL-Rahmenwerks

IT-Services sind Dienstleistungen, die Geschäftsprozesse von betrieblichen oder staatlichen Organisationen unterstützen und unter Anwendung von Informationstechnik erstellt werden [12]. Sie reichen von der Bereitstellung von Rechen- und Speicherkapazitäten bis hin zu entfernt betriebenen Anwendungssystemen, die Nutzern auf Nachfrage bereitgestellt werden („*Software-as-a-Service*“ (SaaS)). Zur Marktrelevanz dieser Dienstleistungen sei an dieser Stelle exemplarisch der SaaS-Markt herangezogen, dessen weltweiter Jahresumsatz nach Zahlen von Gartner noch im Jahr 2012 auf über 14,5 Milliarden US-Dollar ansteigen wird [13].

Das ITSM sieht vor, die primär technisch erbrachten IT-Services in einen organisatorischen Rahmen einzubetten, der eine strukturierte und ganzheitliche Koordination entlang eines Service-Lebenszyklus ermöglicht [14]. Ziel dabei ist, die Servicequalität stetig zu verbessern und eine hohe Kundenorientierung zu erreichen. Das ITSM-Rahmenwerk „ITIL“ ist dazu der de-facto Standard, mit einer vom IT Governance Institute auf 28% geschätzten Verbreitung bei IT-Unternehmen [15]. Als Sammlung von Best Practices in einem prozessorientierten Ansatz besteht ITIL V3 innerhalb der Kernbereiche „*Service Strategy*“, „*Service Design*“, „*Service Transition*“, „*Service Operation*“ und „*Continual Service Improvement*“ aus insgesamt 26 ITSM-Prozessen, die ihre korrespondierenden Phasen im ITIL-Service-Lebenszyklus abbilden [14].

2.2 Stand der Forschung des Green ITSM

Zur Recherche von Beiträgen zu Green ITSM wurden Schlagwortsuchen in Google Scholar ausgeführt, deren Ergebnisse manuell auf inhaltliche Eignung untersucht wurden. Die Literaturverzeichnisse dieser, sowie verschiedener Publikationen zu Green IT (z. B. [4-10]), wurden auf weitere geeignete Arbeiten untersucht. Es konnten sechs Publikationen ermittelt werden, die das Thema eindeutig adressieren.

So wird die Entwicklung von nachhaltigen IT-Services als weiterführender und ganzheitlicherer Ansatz, der auf frühe infrastrukturelle Green-IT-Maßnahmen folgt und u.a. die Geschäftsprozessoptimierung miteinbezieht, betrachtet [16]. Die Verbesserung der Nachhaltigkeit von ITSM-Prozessen kann demnach als potentieller Anwendungsfall des Green BPM betrachtet werden [17]. Die Thematik wird ferner in [18] losgelöst von Rahmenwerken unter dem Schwerpunkt des nachhaltigen Informationsmanagements betrachtet, wozu verschiedene Ansätze, darunter eine Sustainability-IT-Balanced-Scorecard und ein Reifegradmodell, zur Steuerung respektive Beurteilung des operativen Fortschritts entwickelt werden. Für das ITIL-Framework wurde zudem eine systematische Inhaltsanalyse erarbeitet, indem für alle ITIL-Prozesse die Relevanz in Bezug zu grüner Beschaffung, Ressourcenkonsolidierung, Energiemanagement sowie zur Entsorgung obsoleter Produkte untersucht worden ist [19]. Die Verfasser des Beitrages bemerken diesbezüglich, dass ITIL theoretisch in der Lage sei, auch ohne die explizite Aufführung von Green IT dessen Bestrebungen zu unterstützen. Im selben Beitrag stellen sie durch eine Umfrage aller-

dings fest, dass zwar einige Unternehmen in den Bereichen Energieeffizienz und umweltfreundliche Entsorgung aktiv sind, die große Variation der Umfrageantworten aber nicht für eine strukturierte und ausreichende Unterstützung durch das ITIL-Rahmenwerk spricht. Somit erscheint eine Erweiterung von ITIL vorteilhaft, wie in [20-21] vorgeschlagen. So kann das „Incident“-Konzept, das im Kontext von ITIL für eine Beeinträchtigung der Qualität eines IT-Services steht, um den Typ eines „Green Incidents“ ergänzt werden [21]. Dieses Ereignis würde dazu dienen, eine spezifische Gefährdung der nachhaltigen Serviceausführung anzuzeigen. In einem „Green Incident Mangement“-Prozess könnte der Umstand anschließend adressiert und bestenfalls behoben werden.

Insgesamt ist anzumerken, dass – ausgenommen [19] – argumentativ-deduktive Forschungsmethoden dominieren, um die nachhaltige Gestaltung des ITSM zu untersuchen. Erstaunlich ist, dass sich die hier angesprochenen Beiträge kaum aufeinander beziehen – es existieren lediglich 2 Querverweise zwischen den 6 Publikationen. Ebenso wurde der Terminus „Green ITSM“, obwohl in nicht-wissenschaftlichen Beiträgen präsent, bisher nur in [20] aufgegriffen und als solcher noch nicht genügend definiert. In den deutschsprachigen Arbeiten [18], [21] findet die Bezeichnung „nachhaltiges ITSM“ Verwendung. In dieser Arbeit wird jedoch „Green ITSM“ aufgrund seiner Prägnanz und größeren sprachlichen Unabhängigkeit bevorzugt.

2.3 Green ITSM: Entwicklung einer Definition

Das Green ITSM ist die systematische Analyse und Optimierung der Phasen des IT-Service Lebenszyklus' auf Nachhaltigkeit, jeweils auf prozessualer, infrastruktureller und unternehmenspolitischer Ebene. Prozessual stehen die ITSM-Prozesse im Vordergrund, die durch ein Geschäftsprozess-Reengineering an die Nachhaltigkeitsanforderungen angepasst werden müssen. Dazu bietet das Green BPM die nötigen Mittel, um beispielsweise den CO₂-Ausstoß von Geschäftsprozessen zu überwachen und diese dahingehend zu optimieren [10]. Auf infrastruktureller Ebene dominieren Maßnahmen des Green IT Bereichs, also primär die energieeffiziente Konstruktion und Inbetriebnahme von Rechenzentren, aber auch die Optimierung der Systeme an Arbeitsplätzen. Die unternehmenspolitische Ebene des Green ITSM bezweckt, die Zielsetzung zur Nachhaltigkeit organisatorisch zu verankern. Beispiele sind die explizite Formulierung von Nachhaltigkeitszielen, eine auf Nachhaltigkeit optimierte Beschaffungs- bzw. Entsorgungspolitik oder die regelmäßige Sensibilisierung der Mitarbeiter.

3 Theoretische Vorüberlegungen und Vermutungen

3.1 Überblick

Im Rahmen dieser Voruntersuchung stehen im Folgenden zwei Themen des Green ITSM im Fokus. Kapitel 3.2 betrachtet, welche Rolle verschiedene Rahmenbedingungen bisher bei der Realisierung des nachhaltigen ITSM eine Rolle gespielt haben.

Abschnitt 3.3 dagegen ist neuen Konzepten der Green ITSM-Forschung gewidmet, um deren Potenziale sowie Anforderungen besser verstehen zu können.

3.2 Rahmenbedingungen der nachhaltigen Ausrichtung des ITSM

Die nachhaltige Ausrichtung des ITSM ist eine Aufgabe, die nicht in einem Schritt oder durch eine Maßnahme, sondern konsekutiv als Prozess der Entwicklung in Richtung des Ziels stattfindet. Diesbezüglich kann davon ausgegangen werden, dass dieser Weg, bezogen auf verschiedene Unternehmen und ihre Spezifika, auch unterschiedlich bestritten, und v.a. durch verschiedene Rahmenbedingungen beeinflusst wird.

Dem „Green IT Adoption Model“ (GITAM) liegt eine unüblich weite Konzeptualisierung des Green-IT-Begriffs zugrunde, welche die Sichten Beschaffung, Betrieb, Services und Entsorgung berücksichtigt [22]. Green-IT-Adoption wird darin als Intensität und Breite der Auseinandersetzung mit Green IT in diesen vier Sichten definiert. Da diese alle auch im Kontext des Green ITSM eine hohe Relevanz besitzen, werden die Konzepte „Green IT“ und „Green-IT-Adoption“ im Folgenden sinngemäß zu [22] verwendet. Das GITAM unterscheidet ferner sowohl spezifische Beweggründe (ökonomisch, ethisch, regulatorisch), den organisatorischen, technologischen und externen Kontext sowie die Bereitschaft (i.S.v. „readyness“) einer Organisation als Einflüsse auf die Green-IT-Adoption. An dieser Stelle gibt es zum ITSM und der Forschung zum Standardrahmenwerk ITIL erste Ansatzpunkte. So zeigt eine empirische Studie signifikante Zusammenhänge zwischen dem Reifegrad einer ITIL-Implementierung und der Generierung von Vorteilen durch das Rahmenwerk auf [23], die dann auch die nachhaltige Ausrichtung des ITSM begünstigen sollten. Zur Bestimmung des Reifegrads diente dazu das „Maturity Model“, das auf dem Reifegradmodell des „CobiT and Capability Maturity Model Integration“-Frameworks basiert und eine bewährte (u.a. verwendet in [23-25]) Methode darstellt, um Fortschritt und Reife einer ITIL-Implementierung festzustellen [23]. In Anlehnung an [23] können dazu insgesamt 6 Stufen von *nicht existent(0)* bis *optimiert(5)* unterschieden werden, die Tabelle 1 darstellt.

Tabelle 1. Stufen eines ITIL-Reifegradmodells

Stufe und Name	Eigenschaften
0 – nicht existent	ITIL ist nicht implementiert
1 – anfänglich	ITSM-Prozesse sind ad-hoc und desorganisiert
2 – wiederholbar	ITSM-Prozesse sind intuitiv, bieten wiederholbare Ergebnisse
3 – definiert	ITSM-Prozesse als dokumentierte Standardverfahren mit klaren Ergebnissen
4 – kontrolliert	ITSM-Prozesse mit zusätzlicher Messung von kritischen Erfolgsfaktoren („Key Performance Indicators“)
5 – optimiert	Die kontinuierliche Verbesserung auf Basis von Kennzahlen wird zusätzlich angestrebt

In [23] wurde dann festgestellt, dass mit zunehmendem Reifegrad u.a. Schwierigkeiten bei der Implementierung von neuen Prozessen abnahmen, was die Autoren auf Lerneffekte zurückführen. Ebenso wuchs die Unternehmensflexibilität zusammen mit

dem Reifegrad, da der Widerstand gegenüber organisatorischen Veränderungen abnahm und das IT-Business-Alignment verbessert werden konnte. Zusammengefasst sollten sich alle dieser Vorteile auch auf die organisatorische Bereitschaft auswirken und damit im Sinne des GITAM die Green-IT-Adoption begünstigen. Somit kann folgende Vermutung angestellt werden:

V1: Mit zunehmendem ITIL-Reifegrad steigt die Green-IT-Adoption.

Weiterhin ist davon auszugehen, dass im fortschreitenden Prozess der ITIL-Implementierung, zusammen mit höherer Reife, auch solche Kompetenzen erlernt werden, die sich auf Nachhaltigkeitsmaßnahmen übertragen lassen. Befindet sich ein Unternehmen beispielsweise in Reifegrad 4 oder 5, dann sollte es für dieses Unternehmen – aufgrund des bereits implementierten Kennzahlen-Controllings – deutlich einfacher sein, ebenso die Nachhaltigkeit von ITSM-Prozessen zu überwachen, als wenn es sich in einem niedrigeren Reifegrad befände, in dem noch keine quantitativen Kontrollinstrumente implementiert wurden. Dies ergibt eine zweite Vermutung:

V2: Unternehmen bewerten die Implementierung von Green-IT-Maßnahmen mit zunehmendem ITIL-Reifegrad als weniger schwierig.

In einer allgemeineren Betrachtung, die von ITSM-Standards losgelöst ist, kann zudem vermutet werden, dass spezifische Faktoren im Zusammenhang einer erfolgreichen Realisierung eines Green ITSM existieren. Ein naheliegender Erfolgsfaktor könnte beispielsweise in der Unterstützung des (oberen) Managements für Nachhaltigkeitsinitiativen liegen, die dann besonders positiv auf deren Umsetzung wirkt.

V3: Im Kontext der Umsetzung des Green ITSM existieren Erfolgsfaktoren.

3.3 Anforderungen an neue Konzepte des Green ITSM

Wie in Kapitel 2.2 erwähnt wurden im Rahmen des Green ITSM bereits neue Konzepte zur Erweiterung des Rahmenwerks ITIL entwickelt. Ob allerdings in der Unternehmenspraxis tatsächlich ein Interesse an diesen besteht, ist zu diesem Zeitpunkt noch unbekannt und soll daher für verschiedene dieser Konzepte überprüft werden.

V4: Es besteht unter ITSM-Organisationen das Interesse an Erweiterungskonzepten für Rahmenwerke, zur Erreichung einer ökologisch nachhaltigen Ausrichtung.

Ferner soll das neue Konzept der Green Incidents [21], wodurch eine potenzielle Gefährdung der nachhaltigen Ausführung von ITSM-Prozessen angezeigt werden soll, um diese dann in einem anschließenden Prozess beheben zu können, genauer betrachtet werden. So lassen sich argumentativ-deduktiv deutliche Unterschiede hinsichtlich der Dimensionen Erkennbarkeit, Priorität sowie Komplexität (der Behebung) ableiten, die zwischen Incidents und Green Incidents zu bestehen scheinen (siehe Tabelle 2). Diese inhaltlichen Unterschiede charakterisieren einen Green Incident als eigenständiges Phänomen, das nur bedingt dem bisherigen Incident-Konzept gleicht. Somit ergibt sich folgende Vermutung:

V5a: Green Incidents stellen bezüglich der Dimensionen Erkennbarkeit, Priorität und Komplexität eine eigene Incident-Klasse dar.

Tabelle 2. Vermutete Unterschiede zwischen Incidents und Green Incidents

	Priorität	Erkennbarkeit	Komplexität
Incident	Hoch, Kunde kann betroffenen Service evaluieren; ist sich ggf. der Störung bewusst	Der Kunde kann Einbußen in Erreichbarkeit, Performance, etc. feststellen.	In Ausnahmefällen hoch, bei regulären Technikproblemen allerdings gering und wiederkehrend. Eher operatives Vorgehen.
Green Incident	Mittelmäßig, da Green Incidents vom Kunden kaum zu erkennen sind. Green Incidents sind daher unkritischer.	Erkennbarkeit beim Kunden nur bedingt gegeben und abhängig von der Transparenz des Servicebetriebs. Bei überlagerten und kombinierten Services ist die kundenseitige Erkennbarkeit vermutlich nicht möglich.	Hoch, da weniger Erfahrungen auf dem Gebiet existieren. Es muss strategischer und experimenteller vorgegangen werden.

Eine weitere Komponente, die sich aus den vorigen Überlegungen ableitet, ist prozessualer Natur und bezieht sich auf die Behandlung der beiden Incident-Typen. Denn aufgrund der tabellarisch dargestellten Diskrepanzen ist ersichtlich, dass Green Incidents auch anders behandelt werden sollten. Durch die erschwerte Erkennung eines Green Incidents durch den Servicenutzer entfällt beispielsweise ein Hauptauslöser des konventionellen Incident-Management-Prozesses – die Prioritäten müssen hier angepasst werden. Im Gegenzug ist die Güte einer Lösung durch den Kunden nicht zu evaluieren, was im herkömmlichen Incident Management der Regelfall ist und die Kundenzufriedenheit sichert. Im Sinne der Prioritäten und Ziele beim Behandlungsprozess der Green Incidents muss berücksichtigt werden, dass nicht mehr die bloße Schnelligkeit einer Lösung am wichtigsten ist: Vielmehr muss der Incident auf nachhaltige Weise behoben werden. Durch diesen strategischen Charakter ergibt sich auch ein größerer Managementaufwand, da tragende Entscheidungen zuerst diskutiert werden müssen, wenn kein zuvor bereits beschlossenes „Lösungsrezept“ besteht. Infolgedessen kann folgende Vermutung formuliert werden:

V5b: Green Incidents erfordern aus Sicht der Praxis eine spezifische Identifikation und Behandlung.

Hinsichtlich dieses Prozesses ist dann davon auszugehen, dass dort verschiedene Prozesselemente von besonderer Bedeutung sind, z. B. die automatische Erkennung von Green Incidents durch Überschreitung von Toleranzbereichen oder die Verwendung von speziellen Metriken zur Ressourceneffizienz und Nachhaltigkeit.

V5c: Im Green Incident Management gibt es charakteristische Prozesselemente und -eigenschaften, die besonders wichtig sind.

4 Prüfung der Vermutungen auf Plausibilität

4.1 Forschungsmethodik

Zur Überprüfung der Vermutungen wurde eine einmalige standardisierte Befragung ausgewählt und durch die Software LimeSurvey umgesetzt. Die Befragung wurde auf ITSM-Unternehmen in Deutschland eingegrenzt. Im Rahmen der Kapazitäten dieser Voruntersuchung wurde nur eine Teilerhebung angestrebt. Kern der Stichprobe bildete dabei eine Liste der 25 „führenden IT-Service Unternehmen in Deutschland 2011“, herausgegeben von Lünendonk [26]. Aufnahmekriterium für diese ist das Erzielen von mindestens 50% des Umsatzes durch IT-Dienstleistungen. Die Stichprobe wurde anschließend um 75 Unternehmen, die in der Mitgliederliste des deutschen IT-Service-Management-Forums verzeichnet sind, erweitert. Diese mussten u.a. einen klar erkennbaren Bezug zum ITSM und zu IT-Services vorweisen. Schließlich ergab sich eine Stichprobe von 90 zu kontaktierenden Unternehmen.

4.2 Operationalisierung der Konstrukte

Für Vermutungen V1/V2 sind die Konstrukte ITIL-Reifegrad, Green-IT-Adoption sowie Umsetzungsschwere der Nachhaltigkeitsmaßnahmen zu messen. In Anlehnung an [23] wurden sechs verschiedene Aussagen formuliert, die je einen Reifegrad repräsentieren. Durch Auswahl des zutreffendsten Szenarios durch die Befragten sollte sich ein Indikator für den tatsächlichen Reifegrad ergeben. Zur Messung der Green-IT-Adoption wurde ein Index entwickelt. Abgeleitet aus einem Maßnahmenkatalog setzt sich dieser aus 13 grünen Maßnahmen der Bereiche Infrastruktur, Unternehmenspolitik und Mitarbeiterpolitik zusammen. Diese wurden auf Basis deskriptiver Studien zur Green-IT-Adoption zusammengestellt [27-28]. Zusätzlich wurden freie Nennungen von weiteren Maßnahmen ermöglicht. Bzgl. jeder Maßnahme wurde der Status in den vier Stufen „keine Umsetzung vorgesehen“, „Umsetzung vorgesehen“, „Umsetzung begonnen“ und „Umsetzung abgeschlossen“ erhoben. Wurde die begonnene Implementierung einer Maßnahme angegeben, dann wurde für die korrespondierende Dimension die empfundene Schwierigkeit der Umsetzung durch eine 4er-Likert-Skala erfragt. Im Anschluss wurde der Adoptionsindex berechnet, als Summe über alle Teiladoptionen bezüglich der 13 Maßnahmen (und ggf. freien Angaben). Im Hinblick auf V3 wurden Bewertungen der Relevanz zu einer Auswahl an potenziell kritischen Erfolgsfaktoren auf einer 4er-Likert-Skala erhoben. Diese Faktoren wurden auf Basis eigener Überlegungen und in Anlehnung an die in [29] für Anwendungssoftware identifizierten Erfolgsfaktoren zusammengestellt. Bezüglich V4 wurde das Konzept der Green Service Level Agreements (Green SLA) [21] kurz vorgestellt und anschließend auf einer 4er-Likert-Skala erfragt, als wie sinnvoll dieses erachtet wird. Auf gleiche Weise wurde die allgemeine Sinnhaftigkeit der Erweiterung der ITIL erfragt. Anschließend wurde das Konzept der Green Incidents vorgestellt und auf einer 4er-Likert-Skala erhoben, wie sehr das befragte Unternehmen von einem Green Incident-Management-Prozess profitieren würde. Im Hinblick auf V5a wurde die Zustimmung zu jeweils einer Aussage in den Dimensionen Erkennbarkeit, Komplexität und Priori-

tät für je Incidents und Green Incidents auf einer 4er-Likert-Skala gemessen. Annahme dieser Operationalisierung ist, dass, sollte die identische Erkennbarkeit/ Komplexität/ Priorität zwischen Incidents und Green Incidents gegeben sein, auch gleiche Zustimmungen zu den jeweiligen Aussagen gemessen werden müssten. Bzgl. V5b wurde anschließend erfragt, ob die Behandlung von Green Incidents in einem angepassten Incident-Management-Prozess, in einem separaten Prozess oder auf ganz andere Weise (freies Textfeld „Sonstiges“) erfolgen sollte. Für V5c wurde eine Liste mit 9 verschiedenen Prozesselementen aufgestellt, die auf Basis einer 4er-Likert-Skala bewertet werden sollten.

4.3 Durchführung

Die Durchführung fand vom 25.07.2012 bis 24.08.2012 statt. In diesem Zeitraum wurden die 90 ITSM-Unternehmen angeschrieben und zur Teilnahme an der Online-Umfrage eingeladen. Sofern bzgl. eines Unternehmens nach jeweils zwei oder vier Wochen noch keine Antwort vorlag und die vorangegangene(n) E-Mail(s) zustellbar war(en), erfolgte zu diesen Zeitpunkten jeweils eine Erinnerung an die Umfrage.

4.4 Rücklauf und erhaltene Stichprobe

Insgesamt 79 Unternehmen konnten erfolgreich zur Umfrage eingeladen werden. Die Zustellung konnte in 11 Fällen aus technischen Gründen nicht erfolgen. Ein Unternehmen teilte mit, prinzipiell nicht an Umfragen teilzunehmen, wodurch 78 Unternehmen als endgültige Auswahl verblieben. Diesbezüglich lagen zum 27.08.2012 vollständige Antworten von 15 Unternehmen vor, was einer Rücklaufquote von 19,2% entspricht. Dieses Ergebnis ist im Vergleich mit anderen Studien als gut einzustufen. So wurde, in zuvor bereits angesprochener Studie über ITIL, lediglich eine Rücklaufquote von 8% erzielt [23]. Weiterhin ist anzumerken, dass unter den Antworten 6 der nach Lünendonk 25 führenden ITSM-Unternehmen vertreten sind [26]. Im Hintergrund von 8 automatischen Abwesenheitsnachrichten auf das erste Anschreiben und weiteren 7 im Rahmen der Erinnerungen ist ferner zu vermuten, dass die Umfrage in einem anderen Zeitraum noch effektiver gewesen wäre. Dennoch ist auf die verhältnismäßig geringe Anzahl der 15 Antworten – und der damit zu hinterfragenden Aussagekräftigkeit über die Stichprobe hinaus – kritisch hinzuweisen. Als Basis dieser ersten Voruntersuchung ist die Stichprobe geeignet, solange sie nur zum Zweck der Plausibilitätsprüfung verwendet und mit geeigneten Methoden analysiert wird.

Tabelle 3 zeigt das Profil der befragten Unternehmen. Für alle Antworten steht fest, dass sie von einem Bereichsleiter bzw. -manager oder sogar der Geschäftsleitung stammen. Dies sollte sich positiv auf die Validität der Antworten auswirken, sofern die Annahme zutrifft, dass Mitarbeiter in hohen Positionen gut informiert sind und auf eine korrekte Repräsentation ihres Unternehmens Wert legen. Ferner ist die Stichprobe hinsichtlich der Unternehmensgrößen (Mitarbeiter und Umsatz) genügend ausgeglichen. Die Bedeutung des ITSM bei den Unternehmen wurde – mit 2 Ausnahmen – als „eher groß“ bis „groß“ angegeben. Die Stellung von ITIL als de-facto Standard wurde bestätigt, da alle befragten Unternehmen entweder ITIL V2/V3 einsetzen.

Tabelle 3. Profil der 15 befragten Unternehmen

Mitarbeiter	Häufigkeit	Umsatz in Mill. €	Häufigkeit
Unter 250	7	Unter 1	0
250-499	2	1-50	6
500-1499	2	50-499	8
Ab 1500	4	Über 499	1
Befragte Person	Häufigkeit	Geschäftsbereich*	Häufigkeit
Bereichsleiter, Bereichsmanager	11	(IT-)Beratung	Ja: 14 / Nein: 1
		Systemintegration	Ja: 12 / Nein: 3
		Bereitstellung v. IT-Infrastruktur	Ja: 11 / Nein: 4
		Application Hosting	Ja: 10 / Nein: 5
Vorstand, Geschäftsführer	4	Individualsoftwareentwicklung	Ja: 7 / Nein: 8
		Webhosting	Ja: 5 / Nein: 10
		Managed Services	Ja: 1 / Nein: 14
ITSM-Rahmenwerk*	Häufigkeit	Bedeutung des ITSM	Häufigkeit
ITIL V3	Ja: 12 / Nein: 3	Keine Bedeutung	0
ITIL V2	Ja: 5 / Nein: 10	Schwache Bedeutung	0
COBIT	Ja: 2 / Nein: 13	Eher schwache Bedeutung	2
MOF 4.0	Ja: 1 / Nein: 14	Eher große Bedeutung	6
BlueEDGE Methode	Ja: 1 / Nein: 14	Große Bedeutung	7

* Merkmal, bei dem die befragten Personen mehrere Angaben machen konnten

4.5 Prüfung der Vermutungen auf Plausibilität

Bzgl. Vermutung V1 sind die Ausprägungen des ITIL-Reifegrads und des gebildeten Adoptionsindex' zu untersuchen. Zunächst wurden dazu zwei Unternehmen, die lediglich Beratung und Systemintegration als Geschäftsaktivitäten angegeben haben, von der Analyse ausgeschlossen. Dies wurde getan, um eine gleiche Bezugsbasis bzgl. des Adoptionsindex' zu erhalten. So kann davon ausgegangen werden, dass Unternehmen, die keine durch Informationstechnik erstellten Services erbringen, einige der infrastrukturellen Maßnahmen gar nicht adoptieren können, da diese schlicht nicht praktikabel sind. Für die verbleibenden 13 Datensätze wurde zunächst ein Streudiagramm der beiden Variablen erstellt, das folgende Abbildung darstellt.

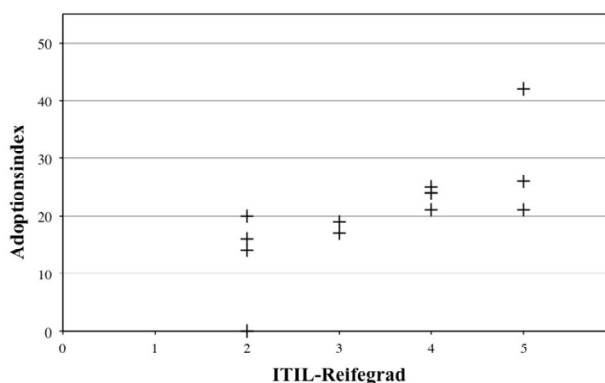


Abb. 1. Streudiagramm für ITIL-Reifegrad und Adoptionsindex

Bereits in Abbildung 1 ist visuell erkennbar, dass in Konformität zu V1 ein positiver Zusammenhang zwischen beiden Variablen vorzuliegen scheint. Um diesen statistisch näher zu untersuchen, wurde das Verfahren der bivariaten Korrelationsanalyse herangezogen. Da die Merkmale Reifegrad und Adoptionsindex beides ordinalskalierte Merkmale sind, wurde der Rangkorrelationskoeffizient nach Spearman gewählt.

Der Rangkorrelationskoeffizient nach Spearman zwischen beiden Variablen zeigt mit 0,864 einen starken positiven Zusammenhang auf. Zudem lieferte ein Test auf statistische Signifikanz einen p-Wert von 0,000. Hinsichtlich V1 lassen sich also deutliche Indizien finden, die es in weiteren Untersuchungen zu bestätigen gilt. Dazu sollten ferner Methoden der Varianzanalyse verwendet werden, um die Adoptionsmittelwerte spezifischer Reifegrade miteinander vergleichen zu können

Weiterhin wurde vermutet, dass die wahrgenommene Umsetzungsschwierigkeit von Nachhaltigkeitsinitiativen mit zunehmendem ITIL-Reifegrad abnimmt (V2). Dazu wurden Rangkorrelationsmaße zwischen den Variablen ITIL-Reifegrad und den jeweiligen Bewertungen für die Umsetzungsschwere in den Dimensionen infrastrukturelle, unternehmenspolitische und mitarbeiterbezogene Maßnahmen gebildet. Diese waren alle mit Werten unter 0,35 schwach und darüber hinaus mit p-Werten von mindestens 0,29 nicht signifikant. Der vermutete Zusammenhang kann demnach auf diese Weise nicht nachgewiesen werden. Hier ist zu hinterfragen, ob die Umsetzungsschwere bzgl. der Nachhaltigkeitsmaßnahmen noch spezifischer erhoben werden muss, um so möglicherweise bestehende Einflüsse des ITIL-Reifegrades in lediglich einzelnen Teilbereichen (z. B. Kennzahlen-Controlling) finden zu können.

Hinsichtlich der dritten Vermutung wurde eine Auswahl an potenziell kritischen Erfolgsfaktoren zusammengestellt, die dann durch die Befragten auf ihre Relevanz im Kontext der Umsetzung des nachhaltigen ITSM bewertet wurden. Folgende Tabelle stellt die sortierten Mittelwerte der Antworten zusammen dar, wobei der Wert 3 für „Sehr relevant“ und der Wert 0 für „Irrelevant“ steht.

Tabelle 4. Bewertung potenziell kritischer Erfolgsfaktoren des Green ITSM

Erfolgsfaktor	Mittlere Bewertung (σ)
Support durch das obere Management	2,87 (0,352)
Klare Zielsetzungen	2,73 (0,458)
Controlling der Projektziele	2,57 (0,488)
Kompetenzen der Projektverantwortlichen	2,47 (0,516)
Projektmanagement	2,33 (0,617)
Orientierung an Best Practices	2,2 (0,561)
Integration v. Nachhaltigkeit in die Unternehmensstrategie	2,2 (0,676)
Kommunikation zwischen Abteilungen	2,13 (0,640)
Einrichtung eines Lenkungsausschusses	1,93 (0,961)
Einbezug von Beratern	1,4 (0,828)

In Tabelle 4 ist zu sehen, dass insbesondere die Faktoren „Support durch das obere Management“, „Klare Zielsetzungen“ sowie „Controlling der Projektziele“ aus Sicht der Praxis für ein erfolgreiches Green ITSM als sehr relevant eingeschätzt werden. Vor allem im Vergleich mit dem als eher irrelevant eingestuftem Faktor „Einbezug

von Beratern“ können diese als wichtig bezeichnet werden. Ob darüber hinaus allerdings auch tatsächlich ein statistisch signifikanter Zusammenhang zu einem erfolgreichen Green ITSM besteht, lässt sich auf diese Weise weder belegen noch abschätzen.

Hinsichtlich V4 wurde erhoben, als wie sinnvoll die Einführung von Green SLAs und die Erweiterung von ITIL um Nachhaltigkeitsaspekte bewertet wird. Zusätzlich wurde gefragt, wie sehr die Unternehmen von einem Prozess zur Behandlung von Green Incidents profitieren würden. Wie Tabelle 5 verdeutlicht, zeigen die Daten ein deutliches Interesse an neuen Konzepten des Green ITSM und der allgemeinen Erweiterung von ITIL. Vermutung 4 kann somit vorläufig bestätigt werden.

Tabelle 5. Interesse der Befragten an Erweiterungskonzepten im Green ITSM

allg. ITIL-Erweiterung	Häufigkeit	Green SLAs	Häufigkeit
Nicht sinnvoll	1 (6,7%)	Nicht sinnvoll	0 (0%)
Eher nicht sinnvoll	1 (6,7%)	Eher nicht sinnvoll	3 (20%)
Eher sinnvoll	11 (73,3%)	Eher sinnvoll	11 (73,3%)
Sinnvoll	2 (13,3%)	Sinnvoll	1 (6,7 %)
Profitieren von Green Incident Management			Häufigkeit
Gar nicht			0 (0%)
Eher wenig			5(33,3%)
Eher viel			10 (66,7%)
Viel			0 (0%)

Um Vermutung 5a zu überprüfen, wurden die Mittelwerte der Zustimmungen (4er-Likert-Skala) zu korrespondierenden Aussagen für Incidents/Green Incidents gebildet. Bereits diesbezüglich ist zu erkennen, dass Incidents als vom Kunden einfacher zu erkennen, durch den Anbieter besser zu lösen und mit allgemein höherer Priorität eingestuft werden als Green Incidents in denselben Dimensionen (siehe Tabelle 6).

Ein t-Test auf jeweils signifikante Unterschiede der 3 Differenzenmittelwerte von Null bestätigte dies, wie die p-Werte in Tabelle 6 zeigen. Die dazu vorausgesetzte Normalverteilung der Differenzen zeigte ein Kolmogorov-Smirnoff Test (H_0 : Normalverteilung) mit p-Werten von (0,367), (0,302) und (0,203) an.

Tabelle 6. Einordnung von Incidents und Green Incidents

	Aussage	Mittel/ σ	p-Wert
Erkennbarkeit	<i>Servicebeeinträchtigungen durch Incidents können vom Kunden i.d.R. erkannt werden</i>	2,2 (0,941)	0,005
	<i>Beeinträchtigungen der nachhaltigen Serviceausführung können vom Kunden i.d.R. erkannt werden</i>	1,13 (0,64)	
Komplexität	<i>Für die Behebung von Servicebeeinträchtigungen durch Incidents gibt es bekannte Maßnahmen</i>	2,33 (0,724)	0,00
	<i>Für die Behebung von Beeinträchtigungen der nachhaltigen Serviceausführung gibt es bekannte Maßnahmen</i>	1,2 (0,676)	
Priorität	<i>Die Behebung von Servicebeeinträchtigungen durch Incidents hat allgemein hohe Priorität.</i>	2,8 (0,414)	0,00
	<i>Die Behebung von Beeinträchtigungen der nachhaltigen Serviceausführung hat allgemein hohe Priorität.</i>	1,53 (0,915)	

Vermutung 5b folgt aus der (zuvor vorläufig bestätigten) Unterschiedlichkeit von Incidents und Green Incidents. Um sie zu überprüfen, wurde gefragt, ob Green Incidents in einem eigenständigen oder angepassten Prozess behandelt werden sollten. In 7 von 15 Fällen wurde die Anpassung des bestehenden Prozesses bevorzugt, 8 von 15 Unternehmen bevorzugten einen neuen Prozess. Somit kann festgehalten werden, dass es keine einstimmige Aussage bzgl. der Behandlung von Green Incidents gibt.

Wie im Falle der kritischen Erfolgsfaktoren von V3 wurde auch hinsichtlich Vermutung V5c eine Aufstellung von potentiell wichtigen Prozesselementen zur Verfügung gestellt, die von den Befragten bewertet werden sollten (siehe Tabelle 7).

Tabelle 7: Bewertungen für Elemente im Green Incident Management

Prozesselement /-eigenschaft	Mittelwert (σ)
Keine Beeinträchtigung der Servicequalität bei der Lösung von Green Incidents	2,67 (0,724)
Effizienzmetriken	2,07 (0,594)
Dokumentation der Green Incidents	1,93 (0,458)
Best Practice basierte Lösungsvorschläge durch Wissensmanagement und Erkennung wiederkehrender Green Incident Typen	1,93 (0,594)
Automatische Erkennung von Green Incidents	1,87 (0,640)
Definierbare Toleranzbereiche zur Auslösung	1,80 (0,862)
Smart-Metering	1,67 (0,724)
Automatische Klassifizierung von Green Incidents	1,53 (0,640)
Priorisierung der Green Incidents	1,33 (0,617)

Wie Tabelle 7 zeigt, ist die Wahrung der Servicequalität, während der Lösung von Green Incidents, deutlich am wichtigsten. Diese Beobachtung ist konsistent mit der in V5a erlangten Erkenntnis, dass herkömmlichen Incidents eine höhere Priorität zugesprochen wird. Ferner wurden die Effizienzmetriken, Dokumentation und Best-Practice basierte Lösungsvorschläge als wichtig eingestuft. Erstaunlich in diesem Kontext ist, dass die Prozesselemente der automatischen Klassifizierung/Priorisierung als eher weniger wichtig bewertet werden, obwohl diese zentrale Bestandteile des herkömmlichen Incident Managements der ITIL sind. Dies unterstreicht V5b, dass die Behandlung von Green Incidents spezifischen Anforderungen genügen muss.

5 Fazit

Im Rahmen dieser Voruntersuchung wurden verschiedene Vermutungen über die konzeptionellen Grundlagen des Green ITSM auf Plausibilität überprüft. Dabei konnten Zusammenhänge zwischen der ITSM-Rahmenwerkverwendung und der Green-IT-Adoption festgestellt werden. So legt die Plausibilitätsprüfung zu Vermutung V1 nahe, dass ein positiver Zusammenhang zwischen dem erreichten ITIL-Reifegrad und der Adoption von Nachhaltigkeitsinitiativen besteht. Dies gilt es in zukünftigen Studien zu bestätigen, die ferner genauer untersuchen sollten, aus welchen Gründen dieser Zusammenhang zustande kommt. In diesem Beitrag wurde als theoretische Grundlage der Vermutung angenommen, dass ein steigender Reifegrad die Green-IT-Adoption begünstigt, da er sich wie in [23] belegt auch positiv auf das IT-Business-Alignment

sowie die Unternehmensflexibilität auswirken kann. Ferner können aber auch weitere moderierende oder medierende Kontextvariablen bestehen, was erforscht werden muss. Nicht bestätigt werden konnte die Vermutung, dass Unternehmen mit größerem Reifegrad die Umsetzung von Nachhaltigkeitsmaßnahmen als leichter bewerten. Dazu sind ggf. subtilere Einflüsse zu untersuchen, da in dieser Studie nur zwischen infrastruktureller, unternehmenspolitischer und mitarbeiterbezogener Dimension der Maßnahmenschwierigkeit unterschieden wurde. Ferner wurde die Relevanz verschiedener Erfolgsfaktoren im Green ITSM erhoben, ob diese allerdings als „kritisch“ einzustufen sind, ist an dieser Stelle noch nicht abzuschätzen.

Die Voruntersuchung hat weiterhin die ersten empirischen Erkenntnisse über vorgeschlagene Erweiterungskonzepte für ITIL, die einem Green ITSM dienen sollen, generiert. Es konnte dazu zunächst gezeigt werden, dass deutliches Interesse an diesen neuen Konzepten seitens der befragten Unternehmen besteht. Daraufhin wurden die in V5a vermuteten Unterschiede zwischen Incidents und Green Incidents in Priorität, Komplexität und Erkennbarkeit bestätigt. Dass die Behandlung von Green Incidents darüber hinaus auch in einem eigenen Prozess erfolgen sollte, konnte jedoch nicht bestätigt werden. Die erhobenen Einschätzungen zu wichtigen Prozesselementen eines Green Incident Managements zeigen, dass es dort große Unterschiede gibt (z. B. zwischen „Keine Beeinträchtigung der Servicequalität“ und „Priorisierung“).

Diese Ergebnisse sollten als Ansatzpunkte weiterer Forschungsvorhaben aufgegriffen werden. Dabei ist vor allem die Stichprobe zu vergrößern, um die Repräsentativität zu steigern und die Anwendung weiterer Analysemethoden (z. B. Varianzanalysen mit Paarvergleichen) zu ermöglichen. Auf diese Weise lassen sich für IT-Service-Unternehmen hilfreiche Erkenntnisse erlangen, welche dort die Umsetzung des Green ITSM voranbringen können und so die Nachhaltigkeit von IT-Services begünstigen.

Literatur

1. OECD: OECD-Umweltausblick bis 2050: Die Konsequenzen des Nichthandelns (2012), <http://www.oecd.org/dataoecd/53/36/49889636.pdf>
2. United Nations: Our Common Future (1987)
3. Koomey, J.G.: Growth In Data Center Electricity Use 2005 to 2010. Stanford University (2007)
4. Murugesan, S.: Harnessing Green IT: Principles and Practices. In: IT Professional 10 (1), 24-33 (2008)
5. Ortwerth, K., Teuteberg, F.: Green IT/IS Forschung – Ein systematischer Literaturreview und Elemente einer Forschungsagenda. In: Proceedings der Multikonferenz Wirtschaftsinformatik (MKWI) 2012, pp. 1501-1513, Braunschweig (2012)
6. Watson, R.T., Boudreau, M.-C., Chen, A., et al.: Information Systems and Environmentally Sustainable Development. MIS Quarterly 34 (1), 23-38 (2010)
7. Gadatsch, A.: Green IT. In: WISU 2/10, 228-237 (2010)
8. Buhl, H.U., Laartz, J., Löffler, M., Röglinger, M.: Green IT reicht nicht aus! In: Wirtschaftsinformatik & Management 1, 54-58 (2009)
9. Facheux, S., Nicolai, I.: IT for green and green IT: A proposed typology of eco-innovation. Ecological Economics 70, 2020-2027 (2011)

10. Ghose, A., Hoesch-Klohe, K., Hinsche, L.: Green Business Process Management: A research agenda. *Australasian Journal of Information Systems* 16 (2), 103-117 (2009)
11. Galup, S.D., Dattero, R., Quan, J.J., Conger, S.: An overview of IT service management. *Communications of the ACM* 52, 124-127 (2009)
12. Zarnekow, R., Brenner, W., Pilgram, U.: *Integrated Information Management: Applying Successful Industrial Concepts in IT*. Springer, Berlin (2006)
13. Gartner: Gartner Says Worldwide Software-as-a-Service Revenue to Reach \$14.5 Billion in 2012, <http://www.gartner.com/it/page.jsp?id=1963815>
14. Buchsein, R., Victor, F., Günther, H., et al.: *IT-Management mit ITIL V3*. 2. Auflage, Vieweg +Teubner, Wiesbaden (2008)
15. IT Governance Institute: *Global Status Report on the Governance of Enterprise IT* (2011), <http://www.isaca.org/Knowledge-Center/Research/Documents/Global-Status-Report-GEIT-10Jan2011-Research.pdf>
16. Harmon, R.R., Auseklis, N.: Sustainable IT Services: Assessing the Impact of Green Computing Practices. In: *Proceedings of the Portland International Conference on Management of Engineering & Technology*, pp. 1707-1717 (2009)
17. Houy, C., Reiter, M., Fettke, P., Loos, P.: Towards Green BPM – Sustainability and Resource Efficiency through Business Process Management. In: Muehlen, M., Su, J. (eds.): *Business Process Management Workshops*. LNBP, Vol. 66, pp. 501-510. Springer (2011)
18. Zarnekow, R., Ere, K.: Nachhaltiges IT-Service-Management – Grundlagen, Vorgehensmodell und Managementinstrumente. *HMD - Praxis der Wirtschaftsinformatik* 45 (264), 7-18 (2008)
19. Cater-Steel, A., Tan, W.-G.: The Role Of IT Service Management in Green IT. *Australasian Journal of Information Systems* 17 (1), 107-125 (2010)
20. Dubey, S., Hefley, W.E.: Greening ITIL: Expanding the ITIL Lifecycle for Green IT. In: *Proceedings of PICMET 2011*, pp. 1-8 (2011)
21. Reiter, M., Fettke, P., Loos, P.: Eine systematische Analyse des Einflusses ökologischer Ziele auf das IT-Service-Management. In: *Proceedings der Multikonferenz Wirtschaftsinformatik (MKWI) 2012*, pp. 1545-1557, Braunschweig (2012)
22. Molla, A.: GITAM: A Model for the Adoption of Green IT. In: *Proceedings of the 19th Australasian Conference on Information Systems*, pp. 658-668, Christchurch (2008)
23. Marrone, M., Kolbe, L.M.: Einfluss von IT-Service-Management-Frameworks auf die IT-Organisation. In: *Wirtschaftsinformatik* 53 (1), 5-19 (2011)
24. Cater-Steel, A., Tan, W.G., Toleman, M.: *itSMF Australia 2007 conference: summary of ITSM standards and frameworks survey responses* (2007)
25. Kießling, M., Marrone, M., Kolbe, L.M.: Influence of IT Service Management on Innovation Management: First Insights from Exploratory Studies. In: D'Atri, A. et al.: *Management of the Interconnected World*. Springer, Berlin (2010)
26. Lünendonk GmbH: *Führende IT-Service-Unternehmen in Deutschland 2010* (alphabetisch), http://lunendonk-shop.de/out/pictures/0/lue_its_2011_f300511_fl.pdf
27. Molla, A., Pittayachawan, S., Corbitt, B.: Green IT Diffusion: An International Comparison. *Green IT Working Paper Series* 1, 1-15 (2009)
28. Schmidt, N.-H., Ere, K., Kolbe, L.M., Zarnekow, R.: Examining the contribution of Green IT to the objectives of IT departments: Empirical Evidence from German Enterprises. *Australasian Journal of Information Systems* 17 (1), 127-140 (2010)
29. Somers, T.M., Nelson, K.: The Impact of Critical Success Factors across the Stages of Enterprise Resource Planning Implementations. In: *Proceedings of the 34th Hawaii International Conference on System Sciences*, pp. 1-10 (2001)

Auswirkungen von In-Memory-Datenmanagement auf Geschäftsprozesse im Business Intelligence

Patrick Wessel¹, Sebastian Köffer², und Jörg Becker²

¹ arvato Systems GmbH, NMA-EB Business Intelligence, Gütersloh, Germany
Patrick.Wessel@bertelsmann.de

² Universität Münster, European Research Center for Information Systems, Münster, Germany
{sebastian.koeffler, joerg.becker}@ercis.uni-muenster.de

Abstract. Für viele Autoren stellt In-Memory-Datenmanagement (IMDM) eine technologische Revolution dar, welche die Art und Weise, wie betriebliche Informationssysteme gestaltet werden, nachhaltig verändern wird. Dieser Artikel trägt zu einem tieferen Verständnis der potenziellen Auswirkungen von IMDM auf betriebliche Geschäftsprozesse bei. Innerhalb einer Literaturanalyse wurden eine Reihe von Publikationen identifiziert, die die Auswirkungen von IMDM auf Geschäftsprozesse thematisieren. Drei zentrale Diskussionsaspekte liegen diesen Prozessänderungen zugrunde: Die bessere Abfragegeschwindigkeit mit der Möglichkeit zur Echtzeit-Analyse, das Zusammenspiel zwischen Big Data und IMDM sowie die mögliche Vereinigung von OLAP und OLTP Systemen. Zudem werden konkrete Beispiele für Prozessveränderungen durch IMDM aus der Literatur genannt. Die Erkenntnisse der Literaturanalyse konnten durch eine Interviewbefragung ergänzt werden. Der Artikel reichert die Prozessperspektive auf IMDM zusätzlich an und nennt weitere Forschungsmöglichkeiten in der Schnittstelle von IMDM und Business Process Management.

Keywords: In-Memory, Business Intelligence, Business Process Management, Real-Time Computing, Big Data

1 Einleitung

“In-memory technology truly marks an inflection point for enterprise applications” – Für Plattner und Zeier (2011) stellt In-Memory-Datenmanagement (IMDM) den Beginn einer technologischen Revolution dar [26]. Zahlreiche Autoren teilen die Meinung, dass sich die Art und Weise, wie betriebliche Anwendungen gestaltet und Unternehmen gesteuert werden, in Zukunft nachhaltig verändern wird [3], [23]. Das Konzept des IMDM sieht vor den gesamten Datenbestand eines Unternehmens primär im Hauptspeicher vorzuhalten. Durch diese technische Innovation sind enorme Leistungszuwächse möglich, um Anforderungen zu erfüllen, die in der Vergangenheit durch einen vergleichsweise langsamen Festplatten-Zugriff verhindert wurden [21].

Bill Gates stellte im Jahr 1994 mit dem Konzept „information at the fingertips“ seine Vision einer Welt vor, in der beliebige Informationen global verfügbar sind

[10]. Die Entwickler der Internetsuchmaschinen sind dieser Vision gefolgt und haben das Potential, große Datenmengen in Echtzeit zu analysieren, bereits unter Beweis gestellt. Der Nutzer erhält schon beim Eingeben einer Suchanfrage sofort relevante Ergebnisse. Mittlerweile erwarten Anwender betrieblicher Informationssysteme ebenfalls Analyseergebnisse innerhalb von einer Sekunde [26]. Dazu operieren Unternehmen heutzutage in immer komplexeren Umgebungen. Es ergibt sich die Notwendigkeit schnell und flexibel reagieren zu können. Eine Anforderung, die allerdings häufig technisch nicht erfüllt werden kann [4], [26]. Derzeit werden viele Entscheidungen auf Basis historischer, vordefinierter Datenstrukturen getroffen. IMDM, eine Kombination aus Software- und Hardware-Innovationen, bietet in diesem Zusammenhang völlig neue Potentiale, die die Vision von Bill Gates in die Realität umsetzen könnten [26].

Durch IMDM entstehen für die Wissenschaft neue Forschungsfelder und für die Praxis ein akuter Bedarf nach Einschätzungen zu ökonomischen Potentialen, Voraussetzungen und Anwendungsgebieten. Das Ziel dieses Beitrags ist, die Auswirkungen von IMDM auf Geschäftsprozesse im Business Intelligence (BI) Umfeld zu untersuchen. Folgende Forschungsfragen sollen untersucht werden:

- RQ1: Welche Literatur zum Thema IMDM wurde bereits im Kontext betriebswirtschaftlicher Prozessänderungen publiziert?
- RQ2: Inwiefern werden Geschäftsprozesse im BI-Umfeld durch IMDM beeinflusst oder verändert?

Der Beitrag ist wie folgt aufgebaut. Zunächst werden in Kapitel 2 die Grundlagen von IMDM erläutert. Kapitel 3 beschreibt die Forschungsmethodik. Zur Beantwortung der Forschungsfragen werden in Kapitel 4 die Ergebnisse zu RQ1 anhand dreier Diskussionsaspekte aufbereitet. Kapitel 5 präsentiert Beispiele zur Veränderung von Geschäftsprozessen durch IMDM aus der Literatur sowie eine erste empirische Validierung der Konzepte anhand einer Interviewbefragung. Die Arbeit schließt mit einer Diskussion über Implikationen und potenzielle Forschungsfelder.

2 In-Memory-Datenmanagement

Die Idee von IMDM ist nicht neu, sondern entstand schon in den 80er Jahren [20]. Zu dieser Zeit verhinderte allerdings die Unzuverlässigkeit des Hauptspeichers und dessen hohe Kosten eine nachhaltige Etablierung [26]. Probleme entstanden durch die Flüchtigkeit des Hauptspeichers, durch Hardwarefehler [18] und durch softwarebedingte Adressierungsfehler [8]. Zusätzlich war es noch nicht möglich, mehrere Prozessoren effektiv zu nutzen und große Speicherbereiche zu adressieren [5]. Durch die Entwicklung von Multicore-Architekturen, der 64-Bit-Technologie und zuverlässigen, großen Hauptspeichern konnten viele Hindernisse überwunden werden [26]. Mit dem Titel „Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King“ beschrieb Gray (2006) die aktuelle Verschiebung der Speicherhierarchie [11]. Entscheidend war bei dieser Entwicklung das sinkende Preisniveau von Hauptspeicher, das die Idee von IMDM erst ökonomisch legitimierte [26].

In traditionellen Datenbanksystemen hat sich das Konzept einer zeilenorientierten Datenorganisation durchgesetzt, in der die Attributwerte eines Datensatzes nebeneinander angeordnet werden [15]. IMDM-Anwendungen setzen im Wesentlichen spaltenorientierte Datenspeicherung ein, die im OLAP-Umfeld schon seit einiger Zeit erfolgreich eingesetzt werden [25]. Dabei werden die Attributwerte eines Datensatzes untereinander auf benachbarte Blöcke verteilt [26]. In Kombination mit IMDM können spaltenorientierte Datenbanken für OLTP-Systeme ebenfalls interessant werden [15]. Beispielsweise können darin Kompressionsalgorithmen besonders effizient implementiert werden, so dass eine signifikante Reduzierung des Datenvolumens bis zum Faktor zehn möglich ist [32]. Auf diese Weise können auch große Datenmengen gespeichert werden [36]. Ebenso sind hybride Ansätze, die die Stärken beider Welten zu einer hybriden OLTP und OLAP-Datenbank vereinen, denkbar und bereits in der Entwicklung [16].

3 Methodik

Zur Identifikation der relevanten Publikationen im Hinblick auf RQ1 wurde eine strukturierte Literaturanalyse nach Webster und Watson (2002) durchgeführt [38]. Dazu wurde in wissenschaftlichen Publikationsdatenbanken (ISI web of knowledge, Google scholar) nach ausgewählten Suchbegriffen gesucht. Nach dem Ende der Suchläufe wurde eine Rückwärtssuche über ein Level durchgeführt, um verwandte und weiterführende Literatur zu identifizieren. Zusätzlich wurde bei besonders relevanten Autoren mit Hilfe einer Vorwärtssuche weitere relevante Literatur des Autors identifiziert. Aufgrund des relativ jungen Forschungsfeldes wurde der Suchprozess nicht auf Journal-Publikationen begrenzt, sondern auf sämtliche Publikationen ausgeweitet, die durch die Suchmethoden identifiziert werden konnten. Als Suchbegriffe wurden „In-Memory“ oder „Main Memory“, in Verbindung mit „Databases“ oder „Data Management“ verwendet. Durch diese Vorgehensweise konnten insgesamt mehr als 70 Publikationen ermittelt werden. Die identifizierten Publikationen wurden im Anschluss von den Autoren gemäß der Relevanz für die Forschungsfragen geprüft. So wurden zum Beispiel Publikationen mit einer rein technischen und nicht prozessualen Sicht auf In-Memory im Sinne der Forschungsfragen ausgeklammert. Auf diese Weise konnten schließlich im Rahmen der Literaturanalyse 25 Publikationen gefunden werden, die einen konkreten Beitrag zur Beantwortung der Forschungsfragen leisten.

Zur anschaulichen Aufbereitung von RQ1, wurden nach Analyse der Publikationen durch die Autoren gemeinsam drei Diskussionsaspekte zu IMDM gebildet: Performance zur Echtzeit-Analyse, Zusammenspiel von Big Data und In-Memory und Integration von OLAP und OLTP. Diese repräsentieren die wesentlichen Aspekte und Auswirkungen von IMDM, die der Veränderung der betriebswirtschaftlichen Prozesse zugrunde liegen. Um einen guten Überblick über die bisher vorherrschenden Meinungen in der Literatur zu erhalten (RQ1) wurde zusätzlich zu jedem Diskussionsaspekt ein Standpunkt definiert. Dadurch konnte die identifizierte Literatur positioniert werden, insofern das die Autoren einem Standpunkt eher zustimmen, diesen eher ablehnen oder neutral sowohl zustimmende als auch ablehnende Aspekte nennen.

Die Identifikation von Geschäftsprozessveränderungen zur Beantwortung von RQ2 vollzieht sich in zwei Schritten. Zunächst wurden Beispiele aus der Literatur ermittelt, die bereits konkrete Geschäftsprozesse nennen, in denen der Einsatz von IMDM zu positiven Veränderungen führen kann. Anschließend wurden die Ergebnisse in semi-strukturierten Interviews mit vier erfahrenen BI-Beratern, die sich bereits mit IMDM auseinandergesetzt haben, einer ersten Überprüfung unterzogen. Die Leitfragen der Interviews orientierten sich an denen im Rahmen der Literaturanalyse ermittelten Diskussionspunkten. Dabei wurde explizit die Prozessperspektive abgefragt, insbesondere welche konkreten betriebswirtschaftlichen Geschäftsprozesse sich durch IMDM verändern. Alle Interviews wurden transkribiert und gemäß der „Open Coding“ Methodik [34] zeilenweise analysiert, um konzeptuelle Codes herauszuarbeiten, die Prozessänderungen durch IMDM repräsentieren.

4 Auswirkungen von In-Memory Technologie

4.1 Performance zur Echtzeit-Analyse

Bereits im Jahr 2002 verkündete der Gartner: „Alles muss schnell gehen. Wir leben in einer Jetzt-Wirtschaft“ das Realtime Enterprise, indem die Zeit zu einem immer größeren Wettbewerbsfaktor aufsteigt [2]. Echtzeit bedeutet in diesem Kontext zum einen die Verarbeitung von Informationen zum Zeitpunkt ihres Entstehens und zum anderen große Datenmengen auf Abfrage leistungsstark auswerten zu können [39]. Hackathorn (2003) beschreibt diesen Zusammenhang zwischen dem Entstehen einer Information und der Reaktion darauf als Aktionszeit, die sich aus der Datenlatenz, Analyselatenz, Entscheidungslatenz und Reaktionslatenz zusammensetzt [12]. Diese Zeit sollte möglichst minimal sein, um einen geschäftlichen Nutzen zu erzeugen. IMDM kann dazu beitragen, die systemseitige Daten- und Analyselatenz zu verringern [35]. Plattner und Zeier (2011) definieren Echtzeit durch die Verfügbarkeit von Informationen innerhalb der „Geschwindigkeit-der-Gedanken“ [26]. Dieses Konzept beschreibt die Reaktionszeit eines Systems, die durch mentale Prozesse bei Menschen als nicht verzögert wahrgenommen werden. Häufig liegt es an der Technologie, dass die Anforderung an eine Reaktionszeit von unter einer Sekunde nicht erfüllt werden kann [4]. Aus den in diesem Zusammenhang geäußerten Meinungen in der Literatur leiten wir den folgenden zentralen Standpunkt ab:

Standpunkt 1: Mit Hilfe von IMDM kann die Vision von Realtime Business Intelligence endlich umgesetzt werden.

Tabelle 1. Einschätzungen in der Literatur zu Standpunkt 1

Quelle	Zustimmung	Neutral	Ablehnung
[1] Acker et al. 2011	x		
[4] Davenport und Snabe 2011		x	
[6] Fabian und Günther 2011		x	
[9] Gartner 2011	x		
[17] Lechtenböcker/Vossen 2011	x		
[19] Lehner 2011		x	
[26] Plattner und Zeier 2011	x		
[27] Richardson 2011	x		
[31] Schütte 2012		x	
[32] Sinzig und Sharma 2011	x		
[36-37] Thiele et al. 2011	x		
[42] Winter et al. 2011	x		
[45] Zeier et al. 2011	x		

Die zitierten Autoren sind im Wesentlichen davon überzeugt, dass der Leistungszuwachs von IMDM einen signifikanten Einfluss auf bestehende Informationssysteme, Geschäftsprozesse und Arbeitsweisen haben wird. IMDM bietet die Möglichkeit, das Qualitätsniveau von bestehenden Anwendungen anzuheben [1], [6], [17], eine mobile Verfügbarkeit in Echtzeit sicherzustellen und neue Anwendungsklassen zu entwickeln, die auf die Zukunft ausgerichtet sind [26], [45]. Dabei ist allerdings ein „Out-of-the-box“-Denken von allen Beteiligten notwendig [45].

Nach den Analysten von Gartner (2011) werden sich kollaborative Entscheidungsprozesse durch IMDM entwickeln, in denen zunehmend Handlungsempfehlungen eine Rolle spielen werden [9]. Mehrere Autoren erkennen im Zusammenhang mit IMDM eine Entwicklung von einem Push- zu einem Pull-BI, wodurch Informationen flexibel und autonom in Echtzeit abgerufen werden können [1], [4], [37]. Dadurch können Unterbrechungen und Kontextwechsel innerhalb von Prozessen reduziert werden, was zu einer effizienteren Arbeitsweise führen kann [26], [32]. Entscheidend ist, dass innerhalb eines Entscheidungsprozesses mehr Fragen gestellt werden können [27], was in der Konsequenz zu einer interaktiven Entscheidungsfindung [42] nach dem Prinzip „Trial-and-Error“ führt [26].

Lehner (2011) sowie Davenport und Snabe (2011) merken dabei einschränkend an, dass die Relevanz von Informationen und der damit zu beantwortenden Fragestellungen von zentraler Bedeutung ist [4], [19]. Demnach müssen Informationen datenflussorientiert zum richtigen Zeitpunkt verfügbar sein. Die einfache Beschleunigung der Informationsverarbeitung wird ohne eine Anpassung der Prozesse hinsichtlich Datenerhebung und Analyse nicht den erhofften Nutzen erzeugen [4], [6]. In diesem Zusammenhang kann die hinzugewonnene Flexibilität und Reaktionsgeschwindigkeit zu intensiveren Prozessen und dem Verlust der wesentlichen betriebswirtschaftlichen Fragestellung führen [4]. Eine weitere Voraussetzung für den Einsatz von IMDM ist,

dass eine schnelle Reaktionszeit erwartet wird, die aufgrund des Datenvolumens nicht mit herkömmlichen Mitteln realisierbar ist [31]. An dieser Stelle müssen die Nutzer bei sehr datenintensiven, neuen Anwendungen eine angemessene Erwartungshaltung entwickeln [19].

4.2 Zusammenspiel von Big Data und In-Memory

Laut einer Studie der International Data Corporation (IDC) verdoppelt sich die globale Datenmenge alle zwei Jahre und wird von 2011 mit 1800 Exabyte auf 7900 Exabyte im Jahre 2015 anwachsen [7]. Das Datenvolumen in Unternehmen wächst ebenfalls kontinuierlich an, so dass sich die Größe von aktuell produktiven Data Warehouses (DWH) in etwa alle zwei Jahre verdreifacht [43]. In diesem Zusammenhang hat die Praxisliteratur den Begriff „Big Data“ geprägt, der sich nicht nur auf das Datenvolumen (Volume) bezieht, sondern zusätzlich auf die Geschwindigkeit (Velocity) mit der Daten erzeugt werden und auf die Vielfalt (Variety) von unterschiedlichen Datenquellen und Datenstrukturen [28]. Zu dieser Entwicklung tragen eine redundante Datenhaltung, Aufbewahrungsfristen und immer detailliertere Daten innerhalb verschiedener Domänen oder Thematiken bei [26]. Dabei werden unternehmensrelevante Daten in Zukunft zum Großteil aus weniger strukturierten oder unstrukturierten Daten wie Texten, Bildern und Videos bestehen [26]. So konnte Russom (2011) innerhalb der kommenden Jahre eine deutliche Verschiebung des Datenintegrationsbedarfs feststellen [29].

Geschwindigkeits- und Wirtschaftlichkeitsvorteile können durch IMDM erzielt werden, sofern sich die Notwendigkeit von IMDM durch große Datenmengen ergibt [42]. Mehrere Autoren greifen daher das Thema „Big Data“ in Zusammenhang mit IMDM auf. Zusammenfassend lässt sich aus den relevanten Publikationen der folgende Standpunkt ableiten:

Standpunkt 2: IMDM sorgt dafür, dass die Herausforderungen durch Big Data im Business Intelligence gelöst werden können.

Die Autoren stimmen überein, dass die Datenmenge, der Integrationsbedarf und die Integrationsgeschwindigkeit in Unternehmen steigen werden. Für die Analysten von Gartner (2011) sind die steigenden Anforderungen mit herkömmlichen Mitteln nur schwer zu bewältigen und neue Technologien wie IMDM von Nöten (9). Speziell Lehner (2011) betont, dass die Performance von IMDM die Ausdehnung der Datenmenge noch verstärken wird [19]. Krueger et al. (2011) sowie Kemper und Neumann (2011) argumentieren, dass die Kapazität von Hauptspeicher schritthalten kann, sofern die Daten auf realen Transaktionen und Ereignissen basieren [13], [16]. Im Kontext schnell anwachsender, unstrukturierter Daten sieht Stonebraker (2011) [33] den Einsatz von IMDM als ausgeschlossen sowie Thiele et al. (2011) und Kossmann (2011) als eher begrenzt an [14], [36]. Die Argumente für IMDM sind an die Voraussetzungen gebunden, dass die Wachstumsgeschwindigkeit der Speicherkapazität von Hauptspeicher dem zu erwartenden Datenvolumen standhalten kann [31] und keine Kostengrenze durch die Anwender gezogen wird.

Tabelle 2. Einschätzungen in der Literatur zu Standpunkt 2

Quelle	Zustimmung	Neutral	Ablehnung
[3] Chamoni 2011	x		
[9] Gartner 2011	x		
[13] Kemper und Neumann 2011	x		
[14] Kossmann 2011		x	
[16] Krueger et al. 2011	x		
[19] Lehner 2011		x	
[21] Ousterhout et al. 2011	x		
[23] Piller und Hagedorn 2011	x		
[26] Plattner und Zeier 2011	x		
[28-29] Russom 2011	x		
[32] Sinzig und Sharma 2011	x		
[33] Stonebraker 2011			x
[36-37] Thiele et al. 2011		x	

Laut Ousterhout et al. (2011) und Piller und Hagedorn (2011) werden in der Zukunft Technologien verfügbar sein, die Datenvolumen im Petabyte-Bereich verarbeiten können [21], [23]. Chamoni (2011) und Russom (2011) attestieren IMDM in der Zukunft hohe Wachstumsmöglichkeiten hinsichtlich Big Data Analytics und Datenintegrationsanforderungen [3], [28-29]. Stonebraker (2011) kann diese Entwicklung noch nicht absehen [33]. Aspekte zur Reduzierung des Datenvolumens stellen die mögliche Integration analytischer und transaktionaler Datenbestände sowie besonders effiziente Kompressionsalgorithmen dar [23], [32]. Sollte die Kapazität dennoch nicht ausreichen, schlagen Plattner und Zeier (2011) Konzepte vor, die lediglich den „aktiven“ Datenbestand In-Memory vorhalten [26].

4.3 Integration von OLAP und OLTP

Unternehmensentscheidungen auf Basis operativer Daten zu treffen, wurde durch die künstliche Trennung von OLAP- und OLTP-Systemen unmöglich gemacht [26]. Die Fortschritte im Bereich von IMDM haben nun in der Literatur eine breite Diskussion ausgelöst, ob diese Trennung nun wieder aufgehoben werden kann. Für die Fusion der Systeme spricht im Wesentlichen die Kostenreduktion durch die Vereinfachung der Architektur [43], [48], die Reduzierung des Ressourcenverbrauchs [13], [25], [41], die flexible Analyse und Planung ohne vorgefertigte Analysepfade [33], [52], die Verfügbarkeit eines aktuellen Datenbestandes [16], [22], die reale Rückkopplung von Informationen in operative Prozesse [16], [36] und die Aufhebung von Medienbrüchen in Prozessen mit operativen und analytischen Anteilen [35]. Allgemein lässt sich der folgende Standpunkt aus der Literatur ableiten:

Standpunkt 3: IMDM wird dafür sorgen, dass OLAP und OLTP-Systeme langfristig integriert werden können.

Tabelle 3. Einschätzungen in der Literatur zu Standpunkt 3

Quelle	Zustimmung	Neutral	Ablehnung
[4] Davenport und Snabe 2011		x	
[13] Kemper und Neumann 2011	x		
[14] Kossmann 2011	x		
[16] Krueger et al. 2011	x		
[23] Piller und Hagedorn 2011	x		
[25] Plattner 2009	x		
[26] Plattner und Zeier 2011	x		
[32] Sinzig und Sharma 2011	x		
[33] Stonebraker 2011			x
[35] Strohmeier 2012		x	
[36-37] Thiele et al. 2011		x	
[40] Winsemann 2011		x	
[41] Winter 2011		x	
[42] Winter et al. 2011		x	
[45] Zeier et al. 2011	x		

Nach Ansicht der Mehrheit der Autoren kann die klare Trennung zwischen OLAP- und OLTP-Systemen, sofern sie technische Gründe zur Leistungssteigerung hatte, mit der Etablierung von IMDM aufgehoben werden. Speziell Plattner (2009) und Krueger et al. (2011) liefern durch die Auswertungen und Tests von produktiven Systemen gute Anhaltspunkte [16], [25]. Lediglich Stonebraker (2011) sieht zwar das Potential von IMDM in betrieblichen Anwendungssystemen, hält aber eine technische Trennung aufgrund der unterschiedlichen Arbeitslasten von OLAP und OLTP weiterhin für richtig [33]. Grundsätzlich kann die Qualität innerhalb der Entscheidungsunterstützung durch die hinzugewonnene Flexibilität im Analyseprozess verbessert werden. Allerdings stellte Strohmeier (2012) fest, dass heutige Analysefunktionalitäten im Kern auf taktische Untersuchungen ausgerichtet sind und Forschungsbedarf nach Konzepten zur operativen Entscheidungsunterstützung besteht [35]. Davenport und Snabe (2011) ergänzen, dass Informationen an die Zykluszeiten der Prozesse gekoppelt sind und diese entsprechend angepasst werden müssen [4].

In Zusammenhang mit der Vereinigung von OLAP und OLTP wird auch die Zukunft des DWH in der Literatur kontrovers diskutiert. Für Kossmann (2011) werden klassische DWHs in den meisten Anwendungsszenarien nicht mehr benötigt [14]. Laut Plattner und Zeier (2011) werden ETL-Prozesse komplett abgelöst und nötige Transformationen „on-the-fly“ berechnet [26]. Damit ist allerdings zur Verifizierung oder Korrektur der Daten keine Zeit mehr vorhanden [4]. Thiele et al. (2011) und Winsemann (2011) erkennen zwar, dass sich die Schichtenarchitektur innerhalb des DWH zunehmend konsolidieren und Aktualisierungszyklen verkürzen werden, allerdings ohne das Konzept des DWH aufzulösen [37], [41]. Außerdem bestehen gerade in heterogenen Systemlandschaften weiterhin die Anforderungen nach Daten-

integration und Harmonisierung [23], [35-36], [41]. Vor diesem Hintergrund sehen Winter et al. (2011) sowie Piller und Hagedorn (2011) IMDM als komplementäre Komponenten, die bei einem hohem Integrationsbedarf im Data Warehousing lediglich die Performancevorteile der Datenanalyse zur Verfügung stellen [23], [42].

5 Auswirkungen auf Geschäftsprozesse

Für die Praxis stellt sich die Frage, inwiefern die vorgenannten Aspekte von IMDM in Wettbewerbsvorteile transformiert werden können. Dem gegenüber steht die Gefahr drohender Fehlinvestitionen und Unsicherheit mangels fehlender interner Expertise [22]. Auch wenn in der Literatur vereinzelt Prozessbeispiele zur Veranschaulichung der Möglichkeiten der Technologie dargestellt werden, gibt es nur wenige Artikel die einen besonderen Fokus auf die Veränderungen der Geschäftsprozesse legen. Piller und Hagedorn (2011) haben in diesem Zusammenhang fünf Kriterien für Geschäftsprozesse definiert, die erste Entwicklungstendenzen aufzeigen und die Analyse nach praktikablen Einsatzgebieten unterstützen sollen [23]. In diesen Geschäftsprozessen lassen sich prinzipiell positive Effekte mittels IMDM erzielen, zum Beispiel eine Steigerung der Auswertungshäufigkeit, -flexibilität und -aktualität, sowie eine Vergrößerung der Datenbandbreite und des Detaillierungsgrads [23].

5.1 Ergebnisse der Literaturanalyse

Tabelle 4 stellt verschiedene Anwendungsfälle aus der Literatur vor, in denen Geschäftsprozessänderungen durch IMDM betrachtet wurden.

In den konkreten Szenarien ist auffällig, dass zu diesem frühen Forschungsstadium im Handelsumfeld durch Schütte (2011) sowie Piller und Hagedorn (2012) Potential für IMDM identifiziert wurde [23], [30]. Dabei ist entscheidend, dass in dieser Branche täglich mit großen Datenmengen umgegangen werden muss [30]. Grundsätzlich wurden Einsatzgebiete im Marketing [1], [23], in der Planung [26], [32], in der Disposition [24], [26], [30], in der Produktion [23] und im Vertrieb [44] genannt. Außerdem können durch IMDM branchenübergreifend neue mobile Applikationen entwickelt werden [44], die einen Einfluss auf zukünftige Geschäftsprozesse nehmen werden [26].

Tabelle 4. Szenarien für Prozessänderungen durch IMDM aus der Literatur

Quelle	Beschreibung
[1] Acker et al. 2011	Durch IMDM können Marketingkampagnen effizienter auf der Basis operativer Daten geplant werden, indem bei der Kundensegmentierung durch verbesserte Performance mehr Iterationen durchgeführt werden können.
[23-24] Piller und Hagedorn 2011/2012	Durch die bessere Verarbeitung von Massendaten leistet IMDM im Einzelhandel einen Beitrag u.a. zur Vermeidung von Out-of-Stock-Situationen, im Qualitätsmonitoring zur Fehlervorhersage und bei der Deckungsbeitragsanalyse auf der Basis von Einzelposten. Mehr Planungsiterationen machen sich positiv bei der Organisation von Marketingkampagnen und in der adaptiven Planung bemerkbar. Zudem regen die Autoren die Verwendung von IMDM zur Auswertung von Endverbraucherdaten (z. B. SmartMeter Analytics) an.
[26] Plattner und Zeier 2011	Durch IMDM können neue Applikationen und Möglichkeiten branchenübergreifend entstehen, z. B. Echtzeit-Einblick in tägliche Prozesse, Risiken im Risikomanagement, Analyse fehlender Produkte (out-of-stock) sowie die Ertragsleitung und Kundennachfrage.
[30] Schütte 2011	<i>Disposition</i> – IMDM bietet Einsatzpotential in der Bedarfsprognoserechnung, die Daten zu Aktionen, Abverkäufen, Marktentwicklungen und Profitabilität erfordert und damit hohe Integrations- und Performanceanforderungen stellt. <i>Aktionen</i> werden auf der Basis von historischen Verkaufsdaten innerhalb eines DWH geplant und in einem transaktionalen System festlegt. Diese künstliche Trennung entspricht nicht der betriebswirtschaftlichen Realität und kann durch den Einsatz von IMDM aufgehoben werden. <i>Absatz- und Umsatzanalysen</i> in Echtzeit nach Märkten und Regionen bei gleichzeitiger Simulation der Wettbewerbssituation sind aufgrund des großen Datenvolumens für IMDM besonders geeignet.
[32] Sinzig und Sharma 2011	Durch IMDM besteht die Möglichkeit, abteilungsübergreifende, integrierte, strategische, operative und finanzielle Planungsmodelle zu konzipieren und in Echtzeit zu berechnen. Die Auswirkung von Teilplanänderungen kann direkt nachvollzogen werden. Der Planungsprozess wird verkürzt und Simulationen zur Entscheidungsfindung werden kooperativ durchgeführt, was zu einer höheren Bindung an die abgestimmte Planung führt.
[44] Wust et al. 2011	Mobile Applikationen ermöglichen die Nutzung von betrieblichen Informationen in einer völlig anderen Art und Weise. Durch schnelle Antwortzeiten können Unternehmen neue Anwendungen entwickeln, die beispielsweise Vertriebsmitarbeiter im Außendienst unterstützen. Die Mitarbeiter erhalten die Informationen, die sie in dem jeweiligen Kontext benötigen. Das kann durch die Nutzung des operativen Datenbestandes zu einer erheblichen Serviceverbesserung führen, wodurch Wettbewerbsvorteile gegenüber Konkurrenzunternehmen erzielt werden können.

Durch die von Piller und Hagedorn (2011) identifizierten Eigenschaften können konkrete Geschäftsprozesse auf ihre potentielle Einsatzfähigkeit für IMDM überprüft werden [23]. Die Auswirkungen von IMDM auf Geschäftsprozesse lassen sich anhand der zu erzielenden Effekte ableiten. Dies ist gerade der Fall, wenn sich große Datenmengen oft, massiv und unvorhersehbar ändern sowie verschiedene Handlungsalternativen in Echtzeit gegeneinander abgewogen werden müssen [32]. Bezogen auf den technischen Hintergrund hängen die meisten Prozessveränderungen mit der stärkeren Analyseperformance zusammen. Diese kann die Auswertungshäufigkeit und -flexibilität ohne vordefinierte Strukturen auf Basis historischer und aktueller Daten auch in komplexen Fragestellungen massiv steigern. Dazu können neue Datenquellen angebunden [24] und neue Applikationen in Geschäftsprozesse eingebunden werden [1], [24], [30], die das Potential eröffnen Prozesse effektiver und effizienter zu gestalten. Wenn eine massive Datenintegrationsanforderung in Echtzeit besteht, würde die vollständige Integration von OLAP und OLTP ebenfalls Prozessveränderungen hervorrufen, und analytische und transaktionale Prozesse konsolidiert werden können [30]. Außerdem würden sich die Bandbreite und der Detaillierungsgrad von Informationen vergrößern, was die Qualität von Analysen und Simulationen verbessern kann [32]. Dazu kann durch die hinzugewonnene Aktualität in operative Prozesse spontan eingegriffen werden [24], [26].

5.2 Vorläufige Ergebnisse der Fallstudienanalyse

Um die erzielten Ergebnisse der Literaturanalyse zu überprüfen und weitere Erkenntnisse zu sammeln, wurde eine qualitative Befragung durchgeführt. Die Interviews wurden bei der arvato Systems GmbH erhoben. Insgesamt wurden auf der Basis eines Fragenkatalogs vier semi-strukturierte Interviews durchgeführt. Bei den Interviewpartnern handelt es sich um erfahrene Berater mit dem Schwerpunkt Business Intelligence im SAP-Umfeld.

Die Interviewpartner haben im Wesentlichen Potential für IMDM an Stellen identifiziert, an denen Massendaten verarbeitet und zeitnah Entscheidungsgrundlagen gelegt werden müssen. Ein Berater beschreibt die Voraussetzung für IMDM als:

„[...] die Situation in der ich in kürzester Zeit schnelle Entscheidungen treffen muss und damit auf eine hohe Performance angewiesen bin. [...] Ich kann aber nur das in Echtzeit darstellen, was im System in Echtzeit verfügbar ist.“

Dabei stellen zwei Gesprächspartner fest, dass IMDM die Möglichkeit bietet, die Analyse vielmehr auf die Zukunft auszurichten. Denkbar sind hier neue noch unbekannte Applikationen, die erst durch IMDM möglich werden und die Entscheidungsqualität erhöhen können. Häufig wurden mobile Applikationen als Anwendungsfeld genannt. Die Modellierung wird sich vereinfachen, indem Schichten konsolidiert und Aggregate verschwinden werden. Ein Interviewpartner ergänzt:

„Der [...] Prozess verändert sich insofern, dass durch die Wiederholung des Prozesses in kürzeren Zeitabständen Informationen gewonnen werden. [...] Wenn ich den

Speicherplatz und die Performance zur Verfügung habe, ist es mir egal, ob ich mehr Informationen aufnehme.“

Das größte Hindernis zur Etablierung von IMDM stellt aus der Sicht der Befragten der verhältnismäßig hohe Preis dar. Zusätzlich wird angemerkt, dass gar nicht so viele Anwendungsfälle eine Echtzeitverarbeitung benötigen und viele Szenarien mit herkömmlichen Mitteln umsetzbar sind. Das Thema Big Data wird weitestgehend als Chance für IMDM wahrgenommen, wo unstrukturierte Daten in strukturierte Informationen transformiert werden. Die Befragten sind nicht davon überzeugt, dass transaktionale und analytische Systeme mittelfristig komplett verschmelzen werden, auch wenn sich das Berichtswesen wieder mehr in das OLTP-System verlagern und langfristig eine Fusion stattfinden könnte. Ein Berater erklärt:

„In einer Welt mit einem System würde im Prinzip nur noch ein „Transform“ stattfinden. [...] Die Auswirkungen einer Buchung werde ich realtime sofort erkennen können.“

Dabei wird von einem Berater angemerkt, dass ohne eine saubere Struktur leicht Verwirrung entstehen kann. Entscheidend ist für alle Gesprächspartner, dass weiterhin die Funktionen des DWH erhalten bleiben und die Möglichkeit besteht, weiterhin zusätzliche Datenquellen anzubinden.

Insgesamt schätzen alle Befragten analog zur Strategie der SAP das Potential von IMDM als hoch ein. IMDM kann die Durchführung von Entscheidungsprozessen kurzfristiger, häufiger, detaillierter und zukunftsorientierter gestalten und mit neuen Applikationen anreichern. Als entscheidender Aspekt wird die gesteigerte Analyseperformance genannt. Die Integration transaktionaler und analytischer Systeme wird aufgrund des erheblichen Integrationsbedarfs kritisch gesehen. Zusätzlich werden steigende Ansprüche im BI-Umfeld beobachtet und IMDM als Teilkomponente zur Erfüllung gerade mobiler Anforderungen identifiziert.

Zusammenfassung und Diskussion

Dieser Artikel trägt zu einem tieferen Verständnis der potenziellen Auswirkungen von IMDM auf betriebliche Geschäftsprozesse bei. Dazu konnten innerhalb einer Literaturanalyse eine Reihe von wissenschaftlichen Publikationen identifiziert werden, die die Auswirkungen von IMDM auf Geschäftsprozesse thematisieren. Drei zentrale Diskussionsaspekte liegen diesen Prozessänderungen zugrunde: Die bessere Abfragegeschwindigkeit mit der Möglichkeit zur Echtzeit-Analyse, das Zusammenspiel zwischen Big Data und IMDM sowie die mögliche Vereinigung von OLAP und OLTP Systemen. Die Erkenntnisse der Literaturanalyse konnten innerhalb einer qualitativen Erhebung vorläufig geprüft und durch zusätzliche Aspekte ergänzt werden.

Auswirkungen auf die Praxis. Die Analyse hat gezeigt, dass IMDM eine technologische Innovation darstellt, die das Potential hat, die wachsenden Herausforderungen im BI-Umfeld zu erfüllen. Dabei sind insbesondere die wachsenden Anforderungen durch den Anwender zu nennen, die aus privaten Erfahrungen mit Endverbraucher-IT

resultieren [27]. IMDM ermöglicht neue Formen von Anwendungsklassen, um Entscheidungsprozesse adressatengenau zu unterstützen. Zu deren Verwirklichung sind allerdings weitere innovative Ideen aus der Praxis notwendig [45]. Ein entscheidender Aspekt dabei ist die Höhe des Integrationsbedarfs. In heterogenen Systemlandschaften mit unterschiedlichen Datenquellen werden umfangreiche Operationen zur Extraktion und Harmonisierung nach wie vor relevant bleiben [31], [45], [51]. Es muss demnach konkrete Prozessanforderungen geben, die Änderungen oder Einschränkungen bei der Datenintegration (z. B. Echtzeitanalysen) rechtfertigen, und einen signifikanten Nutzenzuwachs bieten [6], [8]. Demnach sind die Anwendungsbereiche für IMDM von den Eigenschaften der Geschäftsprozesse und den zu erzielenden Effekten abhängig. Unternehmen können die Potentiale speziell in der Planung, Disposition, Produktion, Vertrieb und Marketing überprüfen und die Auswirkung von IMDM analysieren.

Auswirkungen auf die Forschung. Derzeit lassen sich nur wenig Arbeiten auffinden, die die Auswirkungen von IMDM auf Geschäftsprozesse und Implikationen für das Business Process Management (BPM) thematisieren. Eine Ausnahme sind die Arbeiten von Piller und Hagedorn [23]. Interessanterweise wird das gesamte Thema IMDM primär in Deutschland vorangetrieben, während sich die Forschung im nordamerikanischen Raum eher auf die technischen Aspekte konzentriert. Daraus ergibt sich die Notwendigkeit das Thema auch in internationalen Forschungsbeiträgen zu platzieren. Für eine umfassende wissenschaftliche Analyse ist zudem weitere empirische und vor allem domänen-spezifische Forschung notwendig, um konkrete Verbesserungen von BPM zu erarbeiten. In diesem Zusammenhang regt Winter (2011) an, für aktuelle Herausforderungen aus fachlicher Sicht Referenzszenarien zu spezifizieren [41].

Einschränkungen und Ausblick. Neben den naturgemäßen Restriktionen qualitativer Forschung konnten zunächst nur vier Befragungen im Rahmen der Fallstudienanalyse durchgeführt werden. Nichtsdestotrotz bieten diese einen ersten Ansatz inwiefern die Methodik zu fortschreitenden Erkenntnisgewinnen beitragen kann. Weitere Interviews zur Ergänzung mit verschiedenen BI-Experten sind bereits vereinbart. Wir denken, dass der Artikel das Thema zusätzlich anreichert und sehen ermutigende Potentiale für weitere Forschung in der Schnittstelle von IMDM und BPM.

Literatur

1. Acker, O. et al.: In-memory analytics – strategies for real-time CRM. *Journal of Database Marketing & Customer Strategy Management* 18 (2), 129–136 (2011)
2. Bayer, M.: Wo Datenverarbeitung in Echtzeit sinnvoll ist, <http://www.cio.de/knowledgecenter/bi/2676310/>
3. Chamoni, P.: BI-Strategie zum Ausgleich von Technologie-Push und Business-Pull. In: *Innovative Unternehmensanwendungen mit In-Memory Data Management (IMDM 2011)*. LNI, Vol. P-193, pp. 13–22. GI, Bonn (2011)

4. Davenport, T.H., Snabe, J.H.: How Fast and Flexible Do You Want Your Information, Really?. MIT Sloan Management Review 52 (3) (2011)
5. Eich, M.H.: Main memory database research directions. In: Boral, H., Faudemay, P. (eds.): IWDM '89. LNCS, Vol. 368, pp. 251–268. Springer, Berlin Heidelberg (1989)
6. Fabian, B., Günther, O.: In-Memory-Datenmanagement für Business Intelligence. Wirtschaftsinformatik 53 (6), 388 (2011)
7. Gantz, B.J., Reinsel, D.: Extracting Value from Chaos, <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>
8. Garcia-Molina, H., Salem, K.: Main memory database systems: an overview. IEEE Transactions on Knowledge and Data Engineering 4 (6), 509–516 (1992)
9. Gartner: Gartner Identifies the Top 10 Strategic Technologies for 2012, <http://www.gartner.com/it/page.jsp?id=1826214>
10. Gates, B.: Information At Your Fingertips. Fall/COMDEX., Las Vegas (1994)
11. Gray, J.: Tape is Dead, Disk is Tape, Flash is Disk, RAM Locality is King, research.microsoft.com/en-us/um/people/gray/talks/Flash_is_Good.ppt
12. Hackathorn, R.: Minimizing Action Distance. The Data Administration Newsletter 25 (3), 22–23 (2003)
13. Kemper, A., Neumann, T.: HyPer: A hybrid OLTP&OLAP main memory database system based on virtual memory snapshots. In: IEEE 27th International Conference on Data Engineering, pp. 195–206. IEEE (2011)
14. Kossmann, D.: Einfachheit lautet die Devise. Wirtschaftsinformatik 53 (6), 387–388 (2011)
15. Krueger, J. et al.: Hauptspeicherdatenbanken für Unternehmensanwendungen. DatenbankSpektrum 10 (3), 143–158 (2010)
16. Krueger, J. et al.: Main memory databases for enterprise applications. In: 18th International Conference on Industrial Engineering and Engineering Management, pp. 547–557. IEEE (2011)
17. Lechtenböcker, J., Vossen, G.: Hauptspeicherdatenbanktechnologie: Herausforderungen und weitere Entwicklung. Wirtschaftsinformatik 53 (6), 384–385 (2011)
18. Lehman, T.J., Carey, M.J.: A recovery algorithm for a high-performance memory-resident database system. ACM SIGMOD Record 16 (3), 104–117 (1987)
19. Lehner, W.: „In-Memory Data Management“ – Evolution oder Revolution. Wirtschaftsinformatik 53 (6), 386–387 (2011)
20. Loos, P.: Kommentare zu „In-Memory-Datenmanagement in betrieblichen Anwendungssystemen“ - Einleitung. Wirtschaftsinformatik 54 (4), 209–213 (2012)
21. Ousterhout, J. et al.: The case for RAMCloud. Communications of the ACM 54 (7), 121–130 (2011)
22. Oxford Economics - cooperation with SAP: Real-time Business Playing to win in the new global marketplace. Oxford (2011)
23. Piller, G., Hagedorn, J.: Einsatzpotenziale für In-Memory Data Management in betrieblichen Anwendungssystemen. Wirtschaftsinformatik & Management 3 (5), 18–25 (2011)
24. Piller, G., Hagedorn, J.: In-Memory Data Management im Einzelhandel: Einsatzbereiche und Nutzenpotentiale. Multikonferenz Wirtschaftsinformatik 2012 (2012)
25. Plattner, H.: A common database approach for OLTP and OLAP using an in-memory column database. In: Etintemel, U. et al. (eds.): Proceedings of the 35th SIGMOD international conference on Management of data - SIGMOD '09, pp. 1–7. ACM Press, New York, USA (2009)
26. Plattner, H., Zeier, A.: In-Memory Data Management: An Inflection Point for Enterprise Applications. Springer (2011)

27. Richardson, J.: The Consumerization of BI Drives Greater Adoption. Gartner, Inc. (2011)
28. Russom, P.: Big Data Analytics. The Data Warehouse Institute (2011)
29. Russom, P.: Next Generation Data Integration. The Data Warehouse Institute (2011)
30. Schütte, R.: Analyse des Einsatzpotenzials von In-Memory- Technologien in Handelssystemen. In: IMDM 2011. LNI, Vol. P-193, pp. 1–12. GI, Bonn (2011)
31. Schütte, R.: In-Memory-Technologien: Überlegungen zur Begründbarkeit und zum Einsatz beim Betrieb von großen Systemen. *Wirtschaftsinformatik* 54 (4), 211–213 (2012)
32. Sinzig, W., Sharma, K.R.: In-Memory-Technologie : Verbesserungen bei Planung, Simulation und Entscheidungsunterstützung. *Wirtschaftsinformatik & Management* 5, 18–23 (2011)
33. Stonebraker, M.: Stonebraker on data warehouses. *Communications of the ACM* 54 (5), 10–11 (2011)
34. Strauss, A., Corbin, J.: Basics of qualitative research: Grounded theory procedures and techniques. Sage (1990)
35. Strohmeier, S.: Hauptspeicherdatenbanken in der betrieblichen Informationsversorgung – Technische Innovation und fachliche Stagnation?. *Wirtschaftsinformatik* 54 (4), 209–210 (2012)
36. Thiele, M. et al.: Data-Warehousing 3.0 – Die Rolle von Data-Warehouse- Systemen auf Basis von In-Memory-Technologie. In: Innovative Unternehmensanwendungen mit In-Memory Data Management (IMDM 2011). LNI, Vol. P-193, pp. 57–68. GI, Bonn (2011)
37. Thiele, M., Lehner, W.: Echtzeit-Data-Warehouse-Systeme. *Datenbank-Spektrum* 11 (3), 207–211 (2011)
38. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*. 26 (2), xiii–xxiii (2002)
39. Wieland, C., Lüdeking, M.: Trendmonitor Business Intelligence. RAAD Research (2012)
40. Winsemann, T.: Kriterien für Datenpersistenz bei Enterprise Data Warehouse Systemen auf In-Memory Datenbanken. In: Proceedings of the 23. GI- Workshop on Foundations of Databases, pp. 97–102 (2011)
41. Winter, R.: In-Memory-Appliances – Ein Fortschritt in der Informationslogistik? *Wirtschaftsinformatik* 53 (6), 388–390 (2011)
42. Winter, R. et al.: Revolution or Evolution? Reflections on In-Memory Appliances from an Enterprise Information Logistics Perspective. In: Innovative Unternehmensanwendungen mit In-Memory Data Management (IMDM 2011). LNI, Vol. P-193, pp. 23–34. GI (2011)
43. Winter, R.: Why Are Data Warehouses Growing So Fast ?, www.b-eye-network.com/print/7188
44. Wust, J. et al.: xSellerate: Supporting Sales Representatives with Real-Time Information in Customer Dialogs. In: Innovative Unternehmensanwendungen mit In-Memory Data Management (IMDM 2011). LNI, Vol. P-193, pp. 35–44. GI, Bonn (2011)
45. Zeier, A. et al.: Potenzial von In-Memory-Technologien. *Wirtschaftsinformatik* 53 (6), 385–386 (2011)

Coupling Metrics for EPC Models

Daniel Braunnagel and Florian Johannsen

University of Regensburg, Department of Management Information Systems, Regensburg,
Germany

{daniel.braunnagel,florian.johannsen}@wiwi.uni-regensburg.de

Abstract. Process modeling is a decisive task for modern enterprises. The effectiveness of IS development largely depends on the quality of conceptual models and their understandability. However, process model quality is still a fuzzy concept and not fully understood yet. Recently, coupling became a concept for assessing model quality, but still there is a lack of research in transferring “coupling” to business process modeling. The field of software engineering has shown the importance of measuring coupling as a means for judging the quality of a design. Therefore, this paper collects a range of coupling metrics from the field of software engineering and transfers them to event driven process chains (EPC). Further the metrics are applied to different process models and implications for the process model quality are discussed.

Keywords: Coupling, event driven process chain, process model quality

1 Introduction

Business process modeling is a decisive task for modern enterprises (see e.g. [1-2]). Business process modeling captures employees’ process knowledge so that it can be used for entrepreneurial initiatives. Process models support decisions on IT-investments, the development of information systems and the improvement of processes (see [1]). Moody [3] states that the efficiency and effectiveness of IS development largely depends on the quality of conceptual models that guide IS implementation though their evaluation is more an art than a systematic procedure.

The creation of process models is as a highly subjective process [4]. Usually different persons, such as IT-employees or business analysts, are involved in the design of process models while a generally accepted approach for creating a process model is missing [5]. In addition users demand different levels of abstraction. Whereas a software engineer is interested in details concerning the control flow structure of a process to derive requirements on information systems (see e.g. [6-7]), managers usually prefer more abstract descriptions enabling strategic decision making [8-9]. For utilizing the benefits of process models, e.g. for software development, they need to be easy to understand and maintain [10].

However, the quality of business process models remains a fuzzy topic. According to Mendling et al. [2] quality frameworks, metrics, empirical surveys as well as prag-

matic guidelines were introduced in recent years, dealing with quality aspects of process modeling [2]. Commonly accepted definitions of the term “process model quality” as well as standardized criteria for evaluating process models still are missing.

Frameworks, such as the guidelines of modeling (see e.g. [4], [11]) (GoM) or the SEQUAL model (see [12-13]) deliver criteria, for example “construction adequacy” (see [4]), which can be used for assessing the quality of conceptual models. However an evaluation of conceptual models based on these criteria is strongly affected by the subjectivity of the user [3]. This is because a conceptual model can only be evaluated against user’s expectations and is not to be considered as a “finished product” that can be judged on the basis of a specification [3]. In addition quality frameworks have not been widely accepted in practice and a standard has not yet emerged [3].

As Mendling et al. [2] state, manifold empirical studies on the maturation of business process modeling languages can be found (see e.g. [14-16]) [2]. Several authors (see e.g. [17-21]) introduce criteria that can be used for evaluating modeling languages. However, the object of interest in these studies is the modeling language used, not the process model itself (see [2]).

Pragmatic guidelines that can be found in literature (see e.g. [22-23]) are often too generic (e.g. “keep it simple”) [22] to support a practitioner in a modeling project (see also [2]).

In recent years, literature has focused on the development of metrics enabling an objective evaluation of process models (see e.g. [24-30]). Vanderfeesten et al. [31] assign these metrics to certain categories of process model quality. While manifold metrics for judging a process model regarding complexity or size do exist, coupling metrics for process models are still underrepresented. While coupling of modules is a well-established quality characteristic for information systems (see [32]), research has only begun to transfer this concept to business process modeling thus enabling a new perspective on process model quality. The aim of this paper is to expedite this research by introducing coupling metrics that originate in software engineering specifying them for business process modeling. Afterwards, the metrics that got transferred are applied using an example. The suggestions in this paper focus on event driven process chains (EPCs), since the interpretation of the coupling metrics will be different for different modeling languages varying in language expressiveness (see [33]).

The structure of the paper is as follows. In section 2 the basics are explained. These comprise the event driven process chains and coupling. Section 3 presents the coupling metrics found and describes how they were transferred to the EPC. Section 4 discusses the processes used as examples and the results from applying the metrics. Finally, section 5 summarizes the paper, discusses implications of the metrics for the field of process model quality and shows further work remaining.

2 Basics

2.1 Event Driven Process Chain

The EPC was developed at the University of Saarland in cooperation with SAP AG. The EPC is known for being used as the modeling concept supporting SAP R/3 and for its use as part of the modeling framework ARIS. [34]

The event driven process chain can be defined as a graph, consisting of nodes and directional arcs (see [35-38]). The nodes may be specified as the union set of the set of events, functions, connectors, process interfaces and resources. The set of connectors is the union of the sets of and-connectors, or-connectors and xor-connectors. The set of resources is the union of several sets, with information elements being one of them. Each node in the above set is connected to at least one arc, with each arc being connected with precisely two nodes. No arc connects two functions or two events, they alternate in a path with an arbitrary number of connectors allowed in-between. Resources are connected exclusively to functions, process interfaces exclusively to events. Functions need to be connected with at least two nodes situated among the connectors or events. An EPC model is considered a graph according to the above definition. The control flow is considered the path connecting process interfaces, events, functions and connectors. A model has a beginning, consisting of events without predecessor or process interfaces without a preceding event, and an end, being events without succeeding node or a process interface without succeeding events. The control flow, however, may be continued over multiple models in case of process interfaces referencing each other. [35-38]

2.2 Coupling

The term “coupling” is most generally defined as “being connected for consideration together” [39]. Closer to the context at hand, the field of software engineering presents different more specific interpretations of coupling.

The first interpretation is based on the ontology of Bunge-Wand-Weber [40-41]. Accordingly, two things are coupled if they interact at some moment in time. This interpretation is employed by e.g. the RFC metric. This metric counts the number of methods “that can be invoked in response to a message to an object of the class” [42]. In other words, two objects of an object oriented design are considered as being “coupled” whenever one object calls a method of another object. The metric measures the degree of coupling by counting the number of methods that can be invoked.

A second interpretation focuses graph-theory. Thereby, the graph is analyzed regarding the way its elements are connected. For example, McCabe [43] builds the control flow graph of programs to calculate the cyclomatic complexity. Counting the nodes, arcs and exit nodes, the metric calculates the number of independent circuits in the control flow. The notion of coupling therefore refers to the paths through the programs code. With reference to the EPC, the weighted coupling metric [29] and the cross connectivity metric [30] were presented in this field.

A third interpretation references information theory. This interpretation aims to quantify aspects of coupling using the information content [44]. An implementation of this interpretation is presented by Halstead [45] calculating the “program length”. Their program length is sensitive towards the reuse of statements and therefore considers the coupling between code modules by their reuse of code statements.

3 Coupling in the Context of EPC

To cover a wide range of different coupling interpretations, literature presenting existing metrics was searched for. The well-known literature databases Google Scholar, Computer.org (IEEE Computer Society), AISel and Emerald Insight, that offer a wide range of different electronic sources were queried using the term pair “coupling metrics” “business process model” and “coupling metrics” itself. 47 results were considered as relevant and downloaded, consisting of 33 conference papers and nine journal papers that passed a peer review process. In addition, four technical reports and one book were found.

The metrics covered in these sources and their transferability to the EPC is shown in Table 1.

Table 1. Coupling Metrics

Source	Metric	Transfer	Not transferred because...
[46]	Coupling of a modular system, IntramoduleCoupling	Yes	
[47]	Coupling of a module, Intramodule coupling of a module	Yes	
[48]	PIM	No	... requires dynamic language features (i.e. polymorphism, reflection...).
[49]	PPEP, EMC	No	... requires failure rates of components.
[42]	Depth of inheritance, No. of children	No	.. requires inheritance.
[42]	RFC, CBO	Yes	
[50]	Static / Dynamic Coupling	No	... requires inheritance.
[51]	Direct coupling, indirect coupling, total coupling	Yes	
[52]	Procedure complexity	No	... already transferred [27, 53].
[54]	Object level coupling	No	... requires inheritance / locality of data.
[55]	Interface coupling	No	... equivalent to [56].
[57]	Conceptual coupling of Services	No	... requires locality of data.
[43]	Cyclomatic complexity	No	... already transferred [53].
[58]	Conceptual coupling	Yes	
[59]	CBS ... DCSS	No	... is an implementation of [42] and [43].
[60]	ASSD ... ASPD	No	... requires statefulness.
[61]	CIC ... AMC	No	... requires inheritance.
[56]	Process Coupling	Yes	
[29]	CP	No	... already specified for eEPC.
[30]	CC	No	... already specified for eEPC.

The procedure for transferring the metrics to the EPC can be described as follows. In a first step, the concepts behind the variables of each formula are identified. The description of each concept is then used to identify equivalent concepts within EPC models. Finally, the found concepts are quantified and used to reformulate the original metrics (see Figure 1).



Fig. 1. Transfer procedure

However, in some cases metrics could not be transferred. This was the case whenever the metric comprised constructs for which no equivalent could be found in the EPC. For example, some metrics refer to the inheritance hierarchy of class objects in object oriented programming. An equivalent mechanism for inheritance among process models was not discovered. Further, since modeling takes place on the type level, metrics referring to runtime information or states could not be transferred either.

The metrics that could be transferred using this procedure are discussed in the following.

In [46], Allen et al. present two related metrics called “Coupling of a modular system” and “IntramoduleCoupling”. The motivation of these metrics is the limited capacity of the human short time memory. When the amount of information in a model breaks this limit, a user will not be able to fully realize the model, which will lead to problems in understanding the model. The metrics therefore calculate the information content of a model regarding different aspects of its graph structure. The metric “coupling of a modular system” calculates the excess entropy of the graph structure in modules. The “IntramoduleCoupling” quantifies the excess entropy of the graph structure connecting modules. Transferred to the EPC, the IntramoduleCoupling measures the repetitiveness of the patterns connecting models via process interfaces or a model hierarchy. The coupling of a modular system measures the repetitiveness of patterns in the control flow of separate models.

Allen et al. present a second pair of metrics in [47]. Adapting the prior metrics, the “coupling of a module” calculates the information content in the graph structure in each module. The “Intramodule coupling of a module” calculates the information content of the graph structure connecting the graphs. Transferred to the EPC, the coupling of a module quantifies the amount of information referring to the patterns of arcs in a model a user needs to assess in order to understand the model. The intramodule coupling, on the other hand, regards the information in the connections between models.

Chidamber/ Kemerer present two coupling metrics in [42]. The first metric, CBO, counts the number of classes one class is associated with by calling its methods or variables. The second metric RFC counts the methods in one class and all the methods in other classes that can be called from within. Transferred to the EPC, the RFC counts the functions and interfaces or hierarchies. CBO counts the number of models one model is connected to.

In [51], Gui presents three related metrics, namely the direct coupling metric, the indirect coupling metric and the total coupling metric. The direct coupling metric is calculated for two classes, it is the relation of methods and variables in the second class called by the first class divided by the total number of methods and variables in the second class. As for the second metric, indirect coupling extends the prior metric for classes without direct connection. For any pair of classes for which a connecting

path via called methods or resources can be found, their indirect coupling is calculated by multiplying the direct coupling values for each pair of classes on the path. The third metric is calculated by dividing the sum of indirect coupling values of all pairs of classes by the number of potential class pairs. Transferred to the EPC, two process models are considered coupled if one model contains a process interface or hierarchical function referencing the other. Accordingly, the direct coupling metric, when transferred, divides the number of references from a second model by the number of functions and process interfaces contained in one model. The second and third metric are used like originally described.

Poshyvanyk/ Marcus [58] describe a metric using the information retrieval technique latent semantic indexing to discover a semantic structure among the textual content of source code. The metric is called conceptual coupling metric [58]. They assume that similar concepts are expressed with similar terms. Therefore calculating the co-occurrence of terms indicates how the strength of the relation between concepts. The LSI transformation of a term-document matrix containing variable text from classes as terms and from the class structure as documents presents such co-occurrence of terms. Using the strongest indicators for concepts from this transformation, a new term-document matrix can be built which is then used to calculate the similarity between classes and groups of classes. Transferred to the EPC, the term-document matrix is built using node labels as terms and models as documents. The metric then calculates the conceptual similarity between models and groups of models chosen by the user. These groups can, e.g., belong to one or more processes. The adapted metric therefore calculates the co-occurrence of terms between groups of process models. Assuming the co-occurrence is an adequate measure for a conceptual structure, this indicates the conceptual similarity of process models.

Another approach towards coupling is presented by Reijers/ Vanderfeesten [56]. The metric is defined for a so-called information element structure, which is a graph structure with nodes representing information elements and arcs representing operations. An activity can be described as a partition in the said structure. Operations are considered coupled if they are connected to a common information element. Activities are considered coupled, if they contain coupled operations. Transferred to the EPC, functions take the place of activities, information elements are used as such. Functions are considered coupled whenever they are connected to common information elements.

4 Illustrating Example

The metrics described above are demonstrated using three groups of process models. Each of these groups represents the same situation, but whereas group 1 contains models without syntactical errors, group 2 and group 3 have an increasing number of errors. Figure 2 and Figure 3 show an excerpt from the same process model in group 1 and group 2. The second model contains redundant events after the first function and misses the decision and the corresponding event after “Request Schufa-Report”. Further errors (not pictured) are e.g. missing arcs after splits.

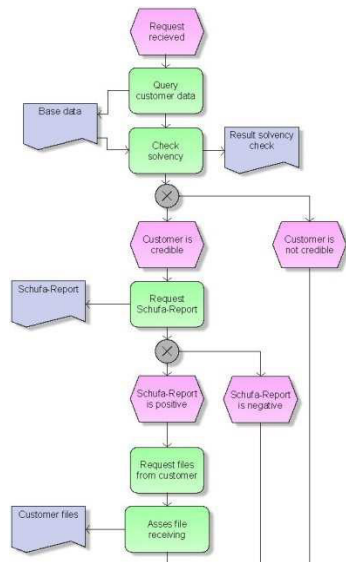


Fig. 2. Validate solvency, group 1

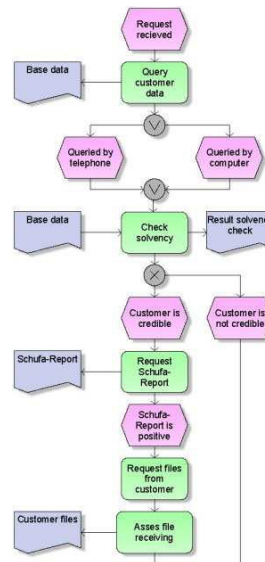


Fig. 3. Validate solvency, group 2

The first two groups of process models comprise three sub-models, the last group four sub-models. The models are quantified in Table 2.

Table 2. Example models

	Group 1	Group 2	Group 3
Credit Application			
Functions	2	2	4
Events	4	4	7
Information elements	0	0	0
Validate solvency			
Functions	13	13	9
Events	11	14	10
Information elements	11	13	7
Final decision			
Functions	9	7	5
Events	8	6	5
Information elements	2	2	4
Proceed credit application			
Functions			4
Events			6
Information elements			8

4.1 Allen et al.

The metrics of Allen et al. (see [46-47]) calculated for the models at hand are shown in Table 3.

Table 3. Metrics of Allen et al.

Metric\Group	Group 1	Group 2	Group 3
Coupling of a Module (Credit application)	33.645	33.612	50.220
IntramoduleCoupling(Final decision)	225.474	175.251	139.152
IntramoduleCoupling(Validate solvency)	458.235	587.723	359.202
IntramoduleCoupling(Credit application)	42.292	42.292	93.603
IntramoduleCoupling(Proceed credit application)			200.088
Coupling of a modular system	11.404	11.939	19.241
Intramodulecoupling of a modular system	296.953	345.767	270.727

The metrics coupling of a module hardly differ between the “group 1” and “group 2” model. The values of the “group 3” models differ due to the higher number of connections between models since this decomposition uses more models.

Regarding the models of “Final decision” and “Validate solvency” the metric suggests that “group 3” and “group 2” models are easier to assess by the model user. The values do, however, mostly result from modeling errors omitting necessary arcs in the case of “group 2” or from a smaller sub-model due to a higher degree of decomposition in the case of “group 3”.

The model “Credit application” is identical in the groups of “group 1” and “group 2”. In the case of “group 3”, the model encompasses more nodes and arcs and therefore has a higher metric value.

The coupling of a modular system hardly differs between “group 1” and “group 2” since both their decomposition comprises the same amount of models. The last group presents a higher value due to the higher amount of sub-models.

The intramodulecoupling mostly represents the number of nodes and arcs in the corresponding sub-models.

4.2 Chidamber/ Kemerer

The metrics of Chidamber/Kemerer (see [42]) calculated for the models are shown in Table 4.

Table 4. Metrics of Chidamber/ Kemerer

Group	Model\Metric	RFC	CBO
Group 1	Credit application	2	2
	Validate solvency	13	
	Final decision	9	
Group 2	Credit application	2	2
	Validate solvency	13	
	Final decision	7	
Group 3	Credit application	4	3
	Validate solvency	9	
	Proceed credit application	4	
	Final decision	5	

The RFC and CBO mostly count objects. Accordingly, their values depend on the number of functions and process interfaces or linked sub-models respectively.

4.3 Gui/ Scott

The Gui/Scott metrics (see [51]) are calculated for the example at hand in Table 5. Since the model “Credit application” is the only model referencing sub-models, it is also the only model that can form a pair of coupled models with the remaining models.

Table 5. Metrics of Gui/ Scott

Metric\Group	1: Credit application	2: Credit application	3: Credit application
Direct Coupling Metric			
Credit application	0	0	0
Validate solvency	0.07692	0.07692	0.11111
Proceed credit application			0.25
Final decision	0.11111	0.14285	0.2
Indirect coupling Metric			
Credit application	0	0	0
Validate solvency	0.07692	0.07692	0.11111
Proceed credit application			0.25
Final decision	0.11111	0.14285	0.2
Total coupling metric	0.03133	0.03663	0.04675

Since all the models in all decompositions are linked to one higher decomposition level, they all have one incoming connection and no other links. Therefore the determining variable for this metric is the number of functions in each model. Accordingly, the values mostly represent the number of activities. A higher number of activities lead to a lower metric value.

The indirect coupling metrics do not differ from the direct ones, since in this case there are no paths of more than two models.

The sum of indirect coupling values is put in relation to the maximally possible number of model pairs for the total coupling metric.

4.4 Poshyvanyk

The metric of Poshyvanyk et al. (see [58]) was used with the exemplary models, though with one limitation. Since the scenario was modeled in only one group with three alternatives, solely the conceptual similarity of models instead of the similarity of model groups were calculated (Table 6 - Table 8).

Table 6. Conceptual similarity of models, Group 1

Model\Model	Final decision	Validate solvency	Credit application
Final decision	1	0.88297	0.73799
Validate solvency	0.88297	1	0.82660
Credit application	0.73799	0.82660	1

Table 7. Conceptual similarity of models, Group 2

Model\Model	Validate solvency	Final decision	Credit application
Validate solvency	1	0.85847942	0.82395738
Final decision	0.858479424	1	0.71707854
Credit application	0.82395738	0.71707854	1

Table 8. Conceptual similarity of models, Group 3

Model\Model	Credit application	Final decision	Credit application	Validate solvency
Credit application	1	0.86429	0.86587	0.90647
Final decision	0.86429	1	0.84521	0.84859
Credit application	0.86587	0.84521	1	0.83160
Validate solvency	0.90647	0.84859	0.83160	1

As can be seen, for all groups the conceptual similarity of models declines along the rising distance in the control flow. In addition to that, the “group 1” and “group 2” models hardly differ. Indeed, since these metrics regard the co-occurrence of labels, their values do not reflect syntactical violations. The last decomposition is hardly comparable since the number of models differs.

4.5 Reijers/ Vanderfeesten

The coupling metric of Reijers/ Vanderfeesten (see [56]) is used in combination with the above process models to calculate the values in Table 9.

Table 9. Metrics of Reijers/ Vanderfeesten

Group 1	Validate solvency	0.06410
Group 2	Validate solvency	0.06410
Group 3	Proceed credit application	0.33333
	Validate solvency	0.02777

The only models containing coupled activities are “Validate solvency” in all groups, and “Proceed credit application” in the group 3 decomposition. There is no difference in the coupling among the first two decompositions since the models do not differ regarding their coupled functions. The last decomposition’s models do differ since their functions are split over two models.

4.6 Discussion

The metrics transferred before were applied to three example processes. It was shown that coupling metrics from the field of software engineering can in fact be transferred to EPC models. Hence the metrics presented in this work extend the existing set of metrics (see [31]).

However, the calculation of these metrics is laborious. E.g. the metrics of Allen et al. require a separate incidence matrix for each node in a number of models. The metric of Poshyvanyk requires a singular value decomposition, which usually can only be

calculated by using specialized software. Therefore, tool support is necessary for calculating these metrics. These tools need to be developed in future work.

Finally, as remarked before, the metric values hardly differ among the models of group one and two. Different reasons can be found for that. For example, the metrics of Allen et al. try to quantify the arbitrariness of patterns among the nodes of process models. However, they do not incorporate syntactical limitations, e.g. that nodes cannot be (directly) connected via two different arcs. Further, the models of group one and group two differ in syntactical errors but they are mostly equivalent structurally. The syntax, however, is ignored by many metrics, e.g. the conceptual coupling, CBO and RFC. On the other hand, group 3 differs regarding its decomposition structure which influences the metric values. In conclusion it may be said that the differences are too subtle, and the models are too small therefore resulting in little differences of the values, too. Regarding the perspective of coupling, the differences between the models are also small again resulting in small differences among the metric values. The lack of difference in the models also matches with the results of a laboratory experiment conducted with 66 students at a German university. They were asked to rate the understandability of the models on a 7-point Likert scale. The results indicated no significant difference in the understandability for all three model groups. However, the relation between understandability and coupling needs further investigation.

5 Outlook and Conclusion

The paper at hand deals with coupling metrics from software engineering. In that field coupling is a well-established concept for judging the quality of information systems. In recent years work has been done transferring the idea of coupling to business process modeling. The main motivation is to assess the fuzziness of the process model quality discipline by the quality dimension “coupling”. Though only a few metrics were introduced for judging a process model regarding coupling (see e.g. [29-31]), there are different perspectives on coupling in software engineering (see section 2.2). These led to a considerable set of different metrics for coupling in software engineering.

When these metrics are transferred to business process modeling they cannot only be used to evaluate business process models but also to infer suggestions for a good process modeling style regarding coupling. The paper at hand thus contributes to this field by transferring corresponding coupling metrics to business process modeling and shows their applicability on an example.

The metrics of Allen et al. generally suggest using repetitive patterns in the structure of nodes and arcs since they are more comprehensible. The actual implementation of the metric, however, merely suggests limiting the size of each model, though no concrete limit is given. The procedures used for the CBO and RFC metrics are easier. CBO counts the number of models one model is connected with by its control flow. Consequently, the metrics suggests limiting the number of connections between models. The RFC further incorporates the size of a model. Therefore, the metric suggests creating models using a low number of functions and connections to other mod-

els. The conceptual coupling of Poshyvanyk suggests to isolate concepts in separate models and to use a distinctive vocabulary. The last metric, Process coupling of Reijers/Vanderfeesten turns out to a good value with only few functions being connected to the same information element. However, one should bear in mind that the modeler's freedom of including or omitting elements should not lead to omitting e.g. information elements solely to realize a good metric value, even though this information would be necessary for the model users.

Still, there is clearly further work to be done. The transfer procedure was influenced by subjectivity in the interpretation of equivalent constructs. This also led to alternative interpretations that need to be discussed. Further, though a range of metrics was presented, the empirical evaluation is still missing. Therefore the practical utility of these metrics remains unanswered. Further, the laborious calculation procedures should be implemented in tool support for a practical use. Furthermore, additional metrics should be searched for and the existing ones should be transferred to more modeling languages (e.g. BPMN).

References

1. Becker, J., Thome, I., Weiß, B., Winkelmann, A.: Constructing a Semantic Business Process Modelling Language for the Banking Sector. *Enterprise Modelling and Information Systems Architectures* 5, 4-25 (2010)
2. Mendling, J., Reijers, H., van der Aalst, W.: Seven process modeling guidelines. *Information and Software Technology* 52, 127-136 (2010)
3. Moody, D.L.: Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data & Knowledge Eng.* 55, 243-276 (2005)
4. Schütte, R., Rothhove, T.: *The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models*. LNCS, Vol. 1507, pp. 240-254. Springer (1998)
5. Kobler, M.: *Qualität von Prozessmodellen - Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung*. Logos, Berlin (2010)
6. Mili, H., Tremblay, G., Jaoude, G.B., Lefebvre, É., Elabed, L., Boussaidi, G.E.: Business process modeling languages: Sorting through the alphabet soup. *ACM Computing Surveys* 43, 1-54 (2010)
7. Aguilar-Savén, R.S.: Business process modelling: Review and framework. *International Journal of Production Economics* 90, 129-149 (2004)
8. Bobrik, R., Reichert, M., Bauer, T.: *View-Based Process Visualization*. LNCS, Vol. 4714, pp. 88-95. Springer (2007)
9. Polyvyanyy, A., Smirnov, S., Weske, M.: Process Model Abstraction: A Slider Approach. In: *Enterprise Distributed Object Computing Conference*, pp. 325-331. IEEE, Munich (2008)
10. Gruhn, V., Laue, R.: Adopting the Cognitive Complexity Measure for Business Process Models. In: Yao, Y., Shi, Z., Wang, Y., Kinsner, W. (eds.): *International Conference on Cognitive Informatics*, pp. 236 -241. IEEE (2006)
11. Becker, J., Rosemann, M., von Uthmann, C.: *Guidelines of Business Process Modeling*. LNCS 1806, pp. 241-262. Springer (2000)
12. Krogstie, J., Sindre, G., Jørgensen, H.: Process models representing knowledge for action. *European Journal of Information Systems* 15, 91-102 (2006)

13. Lindland, O.I., Sindre, G., Sjølvberg, A.: Understanding Quality in Conceptual Modeling. *IEEE Software* 11, 42-49 (1994)
14. Recker, J., Indulska, M., Rosemann, M., Green, P.: Do Process Modelling Techniques Get Better? A Comparative Ontological Analysis of BPMN. In: 16th Australasian Conference on Information Systems, Sydney (2005)
15. Recker, J., Rosemann, M., Green, P., Indulska, M.: Do Ontological Deficiencies In Modeling Grammars Matter?. *MIS Quarterly* 35, 57-79 (2011)
16. Agarwal, S., Teas, R.K.: Cross-national applicability of a perceived quality model. *Journal of Product & Brand Management* 11, 213-236 (2002)
17. Remme, M.: Konstruktion von Geschäftsprozessen - Ein modellgestützter Ansatz durch Montage generischer Prozesspartikel. Gabler, Wiesbaden (1997)
18. Süttenbach, R., Ebert, J.: A Booch Metamodel. Universität Koblenz-Landau (1997)
19. Frank, U., van Laak, B.: Anforderungen an Sprachen zur Modellierung von Geschäftsprozessen. Universität Koblenz-Landau (2003)
20. Prasse, M.: Evaluation of object-oriented modelling languages: A comparison between OML and UML. In: Schader, M., Korthaus, A. (eds.): *The Unified Modeling Language – Technical Aspects and Applications*. Physica, Heidelberg (1998)
21. Zelewski, S.: Eignung von Petrinetzen für die Modellierung komplexer Realsysteme - Beurteilungskriterien. *Wirtschaftsinformatik* 38, 369–381 (1996)
22. Bridges, G.: *Top Ten Tips and Tricks For Business Process Modeling* (2009)
23. Silver, B.: *Ten tips for effective process modeling*. BPMInstitute.org (2008)
24. Cardoso, J.: *Business Process Quality Metrics: Log-Based Complexity of Workflow Patterns*. LNCS, Vol. 4803, pp. 427-434. Springer (2007)
25. Gruhn, V., Laue, R.: Approaches for Business Process Model Complexity Metrics. In: Abramowicz, W., Mayr, H.C. (eds.): *Technologies for Business Information Systems*. Springer, Berlin (2007)
26. Latva-Koivisto, A.: *Finding a complexity measure for business process models*. Helsinki University of Technology (2001)
27. Mendling, J.: *Testing Density as a Complexity Metric for EPCs*. Vienna University of Economics and Business Administration (2006)
28. Mendling, J., Reijers, H.A., Cardoso, J.: *What Makes Process Models Understandable?* LNCS, Vol. 4714, pp. 48-63. Springer (2007)
29. Vanderfeesten, I., Cardoso, J., Reijers, H.A.: *A weighted coupling metric for business process models*. In: Eder, J., Tomassen, S.L., Opdahl, A., Sindre, G. (eds.): *CAiSE 2007*, pp. 41-44, Trondheim, Norway (2007)
30. Vanderfeesten, I., Reijers, H.A., Mendling, J., van der Aalst, W.M.P., Cardoso, J.: *On a Quest for Good Process Models: The Cross-Connectivity Metric*. In: Bellahsene, Z., Lénoard, M. (eds.): *CAiSE 2008*, pp. 480-494, Montpellier, France (2008)
31. Vanderfeesten, I.T.P., Cardoso, J., Mendling, J., Reijers, H.A., van der Aalst, W.M.P.: *Quality Metrics for Business Process Models*. In: Fischer, L. (ed.): *BPM and workflow handbook. Future Strategies*, Lighthouse Point, USA (2007)
32. Weber, R.: *Ontological foundations of information systems*. Coopers & Lybrand, Melbourne (1997)
33. Recker, J., Rosemann, M., Green, P., Indulska, M.: *Do ontological deficiencies in modeling grammars matter?*. *MIS Quarterly* 35 (1), 57-79 (2011)
34. Scheer, A.-W.: *ARIS - business process modeling*. Springer, Berlin et al. (2000)
35. Mendling, J.: *Detection and prediction of errors in EPC Business Process Models*. Institute für Informationssysteme und neue Medien. WU Wien (2007)

36. Mendling, J.: Metrics for process models: Empirical foundations of verification, error prediction, and guidelines for correctness. Springer, Berlin (2008)
37. van Hee, K.M., Oanea, O., Sidorova, N.: Colored Petri Nets to Verify Extended Event-Driven Process Chains. In: Meersman, R., Tari, Z., Hacid, M.-S., Mylopoulos, J., Pernici, B., Babaoglu, Ö., Jacobsen, H.-A., Loyall, J.P., Kifer, M., Spaccapietra, S. (eds.): On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE. LNCS, Vol. 3760, pp. 183–201. Springer, Berlin Heidelberg (2005)
38. Keller, G., Nüttgens, M., Scheer, A.-W.: Semantische Prozeßmodellierung auf Grundlage Ereignisgesteuerter Prozeßketten (EPK). In: Scheer, A.-W. (ed.): Veröffentlichungen des Instituts für Wirtschaftsinformatik, Vol. 89, Saarbrücken (1992)
39. Gove, P.B.: Webster's third international dictionary of the English language. Merriam-Webster, Springfield (1986)
40. Wand, Y., Weber, R.: An ontological model of an information systems. IEEE Transactions on Software Engineering 16, 1282–1292 (1990)
41. Wand, Y., Weber, R.: Mario Bunge's ontology as a formal foundation for information systems concepts. In: Weingartner, P., Dorn, G.W.J. (eds.): Studies on Mario Bunge's Treatise. Rodopi, Atlanta (1990)
42. Chidamber, S.R., Kemerer, C.F.: A Metrics Suite for Object Oriented Design. IEEE Transactions on Software Engineering 20, 476–493 (1994)
43. McCabe, T.J.: A Complexity Measure. IEEE Transactions on Software Engineering 2, 308–320 (1976)
44. Shannon, C.E.: A mathematical theory of communication. The bell system technical journal 27, 379–423 (1948)
45. Halstead, M.H.: Elements of software science. Elsevier, New York (1977)
46. Allen, E.B., Khoshgoftaar, T.M., Chen, Y.: Measuring coupling and cohesion: An information-theory approach. In: 6th International Software Metrics Symposium, pp. 119–127. IEEE Computer Society (1999)
47. Allen, E.B., Khoshgoftaar, T.M., Chen, Y.: Measuring Coupling and Cohesion of Software Modules: An Information-Theory Approach. In: 7th International Software Metrics Symposium. IEEE (2001)
48. Allier, S., Vaucher, S., Dufour, B., Sahraoui, H.: Deriving Coupling Metrics from Call Graphs. In: 10th IEEE Working Conference on Source Code Analysis and Manipulation, pp. 43–52. IEEE (2010)
49. Arshad, F., Khanna, G., Laguna, I., Bagchi, S.: Distributed Diagnosis of Failures in a Three Tier E-Commerce System. ECE Technical Reports. Purdue University (2007)
50. Cho, E.S., Kim, C.J., Kim, S.D., Rhew, S.Y.: Static and Dynamic Metrics for Effective Object Clustering. In: 5th Asia-Pacific Software Engineering Conference, pp. 87–85. IEEE Computer Society (1998)
51. Gui, G., Scott, P.D.: Coupling and cohesion measures for evaluation of component reusability. In: Diehl, S., Gall, H., Hassan, A.E. (eds.): Proceedings of the 2006 International Workshop on Mining Software Repositories, pp. 18–21. ACM Press, New York (2006)
52. Henry, S., Kafura, D.: Software Structure Metrics Based on Information Flow. IEEE Transactions on Software Engineering 7, 510–518 (1981)
53. Cardoso, J.: Process control-flow complexity metric: An empirical validation. In: IEEE International Conference on Services Computing (SCC'06), pp. 167–173 (2006)
54. Hitz, M., Montazeri, B.: Measuring coupling and cohesion in object-oriented systems. In: Proceedings of 3rd International Symposium on Applied Corporate Computing (1995)

55. Wahler, K., Küster, J.M.: Predicting Coupling of Object-Centric Business Process Implementations. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.): Business process management. LNCS, Vol. 5240, pp. 148–163. Springer, Berlin (2008)
56. Reijers, H.A., Vanderfeesten, I.T.P.: Cohesion and coupling metrics for workflow process design. In: Desel, J., Pernici, B., Weske, M. (eds.): Business process management. LNCS, Vol. 3080, pp. 290–305. Springer, Berlin (2004)
57. Kazemi, A., Azizkandi, A.N., Rostampour, A., Haghighi, H., Jamshidi, P., Shams, F.: Measuring the Conceptual Coupling of Services Using Latent Semantic Indexing. In: Jacobsen, H.-A., Wang, Y., Hung, P. (eds.): IEEE International Conference on Services Computing, pp. 504–511. IEEE Computer Society (2011)
58. Poshyvanyk, D., Marcus, A.: The Conceptual Coupling Metrics for Object-Oriented Systems. In: 22nd IEEE International Conference on Software Maintenance, pp. 469–478. IEEE Computer Society (2006)
59. Quynh, P.T., Thang, H.Q.: Dynamic coupling metrics for service-oriented software. *International Journal on Computer Science Engineering* 3 (5), 282–287 (2009)
60. Qian, K., Liu, J., Tsui, F.: Decoupling Metrics for Services Composition. In: Lee, R., Ishii, N. (eds.): Proceedings of the 5th Annual IEEE/ACIS International Conference on Computer and Information Science (ICIS 2006) and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (COMSAR 2006), pp. 44–47. IEEE Computer Society (2006)
61. Rajaraman, C., Lyu, M.R.: Reliability and maintainability related software coupling metrics in C++ programs. In: 3rd international symposium on software reliability engineering, pp. 303–311. IEEE Computer Society (1992)

Information Systems Success - A Quantitative Literature Review and Comparison

Sebastian Dörr, Sebastian Walther, and Torsten Eymann

University of Bayreuth, Chair of Information Systems Management, Bayreuth, Germany
seb.doerr@gmx.de,
{s.walther,torsten.eymann}@uni-bayreuth.de

Abstract. Despite the outstanding economic relevance of information systems in business and society, there exists no consensus among academics and practitioners how to measure the success of information systems. Therefore this paper makes a contribution to structure and classify the most recent research in IS success. This is done by providing an exhaustive overview of literature in the field of IS success capturing the empirical and non-empirical publications between 2007 and 2011. With regard to theoretical foundation, object of analysis, unit of analysis, evaluation perspective, data gathering method, and data analysis, 26 empirical articles are classified. The 11 non-empirical studies are classified by theoretical foundation, methodological type and object of analysis. The results show that in most cases a type of IT or IT application is investigated by applying the DeLone and McLean IS success model. Most of the studies focus on IS users and the individual perspective of analysis, whereas the most prominent data analysis method is structural equation modeling.

Keywords: Information Systems Success, Literature Review

1 Introduction

Information is the basis for economic decisions within the whole value chain, making enterprises dependent on the implementation of modern information systems (IS) to stay competitive [1], e.g. by enabling real-time data access or providing business intelligence functions. Simultaneously, the amount of business realms using sophisticated IS rises: among others, IS comprises e-commerce systems, knowledge management systems, and decision support systems.

In this context, measuring what makes an IS successful is of utmost importance. However, no consensus among practitioners and academics exists, how to measure the success of IS. Therefore, many success models have been developed, complicating the validation and comparison of the antecedents of IS success. Previous research on IS success has found three models to be predominant: the IS success model proposed by DeLone and McLean (D&M success model) [17], which is the most widely used IS success model [65], the updated D&M success model [18], as well as the

Technology Acceptance Model (TAM) [16], which was generally applied in the context of IS adoption.

This paper summarizes the current state of research of IS success by employing a structured literature review according to Webster and Watson [69] with focus on multidimensional IS success models. The review is limited to the years between 2007 and 2011 and includes empirical, as well as non-empirical articles. Chronologically, but also content-related, it continues the literature review provided by Urbach et al. [65].

Empirical articles are analyzed according to theoretical foundation, object of analysis, unit of analysis, evaluation perspective, data gathering, and data analysis. Theoretical articles are classified according to theoretical foundation, object of analysis, and methodological type.

Our paper is built as follows. First of all, the theoretical foundations are introduced. Secondly, the methodology of the structured review is described, including literature selection and framework analysis. The results are then divided into “results of empirical literature” and “results of non-empirical literature”. Finally, the interesting findings and limitations are discussed, whereas the results of our literature analysis are compared to the results of Urbach et al. [65].

2 Theoretical Foundation

During the first International Conference on Information Systems in 1980, Keen highlighted five aspects which are relevant for establishing Management Information Systems (MIS) as a particular field of research. Among them was the quest for the dependent variable to measure IS success.

Measuring IS success poses a challenge to researchers, because its definition varies depending on the perspective of evaluation [65]. Therefore, a multidimensional IS success model is necessary to capture all stakeholder’s perspectives [17], [65]. Furthermore, it is necessary to isolate IS success to enable its explicit attribution to the object of study. In praxis this is frequently not possible or avoided due to the increased effort [17]. Especially the D&M success model [17], the updated D&M success model [18], and TAM [16] are of high importance for this review. Urbach et al. [65] additionally mention the Seddon success model [63]. However, the Seddon success model was found to be irrelevant for the following study.

2.1 DeLone and McLean Success Model

The D&M success model consists of six interdependent variables which are theoretically connected (see Figure 1).

The following definitions are according to [17]. *System Quality* measures the quality of the information processing within the system. The IS output is measured by *Information Quality*. *Use* is seen as the demand or consumption of IS output. *User Satisfaction* describes the reaction of the recipient to the use of the IS output. The impact of information on user / receiver behavior is measured by *Individual Impact*.

Finally, *Organizational Impact* describes the influence of information on overall organizational success.

DeLone and McLean commented on the model: “This success model clearly needs further development and validation before it could serve as a basis for the selection of appropriate I/S measures” [17]. Referring to this limitation, IS success research conducted meta-analyses (e.g. [7], [53]) and standardized measuring methods to validate the IS Success Model (e.g. [60], [47]). Seddon [57] criticizes DeLone and McLean’s attempt to comprise too much in one model. That is why Seddon describes it as confusing and erroneously specified [57]. The IS Success Model gathers success using a depiction of process but also causal factors [17]. Furthermore, Seddon problematizes the ambiguity of *Use*, “Meaning 1: Use as a Variable in a Variance Model of Future IS Use”; “Meaning 2: IS Use as the Dependent Variable in a Variance Model of Future IS Use”; “Meaning 3: IS Use as an Event in a Process Leading to Individual or Organizational Impact” [57]. Seddon clarifies the meaning of *Use* and introduces four new variables (*Expectations*, *Consequences*, *Perceived Usefulness* and *Net Benefits to Society*) [57]. Additionally, a classification of the variables in “Measures of Information and System System Quality”, “General Perceptual Measures of Net Benefits of IS Use”, and “Behavior with Respect to IS Use” are considered [57].

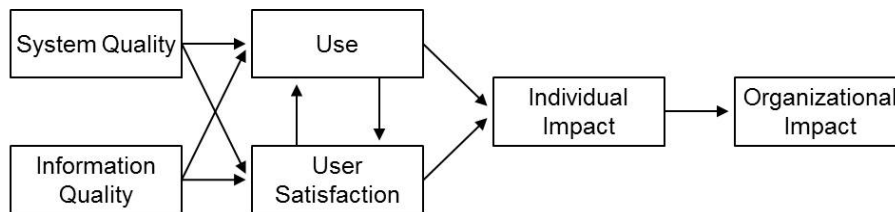


Fig. 1. D&M Success Model

2.2 Updated DeLone and McLean Success Model

In their 10-year-update, DeLone and McLean discussed which hypotheses had been found to be significant [18]. In descending order, these are: *System Use – Individual Impacts*; *System Quality – Individual Impacts*; *Information Quality – Individual Impacts* [18]. With one exception (*System Use – Organizational Revenues*), the other interdependencies have been confirmed as well [18]. DeLone and McLean reject the Seddon’s criticism about *Use* being no success factor [18]. Instead, the difficulty is to be seen in the complexity of the *Use* variable and therefore a missing, simple definition (DeLone and McLean 2003, [16]). Especially e-commerce, where system use by costumers is essential, clarifies the importance of *Use* [14], [31], [45].

Besides *Organizational Impact* and *Individual Impact*, further entities could be affected by IS activities. Therefore, researchers suggested considering *Group Impacts* (e.g. [30], [41]), *Inter-organizational and Industry Impacts* (e.g. [12-13]), *Consumer Impacts* (e.g. [8], [27]), and *Society Impacts* (e.g. [57], [18]). Instead of a model extension, DeLone and McLean decided to consolidate all impacts as *Net Benefits* [18]. As a quid pro quo, this generalization requires a defined frame of reference (e.g.

sponsor, user, stock holder) [18]. Despite the cancellation of *Individual Impact* and *Organizational Impact*, the analysis perspective must still be mentioned (e.g. *Individual Perspective*, *Industrial Perspective*) [18].

Pitt et al. [48] criticized that IS success is focused on products but not on services. Therefore, the system characteristics were extended by *Service Quality* [18]. Figure 2 illustrates the updated D&M success model.

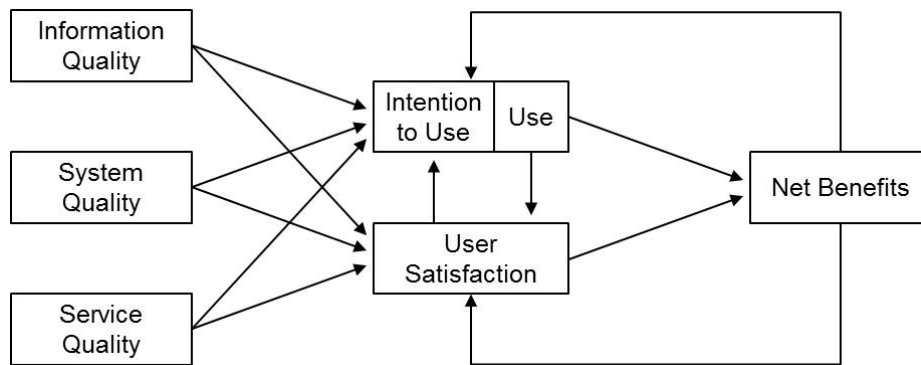


Fig. 2. Updated D&M Success Model

2.3 Technology Acceptance Model

Missing user acceptance often impairs the business' performance [16]. Therefore, Davis developed a model to predict user acceptance. TAM was subject of his dissertation [15] and was published in MIS Quarterly in 1989 [16]. The model is mainly based on the Theory of Reasoned Action (TRA) ([15], [19]).

Davis concentrates his research on the measurement of key constructs [16]. He proposes forecasting *Use (Intention)* by focusing on two theoretical constructs: *Perceived Usefulness* and *Perceived Ease of Use* [16]. According to Davis, users are more willing to use a system or program, if they see an occupational advantage [16]. This variable is measured by *Perceived Usefulness* [16]. Similar to the TRA, the TAM plans to predict behavior. If the system is perceived of being useful, there is still a danger that the system is conceived to be too difficult or complicated [16]. The use would not outweigh the effort. This consideration is measured by *Perceived Ease of Use* [16]. Figure 3 displays the TAM.

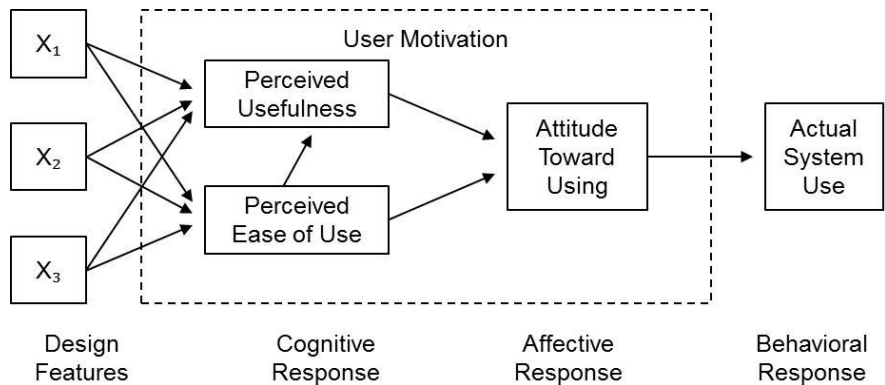


Fig. 3. Technology Acceptance Model

3 Methodology

A well-designed literature review is the foundation for scientific insights [69]. An analysis should simplify the development of theories and models and conclude research areas, where a plethora of research exists, and disclose areas, where research is needed [69]. Due to the rising amount of publications the need to describe, summarize, validate, and categorize the results increases [65].

According to the publication manual of the American Psychology Association, a literature review defines and clarifies the problem, summarizes previous research with intent to convey the current state of research, identifies relationships, contradictions and gaps of relevant literature, and gives further problem solving suggestions [3]. In the following, the steps conducted within the literature review are described in detail.

3.1 Literature Selection

The literature selection is essential for the quality of the analysis. A structured review should cover the relevant articles as exhaustively as possible. In the following, the literature selection is conducted in a three step approach: timeframe definition, source limitation, and selection criteria.

First of all, the timeframe was set. DeLone and McLean reviewed the years between 1981 and 1988, as well as 1992 and 2002 concerning IS success factors ([17], [18]). Urbach et al. analyzed the publications between 2003 and 2007 [65]. This review continues Urbach et al.'s analysis by observing the years 2007 to 2011¹.

To acquire a source selection that is as comprehensive as possible, both, the journals analyzed by DeLone and McLean ([17-18]), and the leading IS journals were taken into consideration. In accordance to Webster and Watson, essential articles are mainly published in leading journals [69]. The literature selection is consistent with the selection by Urbach et al. [65], who used Saunderson's MIS journal-ranking [65].

¹ The reviews overlap in the year 2007.

Journals with a rating value up to 30 were classified as relevant. The ranking by Saunders is a meta-analysis of nine independent journal rankings. Journals, which were ranked by only one of the nine rankings, were excluded for lacking representativeness. Further, some Journals were excluded because of subject specifications [65]. Besides journals, leading conference proceedings were taken into account to consider recent research [65]. Books are not reflected. It is assumed that authors already published their results in journals [65]. In addition to that, books are mostly not subjected to a formal review process [65]. All in all, 34 journals and four conference proceedings are reviewed. A list of the sources is attached to this article.

The sources were examined for topic-related articles. Practically, electronic databases (EBSCOhost, ScienceDirect, and ProQuest) and conference proceedings were searched for the keywords “information systems success”, “IS success”, information systems efficiency” and “IS efficiency”. The query regarded titles, abstracts, and keywords. In total, 78 publications were recorded (39 journal articles, 39 conference articles).

3.2 Review Framework

Intending to continue the review by Urbach et al. [65], the framework was transferred into this study. The framework consists of eight categories: (1) Theoretical Foundation; (2) Research Approach; (3) Object of Analysis; (4) Unit of Analysis; (5) Evaluation Perspective; (6) Data Gathering; (7) Data Analysis; (8) Methodological Type [65].

All publications are sorted by *Theoretical Foundation*. However, only the introduced models are mentioned ([16-18]). Publications using distinct models were classified as “others”. In case that the articles had no underlying models, the articles were labeled as “n/a” (not applicable).

The category *Research Approach* divides articles in empirical and non-empirical approaches. Publications are empirical, if they are based on observations and use empirical methodology of some kind (e.g. survey, interview, laboratory experiment, case study) [2]. Non-empirical studies on the other hand usually are based on ideas, speculations, literature reviews, or frameworks [65]. King and He observed a bias to the review of empirical studies [35]. Therefore, this review also considers non-empirical articles. However, the chosen categories tend to focus on empirical publications [65].

Object of Analysis classifies the articles by type of system and exemplifies the focus of the reviewed publication [65]. Seddon distinguishes between: (1) an aspect of IT use (e.g. single algorithm); (2) a single IT application (e.g. a certain data warehouse); (3) a type of IT or IT application (e.g. knowledge management systems); (4) all IT applications used by an organization or sub-organization [57].

The *Unit of Analysis* can be divided into a micro and a macro point of view [65]. According to Grover et al., for the sake of completeness, both perspectives should be considered for the evaluation of IS success [24]. A micro-perspective of IS success shows, whether the individual needs of co-workers are satisfied [65]. A macro-

perspective reflects the IS' impacts on the organization's competitiveness [65]. Therefore, this review differentiates between an individual and an organizational level.

Stakeholders represent different interests. Thus, there is a possibility that different groups of interest evaluate IS success differently [57]. The *Evaluation Perspective* therefore specifies, which group of interest was surveyed within a particular study and respectively which point of view was represented for IS success measurement [65]. Grover et al. differentiate four perspectives: users, top management, IS personnel, and external entities (e.g. subcontractor, costumers) [24]. Urbach et al. added two additional points of view to allow better differentiation: IS executives and multiple stakeholders [65]. The unit of analysis is independent from the evaluation perspective. All standpoints can be evaluated at an organizational and / or individual level [65].

Data Gathering categorizes empirical methods, which were applied within the particular study [65]. The analysis of data gathering allows statements about reliance and generalizability of study results [65]. Urbach et al. differentiate four "dominant" methods: survey, interview, case study, and laboratory experiment [65]. Further methods are declared as "others" [65].

Data Analysis distinguishes the approaches of data evaluation [65]. According to Urbach et al. for IS research the following analyses are used commonly: structural equation modeling (e.g. linear structural relationships (LISREL), partial least squares (PLS)), regression analysis, factor analysis, and cluster analysis [65]. Further methods (e.g. qualitative analysis) are summarized as "other", studies not using data analysis as "n/a" [65].

Non-empirical studies are also categorized by *Methodological Type*. Palvia et al. differentiate three non-empirical methodological types: framework / conceptual model, speculation / commentary, and library research [46]. Further methods are cumulated as "other".

The literature-pool is classified using this framework. However, certain articles do not mention needed information explicitly [65]. In these cases, the data is based on the author's interpretation [65]. A graphical display of the framework is attached to the appendix.

4 Results

By researching online databases (EBSCOhost, ScienceDirect, ProQuest) and conference proceedings, 78 articles were identified of which 20 journal articles and 21 conference articles were declared as irrelevant. Publications were sorted out when they only covered one-dimensional IS models or did not match the topic of IS success: e.g. the search items also include "IS Outsourcing Success" and "IS Planning Success Factors". The remaining articles were analyzed by referring to the introduced framework. The analysis is divided into two blocks. Firstly, the analysis of empirical literature will be shown. Secondly, the results of the non-empirical literature will be discussed.

4.1 Analysis of Empirical Literature

This review focuses on empirical literature analysis. Therefore, the framework concentrates on empirical literature analysis as well. It is crucial to point out *what* was measured in the particular study [65].

The literature classification under theoretical foundation shows a strong representation of the IS Success Model, especially the updated IS Success Model by DeLone and McLean. Nearly two-thirds of the studies under analysis use this model. Furthermore, 17 studies are to be classified as “other”. It is noticeable that different theoretical foundations are often combined or adjusted. Just 10 of 26 studies are based on a single model.

Regarding the unit of analysis, all studies considered an individual level. Beyond that, nine studies also take account of an organizational level.

The evaluation of object of analysis reveals that in most cases types of IT or IT applications are observed (16 of 26 studies). Eight studies examined a single IT application. Only Leidner et al. examined IT applications in general. The IT function of an organization was the object of study in Santos et al.’s survey. Therefore, just a few universal results can be observed in the timeframe of research.

The perspective of evaluation corresponds with unit of analysis. 18 of 26 studies surveyed IS users. The opinion of users usually correlates with individual success factors. Only Gorla et al. and Lee et al. interviewed users on an organizational level. Three authors questioned IS personnel to determinate IS success. Ifinedo and Leidner et al. addressed IS executives. Multiple stakeholders are only regarded by Gable et al. and Bartis and Mitev.

Except for Bartis and Mitev (case study) every study uses surveys for data gathering. Lee et al. additionally conducted interviews.

For data analysis, structural equation modeling is most common (20 of 26 studies). In nine cases, it is combined with a variance analysis, which is used by 11 studies. Five studies combine it with a factor analysis, whereas nine factor analyses are conducted in total. Of the regarded methods, regression analysis is used least often (five of 25 studies). Teo et al. and Bartis and Mitev use “other” methods. Seven studies used one method of analysis. 14 studies combined two methods. Chiu et al., Gable et al. and Teo et al. used three or more methods.

Figure 4 and figure 5 summarize the result for empirical literature. All in all, the most evaluations are based on a type of IT or IT application and are measuring on an individual level by surveying users. As the theoretical foundation, the IS Success Model is most common, whereas the data is mostly analyzed using structural equation modeling. The complete research data is added to the attachment.

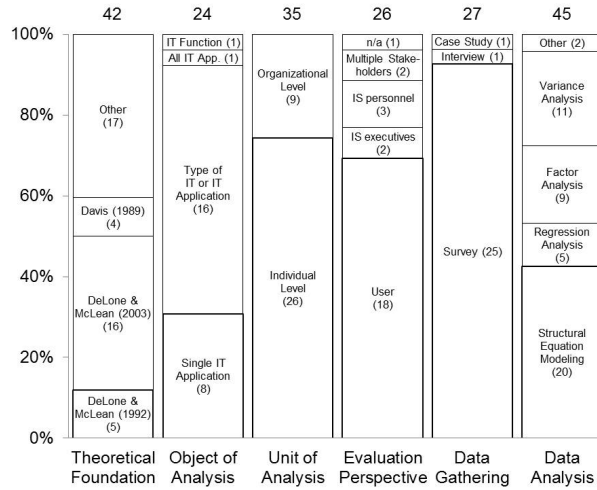


Fig. 4. Classification of Empirical Publications (1)

	Theoretical Background	Object of Analysis	Unit of Analysis	Evaluation Perspective	Data Gathering	Data Analysis
	D&M Success Model Update of D&M Success Model TAM Others	Single IT Application Type of IT or IT Application All IT Applications IT Function of an Organization	Individual Level Organizational Level	User IS Executive IS Personnel Multiple Stakeholders n/a	Survey Interview Case Study	Structural Equation Modeling Regression Analysis Factor Analysis Variance Analysis Others
Bartis & Mitev [4]	x	x	x x	x	x	x
Bernroider [5]	x	x	x x	x	x	x x
Bock et al. [6]	x	x	x x	x	x	x x
Chiu et al. [11]	x x	x	x	x	x	x x x
Floropoulos [20]	x x	x	x	x	x	x x
Gable et al. [22]	x x	x	x x	x	x	x x x x
Gorla et al. [23]	x x	x	x x	x	x	x x
Halawi et al. [25]	x	x	x	x	x	x x
Hong et al. [28]	x	x	x	x	x	x x
Ifinedo [29]	x x	x	x x	x	x	x
Jin and Kim [32]	x	x	x	x	x	x
Khayun & Racham [34]	x x x	x	x	x	x	x
Lee et al. [36]	x x	x	x x	x	x x	x x
Leidner et al. [37]	x	x	x x	x	x	x x
Ng et al. [42]	x x	x x	x	x	x	x
Pak et al. [44]	x x	x	x	x	x	x x
Polancic et al. [49]	x	x	x	x	x	x
Prybutok et al. [50]	x x	x	x x	x	x	x
Saeed & Abdinour-Helm [54]	x x x	x	x	x	x	x x
Santos et al. [55]	x	x	x	x	x	x
Schaupp et al. [56]	x x	x	x	x	x	x x
Song [63]	x x	x	x	x	x	x
Teo et al. [64]	x x	x	x	x	x	x x x
Urbach et al. [66]	x	x	x	x	x	x x
Wang [68]	x	x	x	x	x	x x
Xu et al. [70]	x	x	x	x	x	x x

Fig. 5. Classification of Empirical Publications (2)

4.2 Analysis of Non-Empirical Literature

The non-empirical literature is categorized into theoretical foundation as well (see figure 6). One literature analysis can be based on multiple theoretical foundations. Again, DeLone and McLean are mostly mentioned as theoretical foundation. Six articles use models which are declared as “others”. Three authors do not specify a theoretical foundation.

When categorized by methodology, the publications can be sorted into six frameworks / conceptual models and five literature analyses. Speculations, commentaries, or other methods were not included in the literature pool.

Similar to the empirical literature, the non-empirical literature is classified by object of analysis. Nearly half the articles do not point out a specific object of analysis. Three publications refer to a type of IT or IT application. The remaining three publications contain all IT applications.

	Theoretical Background				Methodology	Object of Analysis
	D&M Success Model	Updated D&M Success Model	Others	n/a	Framework/ Conceptual Literature Analysis	Type of IT or IT Application All IT Application n/a
Chae [10]	x	x	x	x	x	x
Gable and Sedera [21]					x	x
Harris and Weistroffer [26]	x	x	x		x	x
Moeni and Lapointe [40]			x		x	x
Niemi and Pekkola [43]		x			x	x
Petter et al. [47]	x	x			x	x
Raeth et al. [51]			x		x	x
Sedera and Chian [59]			x		x	x
Seen et al. [61]			x		x	x

Fig. 6. Classification of Non-Empirical Publications

4.3 Findings

In the following, the results are compared to the review provided by Urbach et al. [65]. Both reviews confirm a domination of the D&M success. Nevertheless, the studies between 2003 and 2007 preferred the original D&M success model, whereas the studies between 2007 and 2011 mostly referred to the updated D&M success model [65]. This applies to the empirical and non-empirical literature.

In regard to the object of analysis, both reviews come to the conclusion that mainly type of IT or IT application are observed.

The examination of unit of analysis leads to a similar result as well: Urbach et al. confirmed 26 of 28 studies measuring success on an individual level [65]. 12 articles additionally measured success on an organizational level [65].

The results of the evaluation perspective resemble the analysis by Urbach et al.: Both reviews' literature selection interviewed users about IS success. In case of Urbach et al., 19 of 28 articles referred to this perspective [65].

Surveys are the main method for data gathering in both literature analyses.

Finally, there is a similarity between the methods of data analysis, showing the the results between the studies are consistent.

5 Conclusion

This literature review analyzed the state of research of multi-dimensional IS success measurements and models. Therefore, literature between 2007 and 2011 was evaluated. The literature pool of relevant sources consisted of 26 empirical studies and 11 non-empirical articles. The analysis focused on empirical literature, which was analyzed in depth.

The review clarifies that the D&M success models still enjoy huge popularity. By now, the majority has switched to the updated D&M success model published in 2003. The review also shows that different models are often combined to grant justice to the subject of research.

While every empirical article analyzed an individual level, only nine of 26 studies consider an organizational point of view. Studies including both perspectives ensure a more comprehensive success measurement, but also increase the effort of data gathering. Data is usually gathered from users by conducting surveys. In general, a type of IT or IT application is evaluated. The gathered data is mostly analyzed by structural equation modeling.

The limitations mentioned by Urbach et al. [65] are also applicable in this review: on the one hand, the sources are limited to chosen journals and conference articles. Thus, relevant articles may be excluded. Further, books were not considered, as it was assumed that articles of importance were also published in leading journals and conferences. On the other hand, the approach of search using databases could distort the result. Articles of relevance may be excluded if they do not match the criteria of the search requests (referring to title, abstract, keywords). Furthermore, the search requests contain a bias towards the D&M success model. Another limitation is caused by only conducting English queries.

References

1. Al-adaile, R.M.: An Evaluation of Information Systems Success: A User Perspective – the Case of Jordan Telecom Group. *European Journal of Scientific Research* 37 (2), 226-239 (2009)
2. Alavi, M., Carlson, P.: A review of MIS research and disciplinary development. *Journal of Management Information Systems* 8 (4), 45-62 (1992)
3. American Psychology Association: Publication manual of the American Psychological Association. (6th edition), American Psychology Association, Washington (2010)

4. Bartis, E., Mitev, N.: A multiple narrative approach to information systems failure: a successful system that failed. *European Journal of Information Systems* 17, 112-124 (2008)
5. Bernroider, E.W.N.: IT governance for enterprise resource planning supported by the DeLone–McLean model of information systems success. *Information & Management* 45, 257-269 (2008)
6. Bock, G.-W., Suh, A., Shin, K.-S., Hu, A.: The factors affecting success of knowledge-based systems at the organizational level. *Journal of Computer Information Systems* 50 (2), 95-105 (2009)
7. Bokhari, R.H.: The relationship between system usage and user satisfaction: a meta-analysis. *Journal of Enterprise Information Management* 18 (2), 211-234 (2005)
8. Brynjolfsson, E.: The Contribution of Information Technology to Consumer Welfare. *Information Systems Research* 7 (3), 281-300 (1996)
9. Buhl, H.U., Mertens, P., Schumann, M., Urbach, N., Smolnik, S., Riempp, G.: Leserbrief: Stellungnahme zum Beitrag von Urbach et al. aus Heft 4/2009. *Wirtschaftsinformatik* 52 (2), 109-114 (2010)
10. Chae, H.-C.: IS Success Model and Perceived IT Value. In: Hoxmeier, J.A., Hayne, S. (eds.): *Proceedings of the 13th Americas Conference on Information Systems*. Association for Information Systems (2007)
11. Chiu, C.-M., Chiu, C.-S., Chang, H.-C.: Examining the integrated influence of fairness and quality on learners' satisfaction and Web-based learning continuance intention. *Information Systems Journal* 17 (3), 271-287 (2007)
12. Clemons, E.K., Row, M.C.: Limits to interfirm coordination through information technology: Results of a field study in consumer goods packaging distribution. *Journal of Management Information Systems* 10 (1), 73-95 (1993)
13. Clemons, E.K., Reddi, S.P., Row, M.C.: The impact of information technology on the organization economic activity: The "move to the middle" hypothesis. *Journal of Management Information Systems* 10 (2), 9-35 (1993)
14. D'Ambra, J., Rice, R.E.: Emerging factors in user evaluation of the World Wide Web. *Information & Management* 38 (6), 373-384 (2001)
15. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results. Dissertation, Massachusetts Institute of Technology, Sloan School of Management (1986)
16. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13 (3), 319-340 (1989)
17. DeLone, W.H., McLean, E.R.: Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research* 3 (1), 60-95 (1992)
18. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems* 19 (4), 9-30 (2003)
19. Fishbein, M., Ajzen, I.: *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research*. Addison-Wesley, Reading, (1975)
20. Floropoulos, J., Spathis, C., Halvatzis, D., Tsiouridou, M.: Measuring the success of the Greek Taxation Information System. *International Journal of Information Management* 30 (1), 47-56 (2010)
21. Gable, G.G., Sedera, D.: Formative and Reflective Measurement and Validation Mismatch in Survey Research: An Archival Analysis of Information Systems Success Constructs 1985-2007. In: Nunamaker Jr., J.F.; Currie, W.L. (eds.): *Proceedings of the 30th International Conference on Information Systems*. AIS (2009)

22. Gable, G.G., Sedera, D., Chan, T.: Re-conceptualizing Information Systems Success: The IS-Impact Measurement Model. *Journal of the Association for Information Systems* 9 (7), 377-408 (2008)
23. Gorla, N., Somers, T.M., Wong, B.: Organizational impact of system quality, information quality, and service quality. *Journal of Strategic Information Systems* 19 (3), 207-228 (2010)
24. Grover, V., Jeong, S.R., Segars, A.H.: Information systems effectiveness: The construct, space and patterns of application. *Information & Management* 31 (4), 177-191 (1996)
25. Halawi, L.A., McCarthy, R.V., Aronson, J.E.: An Empirical Investigation of Knowledge Management Systems' Success. *Journal of Computer Information Systems* 48 (2), 121-135 (2007)
26. Harris, M.A., Weistroffer, H.R.: A New Look at the Relationship between User Involvement in Systems Development and Systems Success. *Communications of the Association for Information Systems* 24 (1), 739-756 (2009)
27. Hitt, L.M., Brynjolfsson, E.: The Three Faces of IT Value: Theory and Evidence. In: DeGross, J.I., Huff, S.L., Munro, M. (eds.): *Proceedings of the 15th International Conference on Information Systems*, pp. 263-277. Association for Information Systems (1994)
28. Hong, S., Kim, J., Lee, H.: Antecedents of use-continuance in information systems: Toward an integrative view. *Journal of Information Systems* 48 (3), 61-73 (2008)
29. Ifinedo, P.: Examining the influences of external expertise and in-house computer/IT knowledge on ERP system success. *The Journal of Systems and Software* 84 (12), 2065-2078 (2011)
30. Ishman, M.: Measuring information system success at the individual level in cross-cultural environments. In: Garrity, E.J., Sanders, G.L. (eds.): *Information System Success Measurement*. Idea Group, Hershey (1998)
31. Ives, B., Olsen, M., Baroudi, J.J.: The measurement of user information satisfaction. *Communications of the ACM* 26 (10), 785-793 (1983)
32. Jin, S.H., Kim, Y.J.: An Empirical Investigation into the Factors Influencing the Use of E-Banking Services. In: *Proceedings of the 43rd Hawaii International Conference on Systems Science*. IEEE Computer Society (2010)
33. Kettinger, W.J., Lee, C.C.: Perceived Service Quality and User Satisfaction with the Information Services Function. *Decision Sciences* 25 (5-6), 737-766 (1994)
34. Khayun, V., Ractham, P.: Measuring e-Excise Tax Success Factors: Applying the De-Lone & McLean Information Systems Success Model. In: *Proceedings of the 44th Hawaii International Conference on Systems Science*. IEEE Computer Society (2011)
35. King, W.R., He, J.: Understanding the Role and Methods of Meta-Analysis in IS Research. *Communications of the Association for Information Systems* 16, 656-686 (2005)
36. Lee, C.S., Ko, I.S., Jeong, C.: Evaluating the Effectiveness of Information Service for SMEs on Information Orientation and Firm Performance. In: *Proceedings of the 42nd Hawaii International Conference on Systems Science*. IEEE Computer Society (2009)
37. Leidner, D.E., Lo, J., Gonzales, E.: An empirical investigation of IS strategy and IS contribution to firm performance. In: Sabherwal, R., Sumner, M. (eds.): *Proceedings of the 31th Conference on Information Systems*. Association for Information Systems (2010)
38. Li, E.Y.: Perceived Importance of Information System Success Factors: A Meta Analysis of Group Differences. *Information & Management* 32 (1), 15-28 (1997)
39. Mason, R.O.: Measuring information output: A communication systems approach. *Information & Management* 1 (4), 219-234 (1978)
40. Moeni, M., Lapointe, L.: Selecting an Appropriate Operationalization of the System Usage Construct: An IT Artifact Perspective. In: Santana, M., Luftman, J.N., Vinze, A.S. (eds.):

- Proceedings of the 16th Americas Conference on Information Systems. Association for Information Systems (2010)
41. Myers, B.L., Kappelman, L.A., Prybutok, V.R.: A Comprehensive Model for Assessing the Quality and Productivity of the Information System Function: Toward a Theory for Information Systems Assessment. *Information Resources Management Journal* 10 (1), 6-26 (1997)
 42. Ng, B.-Y., Kankanhalli, A., Yip, J.W.L.: Use of Healthcare IS by Multiple User Groups: An Empirical Study of a Medication Management System. In: Galletta, D.F., Liang, T.-P. (eds.): *Proceedings of the 32nd International Conference on Information Systems*, paper 4. Association for Information Systems (2011)
 43. Niemi, E., Pekkola, S.: Adapting the DeLone and McLean Model for Enterprise Architecture Benefit Realization Process. In: *Proceedings of the 42nd Hawaii International Conference on System Sciences*. IEEE Computer Society (2009)
 44. Pak, H.D., Pak, S.M., Jang, I.H., Choe, Y.C.: A Measurement of Non-linear Relationship between IS Qualities and User Satisfaction. In: Santana, M., Luftman, J.N., Vince, A.S. (eds.): *Proceedings of the 16th Americas Conference on Information Systems*. Association for Information Systems (2010)
 45. Palmer, J.W.: Web site usability, design, and performance metrics. *Information Systems Research* 13 (2), 151-167 (2002)
 46. Palvia, P., Mao, E., Midha, V.: Research methodologies in MIS: an update. *Communications of the Association for Information Systems* 14, 526-542 (2004)
 47. Petter, S., DeLone, W.H., McLean, E.R.: Measuring information systems success: models, dimensions, measures, and interrelationships. *European Journal of Information Systems* 17 (3), 236 - 263 (2008)
 48. Pitt, L.F., Watson, R.T., Kavan, C.B.: Service Quality: A Measure of Information System Effectiveness. *MIS Quarterly* 19 (2), 173-187 (1995)
 49. Polančič, G., Heričko, M., Rozman, I.: An empirical examination of application frameworks success based on technology acceptance model. *The Journal of Systems and Software* 38 (4), 574-584 (2009)
 50. Prybutok, V.R., Zhang, X., Ryan, S.D.: Evaluating leadership, IT quality, and net benefits in an e-government environment. *Information & Management* 45 (3), 143-152 (2008)
 51. Raeth, P., Smolnik, S., Kügler, M.: Conceptualizing a very rich usage construct to measure the impact of organizational social web site usage on individual performance. In: Tuunainen, V.K., Rossi, M., Nandhakumar, J. (eds.): *Proceedings of the 19th European Conference on Information Systems*, Helsinki (2011)
 52. Rai, A., Lang, S.S., Welker, R.B.: Assessing the Validity of IS Success Models: An Empirical Test and Theoretical Analysis. *Information Systems Research* 13 (1), 50-69 (2002)
 53. Sabherwal, R., Jeyaraj, A., Chowa, C.: Information systems success: individual and organizational determinants. *Management Science* 52 (12), 1849-1865 (2006)
 54. Saeed, K.A., Abdinnour-Helm, S.: Examining the effects of information system characteristics and perceived usefulness on post adoption usage of information systems. *Information & Management* 45 (6), 376-386 (2008)
 55. Santos, G.D., Takaoka, H., de Souza, C.A.: An Empirical Investigation of the Relationship between Information Quality and Individual Impact in Organizations In: Santana, M., Luftman, J.N., Vince, A.S. (eds.): *Proceedings of the 16th Americas Conference on Information Systems*. Association for Information Systems (2010)
 56. Schaupp, L.C., Bélanger, F., Fan, W.: Examining the success of websites beyond e-commerce: an extension of the IS success model. *Journal of Computer Information Systems* 49 (4), 42-52 (2009)

57. Seddon, P.B.: A Respecification and Extension of the DeLone and McLean Model of IS Success. *Information Systems Research* 8 (3), 240-253 (1997)
58. Seddon, P.B., Staples, S., Patnayakuni, R, Bowtell, M.: Dimensions of information systems success. *Communications of the Association for Information Systems* 2, 1-60 (1999)
59. Sedera, D., Chian, F.T.T.: Reconceptualizing Usage for Contemporary Information System Success. In: Österle, H., Schelp, J., Winter, R. (eds.): *Proceedings of the 15th European Conference on Information Systems*, pp. 1344-1355. University of St. Gallen, St. Gallen (2007)
60. Sedera, D., Gable, G.G.: A factor and structural equation analysis of the Enterprise Systems Success Measurement Model. In: Agarwal, R., Kirsch, L., DeGross, J.I. (eds.): *Proceedings of the 25th International Conference on Information Systems*. Association for Information Systems (2004)
61. Seen, M., Rouse, A., Beaumont, N.: Explaining and Predicting Information Systems Acceptance and Success: An Integrative Model. In: Österle, H., Schelp, J., Winter, R. (eds.): *Proceedings of the 15th European Conference on Information Systems*, pp. 1356-1367. University of St. Gallen, St. Gallen (2007)
62. Shannon, C.E., Weaver, W.: *The Mathematical Theory of Communication*. University of Illinois Press, Urbana (1949)
63. Song, C.: Validating IS Success Factors: An Empirical Study on Webbased State or Local E-government Systems. In: Santana, M., Luftman, J.N., Vinze, A.S. (eds.): *Proceedings of the 16th Americas Conference on Information Systems*. Association for Information Systems (2010)
64. Teo, T.S.H., Srivastava, S.C., Jiang, L.: Trust in Electronic Government Success: An Empirical Study. *Journal of Management Information Systems* 25 (3), 99-131 (2008)
65. Urbach, N., Smolnik, S., Riempp, G.: The State of Research on Information Systems Success – A Review of Existing Multidimensional Approaches. *Business & Information Systems Engineering* 1 (4), 315-325 (2009)
66. Urbach, N., Smolnik, S., Riempp, G.: Improving the Success of Employee Portals: A Causal and Performance-Based Analysis. In: Alexander, P.M., Turpin, M., van Deventer, J. P. (eds.): *Proceedings of the 18th European Conference on Information Systems*, Pretoria (2010)
67. Vanlommel, E., DeBrabander, B.: The Organization of Electronic Data Processing (EDP) Activities and Computer Use. *The Journal of Business* 48 (3), 391-410 (1975)
68. Wang, Y.-S.: Assessing e-commerce systems success: a respecification and validation of the DeLone and McLean model of IS success. *Information Systems Journal* 18 (5), 529-557 (2008)
69. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly* 26 (2) 13-23 (2002)
70. Xu, J., Benbasat, I., Cenfetelli, R.T.: Does Live Help Service Matter? A Empirical Test of the DeLone and McLean's Extended Model in the E-Service Context.

Knowledge Creation in Requirements Engineering – A Systematic Literature Review

Linda Schneider, Kerim Hajji, Andreas Schirbaum, and Dirk Basten

University of Cologne, Cologne, Germany
{linda.schneider, kerim.hajji, andreas.schirbaum}@smail.uni-koeln.de
dirk.basten@uni-koeln.de

Abstract. Requirements engineering (RE) is crucial for software development, yet software requirements are often not properly identified or implemented. As the RE process thereby highly depends on human knowledge (explicit and tacit), this problem can in many cases be attributed to the lack of mutual understanding between customers and developers which is caused by diverging domain knowledge. To solve this problem, we conduct a systematic literature review to identify methods associated with Nonaka's organizational knowledge creation theory. We map eight such methods on six common RE problems as we analyze to which extent the methods overcome the associated problems. Although it is not always obvious which problems apply to specific software projects, the identified methods provide an adequate first approach to reduce the risk of potential RE problems, thus making project failure less likely.

Keywords: requirements engineering, knowledge management, knowledge creation, literature review

1 Introduction

Requirements engineering (RE) is crucial for software development, yet software requirements are often not properly identified or implemented [1]. Flaws in RE often cause a variety of problems, probably most importantly not meeting customer needs. Whereas such flaws can be corrected on later stages, these corrections usually come along with signification rework (60 to 100% of the original costs) [1-2].

Another RE problem concerns the (lack of) mutual understanding between customers and developers [3]. Requirements that are incorrectly communicated, perceived, or understood may lead to diverging perceptions on behalf of developers. Causes for these problems include inadequate RE methods or vague and not clearly postulated needs by the customer [1], [4].

RE highly depends on human knowledge which can be divided into two major categories [4-5]: *explicit knowledge* and *tacit knowledge*. The latter one may thereby be *embodied* or *not-yet-embodied*. However, *explicit knowledge* is the only type a person is fully aware of, completely understands and is able to articulate its meaning [4]. The lack of awareness about the two types of tacit knowledge is supposed to

cause erroneous or incomplete RE. We thus believe that to enhance and sustain the process of organizational knowledge creation helps to address the RE problems. To cope with this issue, we consider the organizational knowledge creation theory [6] which divides the organizational knowledge creation process into four sub-processes. Although a variety of methods exist to implement these processes, their contribution to the RE process has not been scrutinized. To close this gap, we address the following research questions (RQ):

1. Which methods have been proposed that can be applied to RE knowledge creation?
2. To which extent do these methods facilitate and simplify knowledge transfer between customers and developers to overcome common RE challenges in software projects?

To answer these questions, we conduct a systematic literature review [7]. We identify and analyze methods of knowledge creation and determine the degree of their applicability for eliminating common RE problems. We thereby apply organizational knowledge creation theory [6] to categorize the identified methods. Tacit and explicit knowledge play an important role in RE as knowledge is created through their interaction. To facilitate the full potential of (this) knowledge and to solve potential problems, all phases of Nonaka's knowledge spiral should be covered.

With our study, we contribute to theory and practice in the following ways. First, we advance theory of knowledge creation by mapping its methods to common RE problems. Moreover, we identify gaps that need to be addressed to further improve RE. Finally, we provide guidelines that help to solve common RE problems by applying basic knowledge creation methods that dissolve the potential error source of *embodied* and *not-yet-embodied knowledge* as well as flawed communication.

The remainder of this paper is as follows. Next, we refer to common problems in RE, the process of knowledge creation and the interrelatedness of these concepts. We then describe our literature review approach. Subsequently, we present our results. We finally discuss our findings and provide implications for research and practice.

2 Theoretical Background

2.1 RE Problems

Errors during RE inevitably lead to problems in system design and implementation, thus making RE crucial for the software development process [8-9]. These problems are often the result of insufficient requirement analysis. Requirement analysis process includes "deriving the system requirements through observation of existing systems, discussions with potential users and procurers [and] task analysis" [8] and aims to identify and document customer requirements for a planned software [1]. It "has to be organized in close collaboration with all parties involved from outside and inside" [1] and thus requires great coordination and communication effort [3]. Table 1 provides an overview of common problems related to requirements engineering.

Table 1. Problems of requirements engineering [1], [3]

Problems of Requirement engineering	Definition
Lack of information	Lack of information about the needs and requirements of customers and their relevance for the software
Heterogeneous understanding/ Communication gaps	Different views, competencies and experiences resulting in communication problems
Inability to articulate	Difficulties to express or externalize knowledge
Incomplete aggregation	Tacit dimension of knowledge can prevent the complete aggregation of the needs and requirements of the customer
Different project interests	Wrong focus areas and incorrect use of requirements
High degree of coordination and communication	Has to be organized in close collaboration with all involved articles from outside and inside

Throughout the software development process, different views, perspectives and interests need to be aligned among customers and developers [3]. The required communication can thereby be challenging [1], [3]. For instance, customers often determine solutions rather than requirements and their ideas may not reflect the shifting needs of the market [1]. Such problems may be caused by customers' inability to (sufficiently) articulate their wishes as they are in many cases unable to express or externalize their knowledge. Additionally, customers subconsciously have a specific notion of the software to be developed, but lack an adequate technological background. These obstacles make it difficult for the developers to recognize relevant product attributes which are perceived to be important by customers. Moreover, customers may be unable to express all of what they know due to the tacit dimension of knowledge [1]. This tacitness prevents the complete aggregation of needs and requirements of the customer. Furthermore, requirements are often used incorrectly. Instead of generating value for customers they are applied to justify a concrete implementation [3].

2.2 Knowledge Creation in Teams

Contrary to raw data or information, knowledge is personalized and exists only on individuals' minds [10]. When transferring knowledge, it is thus inevitable to transform knowledge into a form that is interpretable by receiving individuals or groups. In this context, two types of knowledge need to be distinguished [6]. Whereas *explicit knowledge* can be articulated, codified and communicated using symbols or language and can for instance be found in books or documents, *tacit knowledge* is about how we do things and therefore difficult to formalize and communicate. Moreover, *tacit knowledge* can be divided into *embodied knowledge* and *not-yet-embodied knowledge* [4-5], [11]. Individuals are aware of their *embodied knowledge* and use it consciously when applying processes and procedures. *Not-yet-embodied* knowledge is characterized by the need of individuals' analytical abilities, reflections, synthetical abilities, logical analysis and creativity. It is reflected in ideas, solutions, innovation and design. Whereas stakeholder's *tacit knowledge (embodied and not-yet embodied)* about requirements needs to be shared with the developer, *explicit knowledge* can be stored and transferred via documents.

Knowledge management can be defined as a “systemic and organizationally specified process for acquiring, organizing and communicating both *tacit* and *explicit knowledge* of employees so that other employees may make use of it to be more effective and productive in their work” [12]. In RE, it is important to achieve a common understanding by mutually sharing knowledge [6], [13]. Thereby, the knowledge creation process can be seen as a spiral based on both, *tacit* and *explicit knowledge* on the one hand and the interplay between both knowledge dimensions on the other hand [6], [10], [14-15]. Each iteration consists of the following phases: *socialization*, *externalization*, *internalization* and *combination* (cf. Figure. 1).

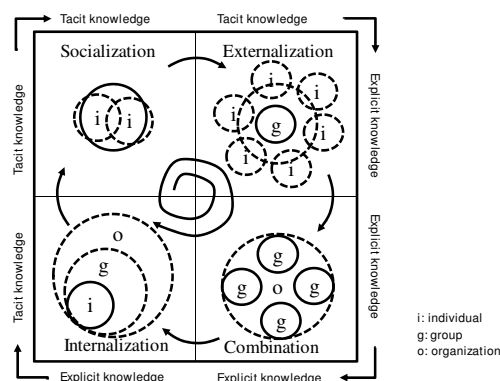


Fig. 1. Nonaka’s knowledge spiral [14]

Socialization describes a process where experiences are shared and by this means *tacit knowledge* is created [14-15]. In the field of RE, direct contact between customers and developers is an example for *socialization*. The process of converting *tacit knowledge* into *explicit knowledge* is called *externalization* [14-15]. During this process, individuals formulate their knowledge utilizing, for instance, easy understandable metaphors or analogies so that previous *tacit knowledge* of an individual is available for others. During *combination*, *explicit knowledge* is transformed into complex *explicit knowledge* [14-15]. A group uses knowledge from sources outside of the group to combine it with the group’s internal knowledge. Usage of knowledge repositories or computerized communications networks can support this process. During *internalization* (also considered as learning), *explicit knowledge* will be transferred into *tacit knowledge* by individuals [14-15]. Usage of documents, manuals or oral stories supports the conversion of knowledge as documenting and verbalizing experiences helps to internalize knowledge.

3 Literature Review

To answer our research questions (cf. section 1), we conducted a systematic literature review [16]. As basis for our search, we selected the following databases:

EBSCOhost, ProQuest, ScienceDirect, IEEE Xplore and AISEL. We thus relied on a multitude of publications, including the extended Senior Scholar's Basket of Journals.

Possibly relevant articles needed to meet the following search pattern in title, abstract, or keywords: (*knowledge management OR knowledge transfer OR knowledge creation OR knowledge worker OR knowledge capture OR knowledge network OR knowledge sharing OR knowledge retrieval OR information management OR information sharing OR information transfer OR information network OR data sharing OR data transfer OR data management OR experience OR know-how OR communication*) AND (*team OR collaborative OR groupware*). The search pattern is based on synonyms and differences in spelling and terminology of the main keywords *knowledge* and *team*. Synonyms and related terms were derived by brainstorming and collecting keywords from initially identified articles. We then verified the identified publications in terms of quality and relevance according to our study objective. In order to ensure the transferability of the results into the context of RE only studies referring to knowledge management in teams with an organizational context were taken into account. For the extraction and categorization of the approaches of knowledge management, a matrix approach was applied. At first the acquired approaches were assigned to the suitable phases of knowledge creation [6]. In the final step, we consolidated the approaches and assigned those to match the problems in RE. That way, we provide different solutions to the various RE problems.

4 Results

In the following, we explain each method and their applications to common RE problems. Table 2 provides an overview of this mapping.

4.1 Information Sharing

Knowledge is transformed to information once it is articulated [10]. Thus, methods of information sharing only refer to articulated knowledge. When using codification strategies, knowledge is unified and stored for further use, for instance, using databases or intranets [17], [20-22], [26]. Thus, knowledge can be applied by anyone in an organization regardless of time and space [18], [26], for instance, to be used in upcoming projects [17], [25]. On the one hand, there are centralized approaches, where a centrally located resource can be used by different clients [22], [26]. On the other hand, information sharing can be realized by using P2P to enable team members to collaboratively work on documents [24], [26]. These so-called people-to-document methods are useful to transfer *explicit knowledge* [18], [26] and especially support *combination* [4], [17]. This corresponds to re-use of knowledge. The centralized as well as the P2P approach encourage that knowledge is gathered from various sources. This forms a broader knowledge foundation, making it less likely that a *lack of information* is encountered.

Table 2. Methods and their mapping on RE problems

Method s= socialization; e=externalization; c= combination; i= internalization	Lack of information	Heterogeneous understanding / Communication gaps	Inability to articulate	Incomplete aggregation	Different project interests	High degree of coordination/ communication
Information sharing [4], [17–27] s, e, c, i	X	X		X		X
Synchronous and asynchronous communication tools [4], [23–31], [31–39] s, e, c	X	X		X		X
Involvement of external sources [19], [21], [24], [30] s, c,	X		X		X	X
Training [17–20], [30], [40] s, c, i	X	X			X	
Search mechanisms [20–22], [24–25], [28], [40] e, c,	X					X
Face-to-face interaction [4], [17–18], [20], [26–33] s, e, c,	X	X	X	X		X
Work guidelines [17], [19], [24], [40] s, e, c, i	X	X	X			
Creative techniques [17–18], [20], [24], [27–32] s, e, c, i			X	X		

Information sharing can support coordination, for example, when documents with tasks, schedules or meeting agendas are made available for team members [24], [26]. If shared revision of the documents is possible, it is also supportive for communication as the work of individuals can be shared immediately and a related discussion might start. *Explicit knowledge*, referring to the content of documents, can be processed and internalized by accessing, comprehending and reflecting [4], [14–15].

Thereby, creating new ideas and opinions helps to master the *internalization* process. Thus, it is likely that new knowledge is understood and applied. The reflection can be helpful to discover own *embodied knowledge* and make it available for the knowledge creation process [4]. In addition and when writing a document, there is more time to think about articulation than during synchronous communication. This makes it more likely that *tacit knowledge* is expressed and therefore indirectly helps *externalization*, thus solving the problem of *inability to articulate*. However, while writing a document, it is certainly possible to express tacit (not-yet-embodied) knowledge if it reflects ideas and opinions [4]. Thus, it is possible to support *socialization* by reflecting on those documents even if knowledge is only available in text form and cannot be transferred by interaction. If customers' ideas and opinions are understood, they can be helpful to interpret the desired requirements, thus facilitating homogeneous understanding. The problem of its *heterogeneous understanding/communication gaps* would at least be partly eased. These ideas and opinions support discussions, which could even create *not-yet-embodied knowledge* [4] and thus solve the problem of incomplete aggregation.

4.2 Synchronous and Asynchronous Communication Tools

This section illustrates the use of computer-mediated communication tools and their effects on the knowledge creation process [29]. The tools can be divided into two categories - synchronous and asynchronous communication tools - and are one of the most important methods *for socialization* besides face-to-face interaction. Moreover, they are also useful in *combination* and *externalization*. Synchronous communication tools, such as video-conferencing, instant messaging and desktop sharing [29] enable users to communicate at the same time at different places. They bring individuals together in real-time and offer immediate feedback. Asynchronous communication tools such as email, knowledge portals [32], wikis and blogs offer a delayed feedback.

Face-to-face contact can be costly and needs much coordination effort [30]. Especially in cases of long distances between the cooperating parties, the costs of meeting on a regular basis are likely to be high. For this purpose, knowledge transfer in geographically distributed teams should be provided with a "support that facilitates the simulation of face-to-face interaction" [4]. Recent advances in information technology have enabled distributed teams to perform knowledge work effectively without meeting face-to-face [26]. They can be effective against various RE problems, particularly in geographically distributed teams that require an effective and low cost communication [25]. If face-to-face contact is not available to the extent necessary to get the job done, communication tools have to be implemented to support the knowledge creation process [4], [29].

Because a deep understanding of team goals is necessary, teamwork highly depends on communication [36]. In order to reach a mutual understanding within a team, IT support is of high importance [24]. The direct interaction through synchronous communication tools helps to mitigate the *lack of information* problem. Tools like IM (Instant Messenger), Chat or VoIP (Voice over IP) can play an important role for interaction and spreading of knowledge in teams. Asynchronous communication

tools maintain an overtime comprehensible stream of knowledge and improve team interaction. However, in RE it is necessary to get to know the partners and learn their needs, goals and work situations [3].

Communication tools have the ability to reduce the problem of *heterogeneous understanding/communication gaps*, by creating an interpersonal relationship, without meeting each other. Video conferencing for instance has the ability to generate the needed media-richness in order to transfer *tacit knowledge* without meeting face-to-face [29].

4.3 Involvement of External Sources

One option to integrate external sources are power users. Those are selected from user groups based on their experience with and knowledge of business processes [30]. While spending time with developers, the developer's technical jargon and the user's domain knowledge can be shared. This leads to *combination* resulting from the *explicit knowledge* exchange. In case of missing knowledge, the problem of *lack of information* may occur. The close contact and collaboration consequently enables *socialization*. As a result, a common goal is developed and a *high degree of coordination and communication* is reached to prevent *different project interests*.

Experts and consultants are other types of external sources [19]. Using that form, *combination* takes place, as own knowledge is extended based on experts' knowledge to prevent a *lack of information* problem. Before the start of a project, not all required knowledge is necessarily available. Hence, it might be possible that the required knowledge is partly not available and suitable experts need to be identified [21], [24]. By the application of, for example, knowledge repositories and knowledge maps (see *use of search mechanisms*), the search can be simplified [21], [24]. This is especially important if the team is homogeneous as they get new perspectives and new routines from external sources [24].

4.4 Training

Ways to implement training include mentoring, lectures, lessons, audits, classroom training, apprenticeship training or power users [17], [19], [30], [40]. The distinct training approaches differ regarding their empowerment to share *tacit* and/or *explicit knowledge*. While traditional classroom training is deemed an effective method to transfer *explicit knowledge*, other approaches like apprenticeship are better suited to share both *tacit* and *explicit knowledge* [30]. Training is one of the most important methods for *internalization* as it is closely related to learning and *internalization* is basically deemed as learning [14-15]. As new information is transferred (*combination*), understood and internalized and can thus be converted to requirements, training is an effective means to solve the *lack of information* problem. It is also a social process and thus facilitates the sharing of common mental models through dialogs and activities [20] leading to *socialization*. In addition, it is possible to develop inter-individual relationships [30]. By strengthening these relations, it is more likely that project goals are better aligned in the future helping to overcome *different project*

interests. Training may improve mutual understanding, as certain terminologies are clarified or mental models are shared and the problems of *heterogeneous understanding/communication gaps* might be solved. On the one hand, developers could better understand user requirements. On the other hand, training with prototypes gives users an understanding of the system and developers' interpretation of requirements [30].

4.5 Search Mechanisms

Search mechanisms are mainly concerned with the issue to retrieve specific knowledge and to simplify its identification. IT support is indispensable in this process. Thus, IT represents the foundation for the following examples of search mechanisms. Assigned to Nonaka's knowledge spiral, search mechanisms can help in the *combination* and the *externalization* phase of the knowledge creation process due to the fact, that only *explicit knowledge* can be created through search mechanisms. One example for such search mechanisms are knowledge maps, which can be understood as a kind of catalogue or index in which for instance an employee's knowledge, know-how, and expertise are managed and made available for searches [21].

Another search mechanism is the peer-to-peer (P2P) approach, which is presented as a network or network-based computing model. Applying this approach, it is possible to mutually provide and access knowledge among the network members [22-23].

Also assigned to the search mechanisms are patterns [40]. These patterns are described as "a powerful mechanism for [...] information-to-knowledge conversion" [40]. They consist of several elements which allow readers to decompose knowledge into its various parts and then to create or assemble a suitable solution for her/his problem out of these decomposed parts [40]. The use of search mechanisms helps to counter two of the identified RE problems. First, it helps in case of *lack of information*. Second, it eases *high coordination and communication* efforts.

Knowledge maps and patterns are predestined to encounter the problem of information lacking. Both organize the knowledge in a way that it can be easily and quickly retrieved and used [21], [40]. Applied to RE, specifically on the interaction between developers and customers, they help gathering all necessary information about the needs and requirements of the customers. To counteract the high degree of coordination as well as the limited contact with customers, developers and customers can rely on the P2P approach. With the help of P2P networks, RE will be organized in close collaboration with all parties involved and the communication and sharing of knowledge can be eased [22-23].

4.6 Face-to-Face Interaction

Face-to-face interaction enables a direct person-to-person knowledge transfer which can be crucial for the creation of knowledge [17]. Especially *socialization* works best with social interaction or demonstration [20]. Moreover, face-to-face interactions can support the *externalization* and *combination* phases of knowledge creation.

There are various activities to implement face-to-face interactions such as project meetings, work related discussions or simple face-to-face conversations during pro-

jects [4], [18]. Groups that use face-to-face communication have access to a number of verbal and non-verbal cues that are important for determining the underlying preferences, positions and emotional experiences of other team members [28], [41]. This may for instance help to align the understanding of the specific attributes of the final product. Consequently, using face-to-face methods in RE might help to reduce the problem of a *heterogeneous understanding/communication gaps*. In the same way, the close interaction achieved by direct face-to-face contact supports the creation of knowledge by building social ties. It also helps team members to gain trust and credibility as well as to build and sustain social networks which are crucial for a successful collaboration [27], [30-31]. Thus, the problem of *different project interests* may be solved. Other activities of *face-to-face interaction* such as discussions may support triggering of unconscious *not-yet-embodied knowledge* [4]. This effect reduces the *inability of the customers to articulate the requirements*. Additionally, creative ideas are the result of discussions and interactions at the conscious level, resulting in *combination*. "Interaction among individuals is an important factor for innovation which, is a result of immediate concepts formation within the individual's consciousness" [4]. These resulting ideas from the knowledge creation process with face-to-face interaction might generate the specific product attributes that provide satisfaction for the customer. Close contact, that is, regular meetings with the customer, might help to mitigate the problem of a *lack of information and high degree of coordination and communication*.

4.7 Work Guidelines

In addition to the above-mentioned methods, there are work guidelines that facilitate the creation of knowledge in teams, or rather standardize it. These work guidelines can create *explicit* as well as *tacit knowledge* and thus can be classified in all four categories of Nonaka's knowledge spiral. Work guidelines are primarily principles or practices that help to successfully distribute knowledge. One type of such work guidelines are the previously presented patterns (cf. section 4.5). Since it is strictly defined of which elements the patterns consist and how they should be drawn up and filled, patterns can be seen as a type of work guideline [40]. Also among work guidelines are best practices for communication norms through which a common understanding should be established regarding aspects like goals, objectives or task requirements [24]. Best practices that help to create such common understanding are for instance the explicit communication of each other's areas of shared understandings or the use of explanations without local jargon. Another work guideline approach is the model of knowledge processes in organizations by Geisler [19]. This model consists of four modes or stages: generation, transfer, implementation and absorption. Each stage describes another condition in the knowledge process. With all four stages terminated, the knowledge process is complete. Additionally, uncomplicated and often applied tools like checklists, templates, methodologies, procedures and previous project documents are also considered as work guidelines [17]. All these approaches and methods are examples for work guidelines that should help to create knowledge and to

transfer it between team members. To prevent *lack of information*, different work guidelines can be applied. One of them are the above mentioned patterns.

Another approach to encounter this problem is the model of knowledge process in organizations [19]. If the four stages of the model are followed as proposed by Geisler [19], *lack of knowledge* should be eliminated especially in the generation stage in which *explicit* as well as *tacit knowledge* is gathered. In order to solve the problem concerning the *inability to articulate* ones needs and to avoid the lacking of knowledge, patterns can be used as well. The fact, that patterns support decomposing knowledge [40] helps to retrieve and understand knowledge better and eases the articulating needs. Next to the patterns, Geisler's model of knowledge process [19] can help counteracting these two problems. In the implementation and in the absorption stage of the model, newly gathered knowledge is integrated with the existing knowledge [19]. One last problem that work guidelines can counteract is the *heterogeneous understanding/the communication gaps*. Here, best practices of communication can be applied since their goal is to create a common understanding [24].

4.8 Creative Techniques

Creative techniques encourage knowledge exchange in teams and call for team members' creativity. Thereby, they help to create *explicit* and *tacit knowledge* and can thus be assigned to all four categories of Nonaka's knowledge spiral.

The storytelling method is hereby one of the most common form of creative techniques. It is a facilitator of knowledge sharing within a group of people [31]. Applying this method, knowledge or experience concerning past or on-going projects is presented in form of stories or narratives that should facilitate team members to understand and to internalize the knowledge [17]. Thereby, storytelling is especially suitable for transferring *tacit knowledge* [30].

An equally common approach of creative techniques is the use of boards of any kind, such as electronic whiteboards, smartboards or system storyboards [20], [24], [28], [39]. They help to spread and retain specific knowledge in teams. The use of such boards often goes hand in hand with the implementation of storyboarding or brainstorming sessions in a team to capture and spread ideas [18], [28], [39].

These examples show that creative techniques create knowledge with the help of creative exchange among team members. The use of creative techniques can encounter two common RE problems. These techniques might help if the customers are *not able to articulate* their wishes due the customer's *lack of knowledge* about what she/he wants and thus can help in case of *incomplete aggregation*.

To counteract the problems of *inability to articulate* the needs and *lack of knowledge*, each identified creative technique can be used. Storytelling as well as boards and thus also storyboarding and brainstorming sessions help to understand and internalize especially *tacit knowledge* so that customers can easily retain knowledge and minimize lacking of the same [17], [30], [39].

The *incomplete aggregation* of requirements can be prevented by the implementation of storyboarding and brainstorming sessions with the use of boards or without.

Storyboarding and brainstorming help to create, submit and tender new ideas so that new needs and requirements may come to customers' minds [39].

5 Discussion

Considering previous studies for KM in the field of RE, software projects prove to be challenging for developers as cross-functional stakeholders, especially in distributed projects, need to specify their requirements regardless of cultural, time zone or organizational barriers [42]. Consequently, a special need for effective KM practices and methods arises.

Many RE techniques are similar to the ones of knowledge acquisition. Considering previous research, Byrd et al. [43] discovered 19 elicitation techniques that are used during knowledge acquisition or during RE. Although those results are helpful in this context, our approach significantly differs: (1) We cope with KM techniques in general. (2) Our study is problem-oriented and does not only focus on techniques related to selected problem domains (i.e., our study includes problems like *different project interests*). (3) We consider all phases of the knowledge spiral.

Our study defines six major problems in the RE and eight methods to counteract them. Concerning Nonaka's knowledge spiral it should be noted that all phases (except *internalization*) are all well-covered. Especially the *combination* phase is supported, as it is much easier to transfer *explicit knowledge* than *tacit knowledge*. Hence, all methods are applicable in the *combination* phase. There are also methods like *use of information sharing* or *use of creative techniques*, which support all four knowledge creation phases. However, this does not necessarily mean that these methods cover an especially high number of problems.

Only four methods are allocated to the *internalization* phase. This means that most methods do not support the learning process. Thus, it might be necessary to implement additional learning opportunities besides the here described KM methods.

Considering *socialization*, face-to-face interaction is most important as *tacit knowledge* is best transferred through personal interaction. Related disadvantages are that it is time-consuming, costly and sophisticated. Furthermore, personal interaction leads to high coordination effort. In some cases, it is even impossible due to geographically distributed teams. However, this study shows that there are other methods to support *socialization*. A cheaper, less time-consuming and easy to implement way are synchronous or asynchronous communication tools.

At least two solutions for each problem could be identified. The problem *different project interests* has by far the fewest possible methods allocated. This problem cannot be solved through simple knowledge transfer or creation. Instead, it is necessary to develop mutual trust to align the developer's and customer's goals.

The problem *incomplete aggregation* concerns the *tacit* dimension of knowledge. Nevertheless, there are five methods that cope with this problem. This is noteworthy as the transfer of *tacit knowledge* is often seen as more critical compared to the transfer of *explicit knowledge*.

Additionally, we emphasize that some approaches were mutually mentioned. Examples include training sessions and power user. External power users have conducted the training, so it might be possible that different approaches reveal their full potential only if mutually applied with other methods. Of course it is also possible that they amplify each other, for instance, to use search mechanisms to identify experts.

6 Conclusion

The first goal of this work was the identification of methods that can be applied to RE knowledge creation. Conducting a literature review, we were able to identify eight such methods. The second goal was to elaborate to which extent these methods facilitate and simplify the knowledge transfer between customers and developers to create new knowledge. Using the Nonaka knowledge spiral it was discovered that any method could help with at least two common RE problems (cf. section 5). In particular, face-to-face interactions and use of information sharing seem to take a special role as they could help solving five problems. Consequently, it seems likely that the identified methods are relevant for practice and that their application is useful during RE.

Although it is not always obvious which problems apply to specific projects, these methods provide an adequate first approach to reduce the risk of potential problems, thus making project failure less likely. To conclude, there are various approaches or methods of KM that are of value for the RE and should be applied in this context to encounter the existing problems.

The results of our work provide a foundation for new research in this area. First, the transferability of the methods to the context of RE has to be verified with the help of empirical validation. Second, we cannot exclude that other problems exist which the identified methods do not mitigate. Furthermore, not all the problems that might arise from the application of these methods were considered. It would be imaginable, for example, that during information sharing, team members do not want to publish their information in a centralized system because they are afraid of losing their information superiority.

References

1. Semmelrock-Picej, M.K.H.: Information Technology based Customer Knowledge Management Externalisation Techniques for Requirements Analysis. In: de Casto Net, M. (ed.): Proceedings of the 4th European Conference on Information Management and Evaluation (ECIME 2010), pp. 353-364 (2010)
2. Pressman, R.S.: Software Engineering. A Practitioner's Approach. McGraw Hill (2001)
3. Herzwurm, G., Mellis, W., Schockert, S.: Qualitätssoftware durch Kundenorientierung. Die Methode Quality Function Development (QFD). Grundlagen, Praxis und SAP R/3 Fallbeispiel. Vieweg, Braunschweig (1997)
4. Feghali, T., El-Den, J.: Knowledge transformation among virtually-cooperating group members. Journal of Knowledge Management 12, 92-105 (2008)

5. Scharmer, C.O.: Organizing around not-yet-embodied knowledge. In: von Krogh, G., Nonaka, I., Nichiguchi, T. (eds.): *Knowledge Creation: A Source of Value*. Macmillan Press (2000)
6. Nonaka, I.: A Dynamic Theory of Organizational Knowledge Creation. *Organization Science* 5, 14-37 (1994)
7. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly* 26, 13-23 (2002)
8. Sommerville, I.: *Software Engineering*. Addison Wesley, Edinburgh (2006)
9. Asghar, S., Umar, M.: Requirement Engineering Challenges in Development of Software Applications and Selection of Customer-off-the-Shelf (COTS) Components. *International Journal of Software Engineering* 1, 32-50 (2010)
10. Alavi, M., Leidner, D.E.: Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly* 25, 107-136 (2001)
11. Scharmer, C.O.: Self-transcending knowledge: sensing and organizing around emerging opportunities. *Journal of Knowledge Management* 5, 137-151 (2001)
12. Alavi, M., Leidner, D.E.: Knowledge Management Systems: Issues, Challenges, and Benefits. *Communications of the Association for Information Systems* 1, 1-37 (1999)
13. Reich, B.H.: Managing knowledge and learning in It projects: A conceptual framework and guidelines for practice. *Project Management Journal* 38, 5-17 (2007)
14. Nonaka, I., Konno, N.: The concept of ba: Building a foundation for knowledge creation. *California Management Review* 40, 40-54 (1998)
15. Nonaka, I., Takeuchi, H.: *The knowledge creating company. How Japanese companies create the dynamics of innovation*. Oxford Univ. Press, New York (1995)
16. Kitchenham, B., Charters, S.: *Guidelines for performing Systematic Literature Reviews in Software Engineering* (2007)
17. Barclay, C., Osei-Bryson, K.-M.: An Exploration of Knowledge Management Practices in IT Projects: A Case Study Approach. In: *Proceedings of the Americas Conference on Information Systems (AMCIS 2010)*, pp. 1-10 (2010)
18. Decker, B., Landaeta, R.E., Kotnour, T.G.: Exploring the relationships between emotional intelligence and the use of knowledge transfer methods in the project environment. *Knowledge Management Research Practice* 7, 15-36 (2009)
19. Geisler, E.: A typology of knowledge management: strategic groups and role behavior in organizations. *Journal of Knowledge Management* 11, 84-96 (2007)
20. Mohamed, M., Stankosky, M., Murray, A.: Applying knowledge management principles to enhance cross-functional team performance. *Journal of Knowledge Management* 8, 127-136 (2004)
21. Gray, P.H.: The effects of knowledge management systems on emergent teams: towards a research model. *The Journal of Strategic Information Systems* 9, 175-191 (2000)
22. Desouza, K.C., Evaristo, J.R.: Managing knowledge in distributed projects. *Communications of the ACM* 47, 87-91 (2004)
23. Gupta, S., Bostrom, R.: Using peer-to-peer technology for collaborative knowledge management: concepts, frameworks and research issues. *Knowledge Management Research Practice* 4, 187-196 (2006)
24. Malhotra, A., Majchrzak, A.: Enabling knowledge creation in far-flung teams: best practices for IT support and knowledge sharing. *Journal of Knowledge Management* 8, 75-88 (2004)
25. Zhuge, H.: Knowledge flow management for distributed team software development. *Knowledge-Based Systems* 15, 465-471 (2002)

26. Gupta, A., Mattarelli, E., Seshasai, S., Broschak, J.: Use of collaborative technologies and knowledge sharing in co-located and distributed teams: Towards the 24-h knowledge factory. *The Journal of Strategic Information Systems* 18, 147-161 (2009)
27. Kotlarsky, J., Oshri, I.: Social ties, knowledge sharing and successful collaboration in globally distributed system development projects. *European Journal of Information Systems* 14, 37-48 (2005)
28. Majchrzak, A., Malhotra, A., John, R.: Perceived Individual Collaboration Know-How Development through Information Technology-Enabled Contextualization: Evidence from Distributed Teams. *Information Systems Research* 16, 9-27 (2005)
29. Schwartz, D.G.: Integrating knowledge transfer and computer-mediated communication: categorizing barriers and possible responses. *Knowledge Management Research Practice* 5, 249-259 (2007)
30. Volkoff, O., Elmes, M.B., Strong, D.M.: Enterprise systems, knowledge transfer and power users. *Journal of Strategic Information Systems* 13, 279-304 (2004)
31. Michailova, S., Sidorova, E.: From group-based work to organisational learning: the role of communication forms and knowledge sharing. *Knowledge Management Research Practice* 9, 73-83 (2011)
32. Thomas, D.M., Bostrom, R.P., Gouge, M.: Making Knowledge work in virtual teams. *Communications of the ACM* 50, 85-90 (2007)
33. Kock, N., Lynn, G.S., Dow, K.E., Akgün, A.E.: Team adaptation to electronic communication media: evidence of compensatory adaptation in new product development teams. *European Journal of Information Systems* 15, 331-341 (2006)
34. Jarvenpaa, S., Staples, D.: The use of collaborative electronic media for information sharing: an exploratory study of determinants. *The Journal of Strategic Information Systems* 9, 129-154 (2000)
35. Kock, N., McQueen, R.J.: Groupware support as a moderator of interdepartmental knowledge communication in process improvement groups: an action research study. *Information Systems Journal* 8, 183-198 (1998)
36. Rafaeli, S., Ravid, G.: Information sharing as enabler for the virtual team: an experimental approach to assessing the role of electronic mail in disintermediation. *Information Systems Journal* 13, 191-206 (2003)
37. Wang, J.-K., Ashleigh, M., Meyer, E.: Knowledge sharing and team trustworthiness: it's all about social ties!. *Knowledge Management Research Practice* 4, 175-186 (2006)
38. Wolfe, M.: Broadband videoconferencing as a knowledge management tool. *Journal of Knowledge Management* 11, 118-138 (2007)
39. Zahniser, R.A.: Design By Walking Around. *Communications of the ACM* 36, 115-123 (1993)
40. May, D., Taylor, P.: Knowledge management with patterns. *Communications of the ACM* 46, 94-99 (2003)
41. Kelly, J.R., Barsade, S.G.: Mood and Emotions in Small Groups and Work Teams. *Organizational Behavior and Human Decision Processes* 86 (1), 99-130 (2001)
42. Damian, D.: Stakeholders in Global Requirements Engineering: Lessons Learned from Practice. *IEEE Software* 24, 21-27 (2007)
43. Byrd, T.A., Kathy, L.C., Zmud, R.W.: A Synthesis of Research on Requirements Analysis and Knowledge Acquisition Techniques. *MIS Quarterly* 16, 117-138 (1992)

Developing an Organizational Culture Supportive of Business Process Management

Sanja Tumbas¹ and Theresa Schmiedel²

¹ University of Münster, The Münster School of Business and Economics, Münster, Germany
sanja.tumbas@uni-muenster.de

² University of Liechtenstein, Institute of Information Systems, Vaduz, Liechtenstein
theresa.schmiedel@uni.li

Abstract. In recent years, a holistic understanding of business process management (BPM) has evolved in Information Systems (IS) research. This understanding considers organizational factors such as strategic alignment, governance, people, and culture as important aspects of BPM beyond Information Technology (IT) and methods. Particularly, organizational culture has been increasingly recognized as a critical success factor for the efficiency and effectiveness of business processes. While existing research already examined which cultural values determine a supportive environment for realizing BPM objectives and how these values can be measured, it has not been explored to date how such a supportive organizational culture can be developed. Based on a review of the literature, we therefore investigate the case of a global IT software provider to gain a first understanding of strategies organizations can implement to develop a supportive cultural setting for their BPM approach. We show how these findings extend recent studies in BPM research.

Keywords: business process management, organizational culture, culture development, BPM culture

1 Introduction

Business process management (BPM) is a field of research which has increasingly gained importance in the Information Systems (IS) discipline. In recent years, a holistic understanding of BPM has evolved, which considers organizational factors such as strategic alignment, governance, people, and culture as important aspects of BPM beyond Information Technology (IT) and methods [1]. Particularly, organizational culture is an acknowledged success factor in BPM initiatives [2-5]. “Having good systems and the right structure in place” [4] without addressing cultural aspects is not sufficient for achieving organizational effectiveness and efficiency. Industry analysts also recognize the relevance of culture: “BPM as a discipline requires an organization to change its culture and work practices” [6].

Culture refers to the shared basic assumptions, values, or beliefs of a group [7]. In recent years, researchers increasingly addressed and systematized the role of culture

in IS research generally and BPM research specifically, differentiating several levels of culture and relations between culture and BPM [8-10]. The academia recognizes that organizational culture should facilitate a business process initiative [3], [11]. Moreover, first results exist on what a supportive organizational culture could look like [12-13]. Nevertheless, it seems to have not been examined yet, how organizations could develop and maintain an organizational culture that supports the efficiency and effectiveness of their business processes. In addition, calls for research regarding this topic can be identified [9], [14]. Therefore, the objective of the study is to provide initial insights to possible measures organizations can take to improve the supportiveness of their culture for BPM. Thus, the paper is guided by the following research question: *How can an organization develop a culture supportive of BPM?*

To address this research question, in section 2 we first conduct a review of the literature to gain a more detailed overview of existing research on the topic. Then, in section 3 we look into the specific case of a global IT software provider. In section 4, we report on the strategy and activities which are derived from interviews with the company to improve the supportiveness of the organizational culture for BPM. The results are discussed in section 5. Finally in section 6, the paper concludes with an overall outlook.

2 Research Background

2.1 Understanding of Culture and Culture Development

In this paper, culture is understood as the shared values of a group that become visible in actions and structures [7], [9]. This means that manifestations of culture can be distinguished according to their visibility. Invisible values represent the defining elements of a particular culture [15]. Actions and structures however, refer to the visible part of culture which is the instantiation of the values that shape a particular culture [7]. These manifestations of culture can be found in any kind of group culture, e.g. nations, organizations, or work groups. In our study, the focus is on organizations. More specifically, we are interested in how change, particularly organizational culture development, can be fostered. Hence, we take a functionalists view, i.e. consider culture as changeable by management according to their objectives. In this sense, culture is a tool of management-driven change in core values which can lead to successful organizational transformation.

2.2 Literature Review on Culture in BPM (-related) Research

After introducing our understanding of the culture concept, an essential step in the research process is the investigation of the literature which provides an overview of the existing body of knowledge [16]. Therefore, a comprehensive literature review is conducted, following a rigorous approach as suggested by vom Brocke et al. [17] to ensure reliability and validity of the results. The review extends recent literature reviews on the topic particularly focusing on measures to develop a BPM-supportive

cultural environment in an organization [9], [18]. We follow the suggestion of previous studies that “a review of culture’s role in BPM-related concepts, like BPR, may provide valuable insights transferable to the BPM concept” [9]. Thus, we set the keywords for identifying relevant literature beyond a narrow BPM focus.

To gain in-depth insights to various research streams, the following search strings were deployed: "Process management" and Cultur*, "Process change" and Cultur*, "Process reengineering" and Cultur*, "Process re-engineering" and Cultur*, BPR and Cultur*, "Process redesign" and Cultur*, "Business Transformation" and Cultur*. The literature search incorporated titles and abstracts of conference and journal papers without restricting the time period. In a first step, central IS and BPM sources were examined, i.e. the Senior Scholars’ Basket of Journals provided by the Association for Information Systems (AIS)¹, the Business Process Management Journal, and the International Conference on Business Process Management. Beyond, the scope of the literature search also included the following journal databases: AIS Electronic Library², Science Direct, and Ebsco. In addition, the following conferences were included: AIS Conferences³ and AIS Affiliated Conferences⁴.

Out of the total 153 articles which fulfilled the basic search criteria, only those articles remained in the scope which fulfilled two key criteria: 1. elaborate on cultural aspects and 2. concerned with an organization as a unit of analysis. This resulted in 23 articles which served as basis for further classification.

A concept-centric approach [19] is used to structure the identified articles (Table 1). First of all, we differentiate between papers which refer to *organizational culture* and those which refer to the *culture of a specific management approach* (e.g. BPM, Business Process Reengineering (BPR), Total Quality Management (TQM)). Second, we distinguish the categories *describe*, *assess* and *develop* to characterize the way in which authors engage with the topic of culture. The majority of identified articles mentions and *describes* organizational culture as an influencing factor for process-related initiatives. There are few articles which apply a certain way of *assessing* culture to determine its typology, while there are much more which mention that the assessment would be desired or necessary. Here only those are included which directly propose a way for assessment. The last category contains articles which make suggestions on how to *develop* a certain culture with respect to process-related initiatives. This category is highlighted in Table 1 since it stays rather unpopulated.

The literature classification provides first evidence for a lack of research on how to develop an organizational culture that supports BPM. For example, Philip and McKeown [20] are in the *develop* category and explain the use of a culture framework originating from social anthropology to identify culture development paths in business transformations. While TQM initiatives are understood as a trigger of cultural change in business transformations, the paper does not reveal what a TQM-supportive organizational culture looks like or how the facilitating culture could be developed.

¹ <http://home.aisnet.org/displaycommon.cfm?an=1&subarticlenbr=346>

² <http://aisel.aisnet.org/communities.html>

³ <http://aisel.aisnet.org/conferences/>

⁴ <http://aisel.aisnet.org/affiliated/>

Yet, Yong and Pheng [21] introduce culture-based TQM implementation strategies. They suggest that either the predominant organizational culture should be modified or the TQM approach should be suited to the culture. In accordance with that, the proposed strategies are to maintain or adjust either the culture or the TQM approach depending on the cultural type of the organization. However, these strategies are very high-level and do not propose concrete activities on how to actually realize a TQM-supportive culture.

Table 1. Categorized Literature Review Results

	Organizational Culture	Culture of Specific Management Approach (e.g. BPM, TQM, BPR)
DESCRIBE culture	Guha et al. [22]; Motwani, Kumar, & Antony [23]; Motwani, Mirchandani, Madan, & Gunasekaran [24]; Motwani, Subramanian, & Gopalakrishna [25]; Nah, Lau, & Kuang [26]; vom Brocke & Sinnl [9]; vom Brocke & Sinnl [27]; vom Brocke & Schmiedel [18]; Rosemann & de Bruin [28]; Rosemann et al. [29]; Hammer [30]; Mosadegh Rad [31]; Zucchi & Edwards [32]; Alibabaei et al. [3]; Baird, Hu, & Reeve [33]; Philip & McKeown [20]; Yong & Pheng [21]; Revenaugh [34]; Kappos & Croteau [35]	de Bruin & Rosemann [1]; Mosadegh Rad [31]; vom Brocke & Sinnl [9]; vom Brocke & Sinnl [27]; vom Brocke & Schmiedel [18]; Cao, Clarke, & Lehaney [36]; Yong & Pheng [21]
ASSESS culture	Schmiedel, vom Brocke, & Recker [13]; Philip & McKeown [20]; Yong & Pheng [21]; Revenaugh [34]; Kappos & Croteau [35]	Schmiedel, vom Brocke, & Recker [13]; Kohlbacher & Gruenwald [37]
DEVELOP culture	Philip & McKeown [20]; Yong & Pheng [21]	

The section presented an overview on the prevailing literature on organizational culture and various process management approaches such as BPM, TQM, and BPR. The analysis of the identified literature strengthens the argument that further empirical research is needed to examine how organizations can develop a culture that supports their BPM approach.

2.3 Theoretical Basis of Culture's Role in BPM

Our empirical study builds on research which we identified in the literature review, particularly it ties in with the BPM-Culture-Model of vom Brocke & Sinnl [9]. Figure 1 shows the three main constructs of the model: (1) BPM culture is understood as a culture supportive of achieving efficient and effective business processes, (2) cultural context refers to the existing cultural setting (e.g. consisting of organizational, national, and work group cultures) which a BPM initiative faces at the beginning, and (3) cultural fit refers to the basic congruence which is required between BPM culture and cultural context for a BPM initiative to be successful.



Fig. 1. BPM-Culture-Model [9]

In recent studies, the concept of BPM culture was examined in more detail. Particularly, the cultural values which determine BPM culture were specified based on literature reviews and a global Delphi study [12], [18]. As a result, the so-called CERT values are identified as directly supportive of achieving BPM objectives [12]. Table 2 illustrates which values define the acronym CERT and which cultural subdimensions are represented by each of the values.

Table 2. CERT Values and their Subdimensions [12-13]

CERT values	Subdimensions	
Customer orientation	external customer	internal customer
Excellence	Innovation	continuous improvement
Responsibility	Accountability	commitment
Teamwork	formal structures	informal structures

To examine how far the CERT values are realized in an organization, a measurement instrument has been recently developed [13]. It measures the degree to which the culture of an organization is supportive of BPM. In case the CERT values are not lived in an organization, cultural resistance can arise. This means that a cultural fit is not achieved and as a consequence two scenarios are possible. Either cultural change facilitates the realization of the CERT values or the culture does not adjust to the requirements of the initiative and the initiative is likely to fail [13]. In our empirical study, we deploy the existing measurement instrument to identify possible shortcomings of the culture in the case organization. Based on this assessment, we then more deeply into the concept of cultural fit by also analyzing how an organization can realize a fit between their cultural context and BPM culture.

3 Empirical Approach

3.1 Case Organization: *ITLeader*

In order to identify strategies for developing an organizational culture supportive of BPM, we look into the specific case of the global IT company *ITLeader* (name changed) which provides software for supporting business processes. The corporation provides a suitable background for identifying how a BPM-facilitating organizational culture can be developed, not only based on their products and services but primarily based on their internal Process Excellence (PE) initiative.

ITLeader is a large-sized enterprise software organization. The headquarter is in Europe, but the 5.500 employees are situated worldwide and serve 10.000 enterprise and public institutions across more than 50 countries. *ITLeader* is one of the 25 largest business software vendors in the world with revenue of around €1 billion in 2011.

The PE initiative in *ITLeader* started in 2010 and is lead by the Audit, Process and Quality (P&Q) team. The defined goals were to identify and maintain a company-wide process landscape; set up and implement the concept of "Process Ownership"; identify and implement standardized target processes; digitize processes using own technology; implement process intelligence; monitor the cultural change and enable continuous improvement. Even though several process improvement projects were conducted and are still planned, the optimization of the product development lifecycle namely *Idea to Product Release* end-to-end process is critical for *ITLeader*'s success. This process is called internally Entire Readiness and is a first priority process in *ITLeader*'s PE initiative. In the following, we introduce the methodological procedure of our empirical analysis at *ITLeader*.

3.2 Methodological Procedure

The purpose of the study is to generate first insights on strategies that organizations can take to develop a BPM supportive organizational culture. Due to the exploratory nature of this study, we chose a qualitative research design as an appropriate lense for this research endeavor [38]. Since the investigation is adressing a contemporary topic which is not heavily discussed in research yet and should be studied in a natural setting [39], interpretive case study is applied as a research method. In our empirical study, the unit of analysis is the *Idea to Product Release* process at *ITLeader* since it represents the central business processes in the PE initiative. In order to examine how *ITLeader* can develop a culture that supports their BPM initiative, the study concentrates on the following two phases:

1. First of all, we analyze the culture development needs of *ITLeader*. For this purpose, we use a previously validated instrument to measure the presence of the BPM-supportive CERT values in the organization [13]. This allows us to gain an initial overview how far the organization is culturally fit for their BPM approach. Specifically, employees involved in the *Idea to Product Release* process were asked to assess their culture with the measurement instrument which we implemented as an online-survey instrument. The data collected from 27 respondents was analyzed in SPSS using average item scores. As the result, the perceived shortcomings of the prevailing culture served as a starting point for identifying strategies on how to further develop the supportiveness of the existing culture for BPM.
2. Second, we conducted five semi-structured interviews with an average length of 45min. The outcome of the assessment from the previous phase served as a basis to discuss with the interviewees how a BPM-supportive culture could be achieved. Based on this, we developed an interview guideline that contained questions adressing all CERT values. For each value dimension, we discussed ways of realizing a culture that supports achieving efficient and effective business processes. The

interviewees were key representatives of the *Idea to Product Release* process such as Project Leader R&D, Senior Vice President Process Quality&Audit, and Quality Manager. The interviews were transcribed and coded with NVivo. The findings from the interviews provide first insights on the activities which can be used to develop a culture that facilitates BPM.

4 Identifying Culture Development Strategies

4.1 Analyzing Culture Development Needs

We use an existing measurement instrument to assess the extent to which employees of *ITLeader* perceive that the CERT values Customer Orientation, Excellence, Responsibility, and Teamwork, respectively the eight cultural subdimensions are lived in the corporation [13]. The findings give first insights on the cultural fitness of the organization for BPM and therefore provide a starting point for examining how a BPM-supportive organizational culture can be realized. Figure 2 gives an overview of the results. While the assessment shows that there is overall still much room to improve the supportiveness of the organizational culture, it becomes immediately obvious that particularly the two cultural dimensions *internal customer orientation* and *accountability* are perceived present in the company below-average. These findings provide a basis for developing strategies how to improve the prevailing culture at *ITLeader*.

Our interview partners confirmed that employees working at the same process, often do not think in terms of *internal customers*. The heads of the initiative received feedback that “it is the first time in these [PE] projects that people from different silos have spoken to each other”. Internal customer orientation demands a shift in thinking, as the interviewee perceives: “The risk if you have processes or formal processes in place is that people say - oh, that is the process and I stick to it, and I exactly did what this process tells me, but I am not looking left and right. And perhaps, you need to look left and right even if it is not in the box that outlines what you should do.”

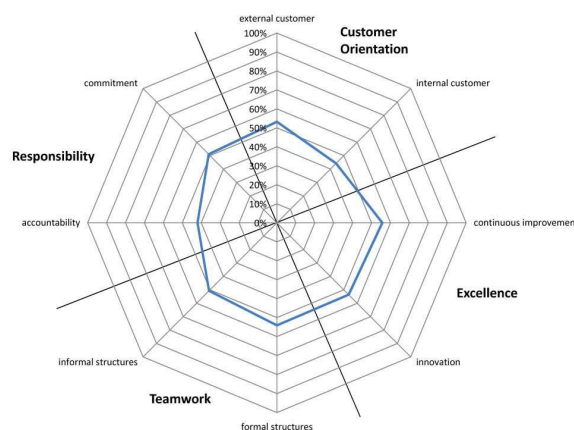


Fig. 2. Cultural Fitness of *ITLeader* for BPM

Regarding *accountability*, which is one of the main goals of the PE initiative, the situation is described as follows: “We could also organize it differently but in our department in our company it’s really that there’s not a single person or department who would be somehow seen as the owner [...] of this process.”; “[W]e also do not have this horizontal really leader role of a process owner. It’s more somebody who does a little bit of the organization”. Moreover, the interviewee regards this deficiency as a part of the current culture “if that is more a culture in the company that you have synchronization, coordination between the department then it’s really difficult to switch to another model with a defined process and the process owner.”

The application of the measurement instrument to analyze the existing organizational culture shows that the instrument is suitable to identify an overall culture development strategy for the organization. Based on the findings which are also confirmed by our interview partners, it can be concluded that activities are required in all eight subdimensions to increase the supportiveness of *ITLeader*’s culture for BPM. Particular attention is required in the two dimensions *internal customer orientation* and *accountability*. In the following, we present culture development activities which we derived on the basis of our empirical study with the organization.

4.2 Deriving Activities to Develop a BPM-supportive Culture

Based on the interviews, we identified several activities which can support *ITLeader* in developing a more BPM-supportive culture. Regarding the realization of such a culture, we present identified activities for all CERT values and their subdimensions in the following. Beyond, we report on some general activities which our interviewees recognized as important drivers for the intended culture change.

Customer Orientation - External Customer

- *Practicing what we preach* is considered essential for the external customer orientation in *ITLeader*. Therefore, the PE team aims at promoting their activities through documented case studies and lessons learned. This knowledge is supposed to provide an important reference point for the employees who then apply the gained knowledge in the contact to the end customer. In this way, there is an intuitive process of learning and sharing own experiences with external customers.
- *One face to the customer* in the delivery is seen as another important side of the relation to the external customer. In this regard, *ITLeader* aims at creating comprehensive offers for the customers and providing the delivery of services in a uniform way. Currently, different product lines have different delivery processes, including separate order and payment possibilities for different products.

Customer Orientation - Internal Customer

- *Trainings* are perceived necessary to form a common understanding of goals between the employees in one business process. Since they don’t have the same view on the process, it is highly relevant to break down the functional silos in the very beginning of the PE initiative and facilitate thinking in terms of internal customers.

- *Involvement* is considered as a critical factor to institutionalize internal customer orientation in the culture of *ITLeader*. This means the voice of the (internal) customer should not only be heard and quality improvements should not only be reflected and implemented on that basis, but it is perceived critical to stay in dialogue with the customer and have potential changes confirmed prior to their implementation. The repeated feedback of internal customers helps avoiding misinterpretations and the realization of unwanted changes. As the Senior Vice President mentions “it means that in case we think about something has to be changed but it is not confirmed by the internal customer, we will not do it. So, this is the reason why I said that everything, the whole initiative is only customer driven.”

Excellence - Continuous Improvement

- *Individual incentive systems* are considered an important means for motivating employees to achieve process goals: “The most important thing that we need to achieve - we have initiated that but it is not fully executed, and this will take us some more years is - putting in process goals, process related goals into the MBO, into the goal achievement of every employee.”
- *Monitoring process performance* is seen as a crucial activity to realize the continuous improvement of business processes. The KPIs should be clearly defined and improvements should be reported in the workflow management application. The fact that the PE team at *ITLeader* receives increasingly more requests is seen as thinking “out of the box” and “departmental silos, and contact us because they have ideas to make the process better. And this is a cultural change we encounter, especially during the last years very much.”

Excellence - Innovation

- *Simple proposal procedures* for process innovation ideas are seen as essential for motivating employees to suggest such innovations. Currently, employees of *ITLeader* need to go through a formal process that requires to know details about the cost centers, project descriptions, organizations effected, financial impact, and similar. As a consequence, ideas are currently communicated in an unstructured way which makes them difficult to handle. A more simple proposal procedure would allow for a structured handling of the ideas. Those suggestions that shall be realized can still be specified later on.
- *Following-up on ideas* for product innovation is considered to be critical for keeping employees creative. At *ITLeader*, the idea generation process was supported by a tool, however, the impression of the employees was not enough care is taken of the suggested innovation ideas. As the Quality Manager explains, “the system itself was not bad, it was just the people who should look at the ideas”. To motivate employees to participate, it is essential to provide feedback on suggestions and manage this process as well.

Responsibility – Accountability

- *Creating awareness* for the role of process owners in cross-departmental projects is seen as a highly relevant task. Not only should employees naturally think of this role in such projects, but the assigned process owner should also support the acceptance of the project in the involved business departments. The foundation of process owner as a role is an essential step for promoting the process orientation.
- *Setting up clear governance rules* is considered an important steering mechanism when it comes to process accountabilities. Particularly, the specification of different roles, such as process sponsor, process owner, and process expert, and the corresponding responsibilities are required. According to the Senior Manager, “By having this split between responsibility and accountability, by having this process ownership, process expert and process sponsor, we can look at the different activities, for one process in a detailed view.” In case of disagreements between the functional heads and process owners, ways of resolving conflicts need to be defined. Established governance structures should be also reflected in documentation available to all employees, such as in process models enriched by RACI (Responsible, Accountable, Consulted and Informed) matrix.

Responsibility - Commitment

- *Motivating through good examples* is understood as an essential driver of individual commitment to process objectives. “[W]hen you look at the process you always have very many functions or departments involved, and this is a risk for everyone because it is not yet so solid in the minds of the people that they can rely on their process chain stakeholders and we have to prove with good projects and good lessons learned that it is worth setting process goals and being optimistic and stick to MBO with process goals.”
- *Rewarding dedication* of employees towards attaining process goals is seen as an important motivator for commitment. The intrinsic motivation of employees is one of the central drivers of goal achievement. The promotion and rewarding of dedicated employees or functions may further stimulate the commitment of others.

Teamwork - Informal Structures

- *Facilitating face-to-face meetings* is considered essential for fostering cross-departmental teamwork. Currently, people often know each other only through impersonal phone or web meetings. It is particularly important to enable teams to build up informal relationships, even more in case of new employees who are not familiar with the culture and working modes yet.
- *Fostering open communication* between departments is seen critical for a comfortable cross-functional working atmosphere. Currently, many different (social) platforms are in place at *ITLeader*. Having one platform which is used by all employees would allow more transparency and would help overcome functional barriers.

Teamwork - Formal Structures

- *Establishing cross-functional meetings* is understood as an important structure to foster thinking beyond departmental boundaries. “[W]hen you think about what are the problems in processes, it is often, or in most cases that the problems occurred between the interfaces, so when the process is cross-functional. [...] To get a common understanding, you need to break these silos [...] in the very beginning.”
- *Defining deliverables* between departments is seen essential for the teamwork between functions. Therefore, service level agreements between different departments in the organization help to structure and achieve the overall common goals of the business processes.

General Culture Development Activities

- *Monitoring the organizational culture* on a regular basis is perceived as a critical factor to sustainable culture development. Particularly in the intensive phase of the PE initiative at *ITLeader*, a continuous evaluation of the organizational culture based on employees’ perceptions should facilitate the intended cultural change.
- *Choosing employees who embrace the desired cultural values* is seen as an important long-term activity to develop the organizational culture. Once project teams share the same cultural values, they stabilize business processes and support maintaining a BPM facilitating organizational culture.

In this section, the derived strategy to address the identified cultural development needs of the case organization was presented. The assessment of BPM-supportive cultural values, more precisely CERT values, was performed by means of a survey. Building on the analysis results, specific guidelines to realize those values were derived based on in-depth interviews for all cultural subdimensions. Considering the tight relation between these empirical findings and theoretical background, we further explicate on particular implications for research and practice.

5 Discussion

The identified activities for developing a BPM-supportive organizational culture in *ITLeader* are linked to the previously introduced BPM-Culture-Model. The culture development strategies provide first insights into how organizations can achieve a cultural fit. As indicated before, understanding the concept of cultural fit is important to study how BPM culture can be facilitated in an organization. To do so, we analyzed the as-is culture in the organizations, examined reasons for shortcomings and derived guidelines for development derived.

5.1 Implications for Research

The study contains several implications for research. First of all, addressing the identified research gap on how to develop a culture that supports BPM extends and

specifies the BPM-Culture-Model of vom Brocke & Sinnl [9]. Despite the fact that the model presents the concept of cultural fit (between BPM culture and cultural context) as the critical element to achieve BPM success, the concept has not been specified beyond a general definition.

While our study certainly does not comprehensively address this research gap, it represents an important first step towards examining how cultural fit can be achieved. Our study provides initial insights into the facets that determine the concept of cultural fit. These may support future research in theorizing how to achieve a cultural fit.

Moreover, the application of the instrument which measures BPM-supportive CERT values showed that the instrument is suitable to identify an overall culture development strategy for organizations [13]. While this is one of the intended use cases of the instrument, it had not been applied to a single case organization to date.

5.2 Implications for Practice

Our study also contains implications for practice. First of all, the research result shows how the use of the CERT values measurement instrument can support organizations in identifying overall culture development strategies. It provides an overview for organizations how far their culture is already supportive of their BPM approach. Based on the perceptions of the eight cultural subdimensions, the organization can identify and focus on exact areas for culture development, and furthermore, they can derive specific activities to improve their culture in the identified dimensions.

Beyond, our research also provides first insights into what these specific activities could look like. Even though it still needs to be assessed how far these findings are transferable to other contexts, our study revealed initial guidelines on how to realize a BPM-supportive organizational culture. These may serve as a starting point for organizations to derive more context-specific guidelines for their own purposes.

5.3 Limitations

We are well aware that our study contains several limitations. First of all, analyzing a single case implies that the generalizability of the results is restricted. Multiple context factors of the examined case could have significant effects on the results. These factors may include the phase of the BPM initiative, the process lifecycle, the industry sector, etc. However, it was not our intention to identify a final set of activities organizations can use to develop their culture but rather to provide first insights on how organizations can derive strategies to implement a BPM-supportive culture.

Besides, our findings are limited to the perceptions of the involved survey participants and interview partners. Involving employees from other processes of the organization or even the entire organization may have yielded different results. Again, for the specific context we focused on, our study provides an initial understanding how organizational culture can be developed in order to facilitate BPM. In this regard, our study contributes to the existing body of knowledge on BPM.

6 Conclusion

Our study aimed at addressing the research question “*How can an organization develop a culture supportive of BPM?*” The research process encompassed both a review of existing research in the field and an empirical investigation of the topic. Particularly, we examined cultural development needs of a specific organization using an existing measurement instrument. On this basis, we conducted in-depth interviews with employees from this organization to derive activities which may help the organization in developing an organizational culture that supports their BPM approach.

In doing so, the findings deepen recent findings and point to facets of achieving a cultural fit. These results extend recent research in the field, particularly the theoretical findings on BPM-Culture-Model and CERT values. However, our findings only provide initial insights to the development of BPM supportive organizational culture and the suggestion of specific guidelines. Further research is still needed to specify these results in different contexts.

References

1. de Bruin, T., Rosemann, M.: Using the Delphi Technique to Identify BPM Capability Areas. In: ACIS 2007 Proceedings, paper 42 (2007)
2. Rosemann, M., vom Brocke, J.: The Six Core Elements of Business Process Management. In: vom Brocke, J., Rosemann, M. (eds.): Handbook on Business Process Management 1. Springer (2010)
3. Alibabaei, A., Bandara, W., Aghdasi, M.: Means Of Achieving Business Process Management Success Factors. In: MCIS 2009 Proceedings (2009)
4. Zairi, M.: Business process management: a boundaryless approach to modern competitiveness. *Business Process Management Journal* 3, 64–80 (1997)
5. Baumöl, U.: Cultural Change in Process Management. In: vom Brocke, J., Rosemann, M. (eds.): Handbook on Business Process Management 2. Springer (2010)
6. Dixon, J., Jones, T.: Gartner Hype Cycle for Business Process Management (2011)
7. Schein, H.E.: *Organizational Culture and Leadership*. John Wiley & Sons (2004)
8. Leidner, D.E., Kayworth, T.: Review: A Review of Culture in Information Systems Research: Towards a Theory of Information Technology Culture Conflict. *MIS Quarterly* 30, 357–399 (2006)
9. vom Brocke, J., Sinnl, T.: Culture in business process management: a literature review. *Business Process Management Journal* 17, 357–378 (2011)
10. Kappos, A., Rivard, S.: A Three-Perspective Model of Culture, Information Systems, and Their Development and Use (2008)
11. Al-Mashari, M., Zairi, M.: BPR implementation process: an analysis of key success and failure factors. *Business Process Management Journal* 5, 87–112 (1999)
12. Schmiedel, T., vom Brocke, J., Recker, J.: Which cultural values matter to business process management? Results from a global Delphi study. *Business Process Management Journal* 19 (2) (2013)
13. Schmiedel, T., vom Brocke, J., Recker, J.: Is your Organizational Culture Fit for Business Process Management?. *Business Process Trends*, 1–5 (2012)
14. vom Brocke, J., Becker, J., Braccini, A.M., Butleris, R., Hofreiter, B., Kapočius, K., De Marco, M., Schmidt, G., Seidel, S., Simons, A., Skopal, T., Stein, A., Steiglitz, S., Suomi,

- R., Vossen, G., Winter, R., Wrycza, S.: Current and future issues in BPM research: a European perspective from the ERCIS meeting 2010. *Communications of the Association for Information Systems*, 393–414 (2011)
15. Hofstede, G.H., Hofstede, G.J.: *Cultures and Organizations: Software of the Mind*. (2005)
 16. Cooper, H.M.: Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society* 1, 104–126 (1988)
 17. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: 17th European Conference on Information Systems, pp. 1–13 (2009)
 18. vom Brocke, J., Schmiedel, T.: Towards A Conceptualisation Of BPM-Culture: Results From A Literature Review. In: PACIS 2011 Proceedings (2011)
 19. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* 26 (2), pp. xiii–xxiii (2002)
 20. Philip, G., McKeown, I.: Business Transformation and Organizational Culture: The Role of Competency, IS and TQM. *European Management Journal* 22, 624–636 (2004)
 21. Yong, K.T., Pheng, L.S.: Organizational culture and TQM implementation in construction firms in Singapore. *Construction Management and Economics* 26, 237–248 (2008)
 22. Guha, S., Grover, V., Kettinger, W.J., Teng, J.T.C.: Business process change and organizational performance: exploring an antecedent model. *Journal of Management Information Systems* 14, 119–154 (1997)
 23. Motwani, J., Kumar, A., Antony, J.: A business process change framework for examining the implementation of six sigma: a case study of Dow Chemicals. *The TQM Magazine* 16, 273–283 (2004)
 24. Motwani, J., Mirchandani, D., Madan, M., Gunasekaran, A.: Successful implementation of ERP projects: Evidence from two case studies. *International Journal of Production Economics* 75, 83–96 (2002)
 25. Motwani, J., Subramanian, R., Gopalakrishna, P.: Critical factors for successful ERP implementation: Exploratory findings from four case studies. *Computers in Industry* 56, 529–544 (2005)
 26. Nah, F.F.-H., Lau, J.L.-S., Kuang, J.: Critical factors for successful implementation of enterprise systems. *Business Process Management Journal* 7, 285–296 (2001)
 27. vom Brocke, J., Sinnl, T.: Applying the BPM-Culture-Model – The Hilti Case. In: 21st Australasian Conference on Information Systems, paper 16 (2010)
 28. Rosemann, M., de Bruin, T.: Towards a business process management maturity model. In: ECIS 2005 Proceedings, pp. 521–532 (2005)
 29. Rosemann, M., de Bruin, T., Hueffner, T.: A Model for Business Process Management Maturity. In: ACIS 2004 Proceedings, paper 6 (2004)
 30. Hammer, M.: The process audit. *Harvard Business Review* 85, 111–119, 122–123, 142 (2007)
 31. Mosadegh Rad, A.M.: The impact of organizational culture on the successful implementation of total quality management. *The TQM Magazine* 18, 606–625 (2006)
 32. Zucchi, F., Edwards, J.S.: Human resource management aspects of business process reengineering: a survey. *Business Process Management Journal* 5, 325–344 (1999)
 33. Baird, K., Hu, K.J., Reeve, R.: The relationships between organizational culture, total quality management practices and operational performance. *International Journal of Operations & Production Management* 31, 789–814 (2011)
 34. Revenaugh, D.L.: Business Process Re-engineering: The Unavoidable Challenge. *Management Decision* 32, 16–27 (1994)

35. Kappos, A., Croteau, A.-M.: Organizational Change and Culture: Insights on BPR Projects. In: AMCIS Proceedings, paper 284 (2002)
36. Cao, G., Clarke, S., Lehaney, B.: A critique of BPR from a holistic perspective. *Business Process Management Journal* 7, 332–339 (2001)
37. Kohlbacher, M., Gruenwald, S.: Process orientation: conceptualization and measurement. *Business Process Management Journal* 17, 267–283 (2011)
38. Creswell, J.W.: *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Sage Publications Ltd. (2012)
39. Benbasat, I., Goldstein, D.K., Mead, M.: The Case Research Strategy in Studies of Information Systems. *MIS Quarterly* 11, 369 (1987)

IT-basierte Unterstützung virtueller und realer Selbsthilfegemeinschaften in Katastrophenlagen

Oliver Heger und Christian Reuter

Universität Siegen, Institut für Wirtschaftsinformatik, Siegen, Germany
{oliver.heger, christian.reuter}@uni-siegen.de

Abstract. Social Media wird zunehmend von Betroffenen und Selbsthilfegemeinschaften in Katastrophenlagen genutzt. In diesem Beitrag werden Studien zu ‚virtuellen‘ und ‚realen‘ Selbsthilfeaktivitäten dargestellt. Untersucht wurden zum einen ‚virtuelle‘ Selbsthilfeaktivitäten in Twitter während der Tornadokatastrophe in den USA am 27./28.4.2011 mit besonderem Fokus auf die zeitlichen Verläufe und Nutzergruppen; zum anderen ‚reale‘ Selbsthilfe anhand Interviews mit Behörden und Organisationen mit Sicherheitsaufgaben (BOS). Hierauf aufbauend werden Gestaltungsaspekte und Anforderungen für Social Media zur Unterstützung von Krisenhelfern bzw. Selbst- und Nachbarschaftshilfe und zur Kombination von virtuellen und realen Aktivitäten im Katastrophenmanagement diskutiert.

Keywords: Social Media, CSCW, Katastrophenmanagement, Selbsthilfe, Nachbarschaftshilfe, Twitter

1 Einleitung

Das Ausmaß von Katastrophen, wie dem Love-Parade-Unglück in Duisburg 2010, der Tsunami- und Erdbebenkatastrophe in Japan 2011 oder dem Hurrikan Sandy in den USA 2012 ist so weitreichend, dass die für die Bewältigung zuständigen Behörden und Organisationen oftmals überfordert sind und aus diesem Grund dringende und lebensrettende Maßnahmen teilweise nicht schnell genug durchgeführt werden. Umso wichtiger ist es daher, dass sich diejenigen Menschen, die nicht am offiziellen Katastrophenmanagement beteiligt sind – das sind vor allen Dingen die betroffenen Bürger – selbst helfen können. Selbsthilfegemeinschaften (SHG) bestehen aus Zivilpersonen, die zusammenarbeiten, um gemeinsame Ziele hinsichtlich akuter oder potenzieller Katastrophen zu verfolgen, deren Organisation aber noch nicht institutionalisiert wurde [1]. Gemäß der Krisen-Kommunikationsmatrix [2] stellt dies die Kommunikation unter Akteuren der allgemeinen Öffentlichkeit, also den (betroffenen) Bürgern dar. Für die Entstehung solcher Gemeinschaften sind laut Quarantelli [3] fünf Faktoren entscheidend: (a) ein übergemeinschaftliches Umfeld, das die SHG legitimiert; (b) ein entscheidendes Ereignis, das als Gefahr erkannt wird; (c) ein unterstützendes soziales Klima mit positiven Werten, Normen und Überzeugungen hinsichtlich der Notwendigkeit gemeinschaftlichen Handelns; (d) ein bestehendes soziales Netzwerk, sodass

Kommunikation stattfinden kann; (e) und verfügbare Ressourcen, wie z. B. Informationen, Wissen oder Fähigkeiten. Entgegen anders lautender Mythen verfallen Bürger in betroffenen Gebieten selten in Panik, sind nicht hilflos und abhängig von externen Rettungskräften und plündern nicht [4]. Sie sind meist die ersten, die sich um Opfer kümmern und Such- und Rettungsaktionen durchführen. Die primäre Motivation, sich an einer SHG zu beteiligen, ist ein starker Drang, zu helfen, Opfern beizustehen und der Wunsch, „irgendetwas zu tun“ [5]. Es können drei Gruppen unterschieden werden: Der kleine aktive Kern, (~1%); der größere Kreis von Unterstützern für spezifische Aufgaben (~10%); und die große Gruppe von nominalen Unterstützern, die gelegentlich mithelfen und eher passiv sind (~90%) [3]. Abhängig von den Fähigkeiten findet eine Arbeitsteilung unter den Beteiligten statt. Hierarchien sind entweder kaum vorhanden oder sehr flach angelegt. Die Führung der Gruppe kristallisiert sich langsam und informell heraus [3]. Zur Koordination bedarf es einer intensiven Kommunikation zwischen den Mitgliedern. Laut Voorhees [6] baut das Kommunikationssystem einer SHG auf den sog. „semi-permanent volunteers“ auf, also den Beteiligten, die bei den Aktivitäten regelmäßig anwesend sind und vermutlich zum aktiven Kern gehören. Nach einer gewissen Zeit können sich diese Mitglieder gegenseitig erkennen und Kommunikationskanäle schaffen, sodass ein informelles Netzwerk entsteht. In den vergangenen Jahren hat die Evolution von Social Software Infrastrukturen für diese informellen Netzwerke geschaffen und virtuelle Aktivitäten ermöglicht [7].

So wie es üblich ist, reale und virtuelle Communities zu unterscheiden, erscheint auch bei SHG eine solche Unterscheidung sinnvoll: *Virtuelle SHG* haben ihren Ursprung im Internet und führen ihre Aktivitäten hauptsächlich online durch. *Reale SHG* treten in Form von Nachbarschaftshilfe auf und bekämpfen die Auswirkungen einer Katastrophe vor Ort. Sie nutzen das Internet als ein Unterstützungsmittel, sehen es jedoch nicht als ihr Zentrum an. Ein wesentliches gemeinsames Merkmal ist ihre emergente Struktur, die sich durch eine Katastrophenlage ad hoc mit dem Ziel herausbildet, deren Auswirkungen zu bekämpfen.

Die bisherige Forschung hat reale und virtuelle Selbsthilfephänomene unter den Begriffen „emergent groups“ [1] und „digital volunteers“ [8] getrennt voneinander betrachtet. Im Fokus haben einerseits die Verhaltens- und Funktionsweisen von SHG und andererseits die Verwendung von IT, insbesondere Twitter, gestanden. Die begriffliche Unterscheidung von realen und virtuellen SHG, die diesem Beitrag zugrunde liegt, erkennt einerseits an, dass beide Formen Unterschiede in ihren Schwerpunkten und ihrer Arbeitsweise aufweisen, erlaubt jedoch andererseits eine kombinierte Betrachtung, um Gemeinsamkeiten und Verbindungen aufzudecken.

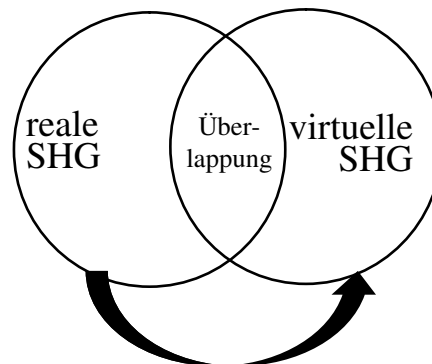


Abb. 1: Reale und virtuelle Selbsthilfegemeinschaften

Die Forschungsfrage in diesem Beitrag lautet: Welchen Beitrag kann IT in der Zusammenarbeit virtueller und realer Selbsthilfegemeinschaften leisten? Um mögliche Punkte der Kooperationen realer und virtueller SHG zu analysieren, werden in diesem Beitrag zunächst beide Formen einzeln untersucht und anschließend ihre Verknüpfung diskutiert. Der erste Teil der folgenden empirischen Untersuchung widmet sich vor allem virtuellen SHG, indem Twitterdaten, die während einer Katastrophe entstanden sind, ausgewertet werden. Da Twitter in den USA umfassender genutzt wird, wurde eine dort verortete Katastrophe gewählt. Der zweite Teil der Untersuchung analysiert Interviews mit Akteuren aus Behörden und Organisationen mit Sicherheitsaufgaben (BOS) und konzentriert sich dabei auf reale SHG. Die daran anschließende Diskussion geht auf Gemeinsamkeiten und Unterschiede und mögliche IT-Implicationen zur Förderung von Selbsthilfe ein.

2 Twitter-Studie: Virtuelle Selbsthilfeaktivitäten am Beispiel der Tornado katastrophe in den USA (27./28.4.2011)

2.1 Grundlagen: Virtuelle Selbsthilfe und Social Media

Seit einigen Jahren werden das Internet und Social Media verstärkt zur Kooperation und Kommunikation verwendet. Untersuchungen zeigen, dass *Microblogging* während Katastrophen eingesetzt wird, um Informationen zu sammeln sowie zu verteilen, zu kommunizieren und Hilfsanfragen zu bearbeiten [8]. Twitter dient als Quelle für Situationsupdates [9] und als Plattform für die Koordination von Aktionen, für den Meinungsaustausch und für die emotionale Verarbeitung einer Katastrophe [10]. Für eine intensivere Koordinationsarbeit wird auf andere Software, wie z. B. Skype, ausgewichen [8]. *Soziale Netzwerke* bieten ihren Nutzern, die mittels sogenannter Profile vertreten und mit anderen Nutzern verbunden sind, vielfältige Funktionalitäten an, wie z. B. das Senden von Nachrichten, das Teilen von Fotos und Videos, das Bereitstellen von Informationen, das Melden des aktuellen Status, das Ankündigen von Veranstaltungen und das Diskutieren in Foren [11]. Existierende soziale Netzwerke, wie Facebook, bieten den Vorteil, dass ihre Funktionalitäten in Katastrophen nicht neu erlernt werden müssen und sie bereits vor deren Eintreten ein soziales Geflecht aufweisen [12]. *Wikis*, wie Emergency Wiki, Quake Help Wiki oder Scipionus dienen der kollaborativen Sammlung von Informationen und Wissen [7]. Sie sind besonders nützlich, um kollektive Intelligenz zu schaffen, haben jedoch Defizite in der Kommunikation und folglich auch der Koordination. Sahana, Ushahidi und Google Crisis Map sind Beispiele für *katastrophenspezifische Plattformen*, die speziell gestaltet wurden und mehrere webbasierte Applikationen integrieren.

Für die Erforschung von virtuellen SHG eignet sich besonders Twitter, da diese Plattform in Katastrophen intensiv genutzt wird und die Kommunikation ihrer User zumeist öffentlich stattfindet. Viele Studien untersuchen die Nutzung von Twitter in Katastrophen als Informations- und Kommunikationsplattform. In der Betrachtung der Verhaltensweisen von Twitterern wurden sogenannte 'information brokers' [13] identifiziert, die Informationen von verschiedenen Quellen sammeln um den Betrof-

fenen zu helfen. In solchen Situationen spielen ‚broadcast‘ (Übermittlung an alle Teilnehmer eines Netzes) und ‚brokerage‘ (Informationsvermittlung) eine größere Rolle als in der alltäglichen Twitter-Nutzung [13]. Die Informationen selbst können in vier Typen unterschieden werden [14]: Die generativen, synthetischen, derivativen und innovativen Informationen. ‚Retweets‘, die Wiederholungen einer Nachricht eines anderen Users, gelten als wichtiges Weiterempfehlungsinstrument, und sind den derivativen Informationen zuzuordnen [15].

Ziel der folgenden empirischen Analyse von Twitterdaten ist es, die aktuelle Forschung um die Punkte zeitliche Entwicklung des Aktivitätsfokus und eine Kategorisierung von Nutzertypen zu erweitern, um Zeitpunkte und Nutzertypen für die Kooperation realer und virtueller Gruppen zu identifizieren.

2.2 Methode

Am Mittwoch, 27.04.2011, und am darauffolgenden Tag wüteten in den USA nach Schätzungen des US-Wetterdienstes 211 einzelne Tornados. Mehr als 340 Menschen kamen durch diese Tornadokatastrophe ums Leben. Auch erhebliche materielle Schäden an Häusern, Straßen und Autos wurden gemeldet. In Alabama hatten etwa eine Millionen Menschen keinen Strom. Aufgrund dieser katastrophalen Lage suchten in vielen Städten Rettungskräfte nach Toten und Überlebenden. Die Spitze der Tornado-katastrophe fand am Mittwoch statt. Am Donnerstag wurden weitere sintflutartige Regenfälle und Tornados verzeichnet. Die zu analysierenden Twitterdaten stammen von Donnerstag, als noch einzelne Tornados wüteten und es „tornado warnings“, d. h. Warnungen vor existierenden Tornados und „tornado watches“ gab, die die Bevölkerung aufgrund der Wetterlage vor möglichen Tornados warnt. Die Anzahl und die geographische Ausbreitung der warnings und watches ließen im Tagesverlauf nach.

Mithilfe des Programms „The Archivist“ (<http://visitmix.com/work/archivist-desktop/>), das die Twitter Search API verwendet, wurden mit dem Suchbegriff „Tornado“ Twitter-Einträge erhoben. Um eine möglichst breite Palette an unterschiedlichen Aktivitäten zu erfassen, wurde ein Zeitraum gewählt, der der Warnphase mit aktuellen Tornadowarnungen, der Notfallphase mit gegenwärtigen Tornados und der Wiederaufbauphase in den Gegenden, in denen Tornados schon am Mittwoch gewütet haben, entspricht. Da die Tweets keine geographischen Daten enthalten und die einzelnen Phasen überlappen, können die Phasen zeitlich nicht voneinander abgegrenzt werden. Insgesamt liegen 79.318 sogenannte Tweets vor, die vom 28.04.2011, 12:17 (EDT) bis 29.04.2011, 03:16 (EDT) von 59.282 unterschiedlichen Usern verfasst wurden. Der Zeitraum entspricht der Warnphase mit aktuellen Tornadowarnungen, der Notfallphase mit gegenwärtigen Tornados und der Wiederaufbauphase in den Gegenden, in denen Tornados schon am Mittwoch gewütet haben. Die Tweets liegen in Form einer Excel-Tabelle vor; Angaben zum Usernamen, zum Veröffentlichungsdatum und -uhrzeit der jeweiligen Tweets sind vorhanden. Die Abkürzung RT innerhalb einer Nachricht kennzeichnet, ob sie Retweets enthält bzw. ein Retweet ist. Der Ausdruck @Benutzername gibt an, an wen eine Nachricht adressiert ist. Da dieser Beitrag neuartige Anforderungen zur Unterstützung von SHG erforschen soll, wird folgenden Fragen nachgegangen: (1) Welche zeitlichen Entwicklungen und Verschie-

bungen können beobachtet werden? (2) Welche Inhalte werden getweetet? Und darauf aufbauend: (3) Welche Verhaltensweisen der Twitter-Nutzer können unterschieden werden?

Zeitliche Entwicklungen. Um zeitliche Entwicklungen und Unterschiede deutlich machen zu können, werden die Nachrichten in einstündige Intervalle eingeteilt. Durch den 15-stündigen Untersuchungszeitraum ergeben sich also 15 Intervalle. Dabei werden sämtliche Nachrichten nach Schlüsselwörtern durchsucht, die einen Teilaspekt des Nachrichteninhalts wiedergeben können. Erfasst werden nicht nur das Wortselbst, sondern auch Wörter, die das gesuchte Wort enthalten. Beispielsweise wird das Wort „helper“ beim Suchbegriff „help“ mitberücksichtigt.

Qualitative Inhaltsanalyse. Zum Zwecke eines besseren Verständnisses über die Arten der Selbsthilfeaktivitäten in Twitter wurden neben der Verwendung quantitativer analytischer Verfahren mehrere Hundert Tweets selektiert, gelesen und manuell mithilfe von qualitativer Codierung in Gruppen gleichen Inhalts klassifiziert. Hier werden exemplarische Tweets gezeigt, die den Inhalt widerspiegeln.

Rollenanalyse. Nach der Betrachtung zeitlicher Entwicklungen und der Inhalte werden die Nutzer selbst kategorisiert und deren Rollen beschrieben. Eine solche Betrachtung fehlt in anderen Studien. Aus Gründen der Nachrichtenmenge wurden nicht sämtliche User einzeln durchleuchtet und eingeteilt, sondern die Aktivitäten derjenigen User untersucht, die sich durch eine besonders hohe Anzahl an Tweets auszeichnen und wahrscheinlich zum aktiven Kern gehören [3]. Diese wurden Schritt für Schritt bis zu einer subjektiv festgestellten Sättigung, bei der keine grundlegend neuen Charakteristika mehr zu erwarten waren, erfasst. Es werden daher die 41 User mit den meisten Tweets (1982 Tweets, 2,50% aller Tweets) und die 51 User, die am häufigsten retweetet werden (7742 Retweets, 22,32% aller Retweets), analysiert. Die letztgenannte Zahl zeigt, dass sich sehr viele Retweets, die unter anderem als Weiterempfehlungssystem für Informationen und Personen genutzt werden, auf sehr wenige Nutzer fokussieren, deren Informationen vermutlich besonders wertvoll sind. Insgesamt werden 85 der auffälligsten User analysiert. Sie ergeben nicht die Summe aus 41 und 52, da es Überschneidungen gibt. Es wurde Open Coding [16] verwendet, um das Material zu analysieren und auffällige Muster zu betrachten. Anschließend wurde versucht, verschiedene Kombinationen, die die Verhaltensweisen oder Rollen charakterisieren, unter einem beschreibenden Titel zusammenzufassen.

2.3 Ergebnisse (1): Zeitliche Entwicklungen

Zunächst kann festgestellt werden, dass die Anzahl der Tweets fast stetig abnimmt. Während am frühen Nachmittag noch über 10.000 Nachrichten in einer Stunde getwittert werden, werden am frühen Abend nur noch 5494 Einträge pro Stunde und zwischen 2:17 und 3:16 1209 Einträge veröffentlicht. Dies geht mit der oben beschriebenen Tatsache einher, dass die Anzahl der Tornados und Tornadowarnungen im Tagesverlauf sinkt. Je höher also die Gefahr bzw. die Gefahrenwahrnehmung ist, desto aktiver sind virtuelle SHG. Die abnehmende Anzahl der Tornadowarnungen spiegelt sich auch dadurch wider, dass die Begriffe „warning“ und „tornado watches“ in den Tweets immer seltener vorkommen (Abbildung 2). Die gegenteilige Entwick-

lung kann man feststellen, wenn man nach dem Wort „help“ sucht. Aus diesen beiden gegensätzlichen Verläufen der Warn- und Hilfsaktivitäten kann gefolgert werden, dass eine allmähliche Verlagerung des Aktivitätsfokus stattfindet. Hilfsaktivitäten sind also für virtuelle SHG erst dann verstärkt von Interesse, wenn die möglichen Gefahren abgeklungen sind.

Bei der Analyse fällt weiterhin auf, dass Retweets und die Verwendung von Links eine herausragende Rolle spielen. 39,08% der Nachrichten enthalten Links und 37,27% der Tweets sind oder enthalten Retweets, die ein wichtiges Instrument der Informationsverteilung sind. Als die Hilfsaktivitäten beginnen in den Fokus zu rücken, nimmt das Verlinken von externen Seiten (synthetische Informationen) deutlich zu, während der Anteil an Retweets (derivative Informationen) jedoch deutlich sinkt (Abb. 3). Eine mögliche Erklärung ist, dass das externe Angebot an Nachrichten, Bildern, Videos und Hilfsmöglichkeiten immer größer wird und damit auch die Notwendigkeit der Verteilung von Informationen innerhalb des Twitter-Raums schwindet. Bei näherer Durchsicht derjenigen Inhalte, auf die die Links verweisen, fällt auf, dass auf eine große Vielfalt von unterschiedlichen Webseiten verwiesen wird, die Selbsthilfeaktivitäten betreffen, jedoch keine zentrale Anlaufstelle existiert.

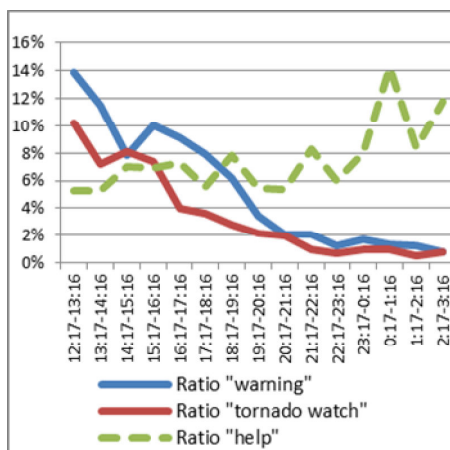


Abb. 2. „warning“, „tornado watch“ und „help“: Wenn Hilfeaktivitäten in den Fokus rücken, steigt der Anteil an verlinkten Seiten, während der Anteil an Retweets sinkt.

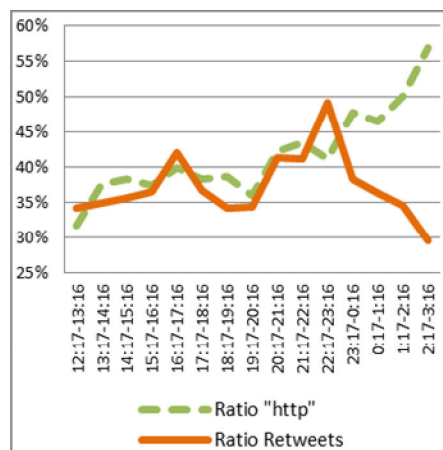


Abb. 3. Links und Retweets: Für virtuelle SHG beginnen die Aktivitäten besonders dann, wenn die eigentlichen Gefahren vorüber sind.

2.4 Ergebnisse (2): Qualitative Inhaltsanalyse

Beim Lesen der Tweets kann eine ganze Reihe von unterschiedlichen Aktivitäten festgestellt werden. Wie bereits weiter oben beschrieben, können Informations- und Hilfsaktivitäten sowie zwischen persönlicher Kommunikation mit persönlichen Informationen oder direkten Adressaten und zwischen unpersönlicher Kommunikation mit allgemeinen Informationen unterschieden werden.

Eine weit verbreitete, unpersönliche Informationsaktivität ist das Verlinken von Videos, Bildern und Nachrichten, in denen es häufig um Schadensberichte oder Tornadowarnungen geht. Beliebt sind auch kurze Tweets über die aktuelle Wettersituation, mal mit und mal ohne Link: „BRNCOiRADIO (28.04.2011, 12:28(EDT)): *The National Weather Service has issued a TORNADO WATCH IN CUMBERLAND COUNTY. That means conditions are favorable... <http://fb.me/A1ZPVcLc>*“. Weitere Informationen findet man über Stromausfälle oder Statistiken über Todesopfer. Informationsaktivitäten finden jedoch auch auf persönlicher Ebene statt. Hier werden individuelle Erlebnisse, Schadensmeldungen oder Informationen über den persönlichen Zustand weitergegeben. Häufig sind auch Familie und Bekannte Thema von persönlichen Tweets: „Rockt3 (28.04.2011, 12:20(EDT)): *@limitlessdreamz definitely grateful that the tornado storm did not take any of my fam or friends & prayed for the ones that were hit #ATL*“. Auch Tornadowarnungen finden nicht nur als allgemeine Informationstätigkeit statt, sondern richten sich auch häufig direkt an Bekannte, z. B. mit der Empfehlung, ihren derzeitigen Aufenthaltsort nicht zu verlassen.

Ein zweites großes Feld der Tweets betrifft Hilfsaktivitäten. Bei ihnen handelt es sich oft um Aufrufe zur Hilfe oder um Hilfsangebote. Es werden Hinweise geliefert, wie man helfen kann, sowie an welche Organisationen man sich wenden kann. Häufig enthalten diese Tweets Links auf externe Seiten. Vielfach zeigen sie Möglichkeiten zur Spende von Geld und Kleidung oder zur freiwilligen Teilnahme an Such- und Rettungs- oder Aufbaumaßnahmen auf. Diese Art der Aktivität findet meist, aber nicht immer, unpersönlich statt: „Farrah_Kaye (28.04.2011, 14:25 (EDT)): *RT @Amie_Theresa: @bscottracing Please RT! Help Alabama tornado victims! <https://chapterneighborhood.wufoo.com/forms/online-donations-midalabama-chapter>*“. Dieses Beispiel weist sowohl einen Aufruf auf, die Nachricht an andere zu verteilen, als auch die Bitte, Opfern zu helfen. Zusätzlich wird eine Seite verlinkt, auf der die Hilfe organisiert wird. Diese Nachricht wird im untersuchten Zeitraum insgesamt 18 Mal retweetet. Wie man am letzten Beispiel sehen kann, ist das Verlinken auf Seiten, auf denen Hilfe oder Ressourcen koordiniert werden, eine gelegentlich anzutreffende Tätigkeit. Eine zentrale Anlaufstelle ist bei der Analyse nicht aufgefallen, es wird eine große Bandbreite an Links angeboten. Twitter ist, wie bereits erwähnt, für eine komplexe Koordination von Hilfsaktivitäten wegen der begrenzten Nachrichtenlänge weniger geeignet. Es kommt aber vor, dass einzelne User, die in Hilfsaktivitäten beteiligt sind, ihren Status über Twitter bekanntgeben: „donnajean822 (28.04.2011, 13:17 (EDT)): *On my way to Alabama to deliver generators to the tornado-ravaged areas...*“. Hilfsaufrufe werden auch durch persönliche Kommunikation verbreitet oder es wird sich persönlich für den Einsatz bedankt: „jefferybiggs (28.04.2011 12:14 (EDT)): *@redcrossokc Experienced volunteers with tornado experience most welcome! Thanks for your help!*“. Eine weit verbreitete Hilfsaktivität bezieht sich auf die emotionale Unterstützung betroffener Menschen. Betroffene äußern ihre Angst, Helfer drücken ihr Mitgefühl für Opfer aus und rufen zu Gebeten auf. Dies lässt sich auch dadurch belegen, dass 8,11% aller Nachrichten den Begriff bzw. den Wortstamm „pray“ enthalten: „GrahamBlog (28.04.2011, 12:17 (EDT)): *Our prayers are with those who have been affected by the tornado's and storms across the South.*“. Es werden aber nicht nur Informationen und Hilfsangebote bereitgestellt, sondern auch kon-

kret angefordert: „jenconnic (28.04.2011, 12:20(EDT)): There's a tornado warning in portions of north Jersey. If you have news, photos, video, tweet us at @njdotcom.“.

Zusammengefasst lässt sich sagen, dass die gesammelten Twitterdaten eine breite Palette an Selbsthilfetätigkeiten während der Tornadokatastrophe zeigen können. Dazu gehören eine Vielzahl von Informations- und Hilfsaktivitäten, die durch persönliche und unpersönliche Kommunikation dargestellt werden. Welche Selbsthilfeaktivitäten noch stattfinden und welche Rollen Twitter-Nutzer einnehmen können, kann eventuell durch eine Analyse der besonders aktiven Twitterer beantwortet werden.

2.5 Ergebnisse (3): Rollenanalyse

Die Analyse nach Rollen liefert vier verschiedenartige Typen von Usern, die voneinander abgegrenzt werden können und sich in ihrem Twitterverhalten und im Inhalt Ihrer Tweets unterscheiden. Ihnen werden folgende Namen gegeben: Reporter, Retweeter, Wiederholer und Helfer (siehe Tabelle 1). Da die Kategorien nicht disjunkt sind und User zu mehr als einer Kategorie gehören können, ist die Summe ihrer prozentualen Anteile größer als 100%.

Helfer. Die Helfer sind in allerlei Hilfsaktivitäten involviert, und lassen sich besonders durch ihre Tweetinhalte abgrenzen. Meist stellen sie Hilfsangebote zur Verfügung oder rufen zur Hilfe auf. Vielfach verweisen sie auf Möglichkeiten zur Spende von Geld und Kleidung oder auf freiwillige Teilnahmen an Such- und Rettungs- oder Aufbaumaßnahmen. Unter ihnen befinden sich auch Katastrophenerfahrene, die generelle Tipps und Handlungsanweisungen für Betroffene geben. Typisch ist ebenso das Verhalten, Opfern emotional beizustehen. Auch Beteiligte realer SHG zählen in dieser Kategorisierung zu den Helfern, die Twitter beispielsweise dafür nutzen, ihren Status bekanntzugeben. Da die Palette an Hilfsaktivitäten sehr breit gefächert ist, erscheint es möglich, die Helfer in weitere Unterkategorien zu unterteilen.

Reporter. Der Reporter sorgt dafür, dass Informationen in den Twitter-Raum gelangen. Er nutzt entweder externe Quellen, oder ist selbst als Nachrichtenkanal die Originalquelle. Er produziert die generativen und synthetischen Informationen. Häufig entsprechen Reporter organisierten Nachrichtenkanälen, die ja als Institution per Definition nicht Teil einer SHG sind und unter anderem als Intermediäre zwischen BOS und SHG betrachtet werden können. Es kann festgehalten werden, dass Medien sehr intensiv als externe Informationsquelle für Selbsthilfeaktivitäten genutzt werden. Nicht nur Nachrichtenkanäle, sondern auch Augenzeugen stellen Originalquellen dar, die wiederum ein wichtiger Bestandteil von SHG sind.

Retweeter. Viele User beschränken sich in ihren Twitteraktivitäten auf das reine Retweeten von Informationen, die die Reporter in den Twitter-Raum hineingebracht haben. Dieser Userstyp wird hier Retweeter genannt und hat die Aufgabe, die von anderen Usern generierten und aus seiner Sicht wichtigen Informationen an seine Follower oder direkt an bestimmte User zu verteilen. Follower sind diejenigen Twitter-Nutzer, die den Nachrichten eines anderen Users folgen. Der Retweeter stellt die derivativen Informationen her und fällt durch eine hohe Anzahl von Tweets auf. Selten ist er auf ein bestimmtes Thema fixiert, häufiger retweetet er vielfältige Hilfsaufrufe, Warnungen, Kurzinfos oder Links mit Nachrichten. Selbst wird er nicht

retweetet, da er nie die Originalquelle ist. Der Retweeter entspricht in etwa dem bereits identifizierten ‘information broker’ [13].

Wiederholer. Aktive Twitter-Nutzer können auch die Rolle des Wiederholers einnehmen, der meist nur eine einzige Hauptbotschaft besitzt, die er verbreiten möchte. Um das zu erreichen, wiederholt er die gleiche Nachricht immer wieder, meist mit direkten Adressaten, um gehört zu werden. Daher fällt er durch eine besonders hohe Zahl an Tweets auf. Er sorgt demzufolge sowohl für die Generierung als auch für die Verbreitung einer Botschaft bzw. sehr weniger unterschiedlicher Botschaften. Beliebt ist unter ihnen die Methode, prominente Twitterer anzuschreiben, die eine sehr große Anzahl von Followern aufweisen, um so die Aufmerksamkeit zu steigern. Die Hauptbotschaft enthält oft einen Aufruf zur Spende, kann aber auch eine aus Sicht des Wiederholers wichtige Nachricht oder eine politische Meinung sein.

Tabelle 1. Rollen der aktiven Twitter-Nutzer

Rolle	Strukturelle Eigenschaft	Beschreibung	%
Der Helfer	Hohe Anzahl an Tweets und wird oft retweetet	Involviert in Hilfsaktivitäten	19%
Der Reporter	Wird häufig retweetet	Generiert Informationen	28%
Der Retweeter	Hohe Anzahl an Tweets	Verteilt Informationen	68%
Der Wiederholer	Hohe Anzahl an Tweets	Verbreitet eine Botschaft	16%

Es kommt natürlich vor, dass Twitter-Nutzer mehrere Rollen gleichzeitig einnehmen. Beliebt ist die Kombination des Reporters und Retweeters, also eine Person, die sich auf das Generieren und Verteilen von Informationen beschränkt. Eine andere gängige Kombination ist die des Helfers und Wiederholers. Die Analyse offenbart, dass ein Twitterer, der in Selbsthilfeaktivitäten involviert ist, immer mindestens einen dieser Rollen einnimmt. Unter den 46 Reportern sind eine überwiegende Mehrheit Nachrichten- und Wetterkanäle. Es bleibt jedoch festzuhalten, dass Beteiligte von Selbsthilfeaktivitäten selten sämtliche ihrer Möglichkeiten ausschöpfen, sondern nur einen Teilbereich abdecken, sei es der Informations- oder Hilfsbereich.

3 Interview-Studie: Selbsthilfeaktivitäten in der Wahrnehmung von BOS

3.1 Grundlagen: Reale Selbsthilfeaktivitäten

Das vorherige Kapitel betrachtet zeitliche Entwicklungen, Aktivitäten und verschiedene Rollen-Muster virtueller SHG. Dieser Absatz fokussiert auf reale SHG und deren Wahrnehmung durch offizielle Stellen. Einige Studien thematisieren die Sichtweise offizieller Hilfsorganisationen auf reale Selbsthilfe [1], [17-20]. Diese Ergebnisse sind zum Teil widersprüchlich und zeigen eine ambivalente Haltung gegenüber Selbsthilfe. Die Existenz von SHG wird einerseits als negativ empfunden [18], [20] und andererseits als ein entscheidender Faktor in der Katastrophenbewältigung bewer-

tet [19]. Weiterhin wird zum einen festgestellt, dass offizielle Planungen Selbsthilfe nicht berücksichtigen [1], [17] und zum anderen, dass Selbsthilfe ein wichtiger Bestandteil offizieller Hilfsaktionen ist [19]. Diese Widersprüche entstehen vermutlich durch örtliche und zeitliche Unterschiede der Untersuchungen. Ziel der folgenden empirischen Untersuchung ist es daher festzustellen, ob und warum deutsche BOS Selbsthilfe als wichtig empfinden, ob Schnittstellen für eine Zusammenarbeit geschaffen werden können und welche Grenzen für deren Wirken bestehen. In der anschließenden Diskussion wird die Schaffung von Schnittstellen nochmals aufgegriffen.

3.2 Methode

Die Untersuchung, die den Fokus auf Informations- und Kommunikationspraktiken im Krisenmanagement hatte [21-22], wurde in zwei Kreisen Nordrhein-Westfalens durchgeführt, bei deren Auswahl darauf geachtet wurde, dass sie geographisch und strukturell verschieden sind. Es wurde ein ländliches (Kreis A) und ein urbanes Gebiet (Kreis B) gewählt. Es wurde ein Grounded Theory orientierter Ansatz [16] durchgeführt, bei dem das Feld nicht nach vordefinierten Kategorien untersucht, sondern die Kategorien auf Basis des empirischen Materials gebildet wurden. Zur Rekonstruktion der Praxis wurden verschiedene qualitative Methoden wie Dokumentenanalysen, 4 Beobachtungen, 22 teilstrukturierte Interviews und 4 Gruppendiskussionen ausgewertet. Ziel der *Dokumentenanalyse* ist die Untersuchung offizieller Arbeitspraktiken. Es wurden Dokumente analysiert, welche die Arbeit im Krisenmanagement darstellen (Gesetze, Erlässe, Verordnungen, Richtlinien, Lehrgangsunterlagen). Die *Beobachtungen* dienten dazu, Erkenntnisse über die praktische Arbeit im Krisenmanagement zu erhalten. Diese wurden in der Kreisleitstelle während eines normalen Arbeitstages (9h), im Krisenstab und der Einsatzleitung einer Krisenübung (4h) und eines Großereignisses (6h) durchgeführt. Die *Interviews* (n = 22 je 1-2h; I01 bis I22) dienten der Erfassung der Arbeitspraxis, der Identifikation möglicher IT-Unterstützung und interorganisationaler Informationsbedarfe. Die *Gruppendiskussionen* (n = 4 je 4h) dienten der Erfassung kommunikativer Arbeitspraxen. Ziel der empirischen Untersuchungen ist es, Informations- und Kommunikationsprozesse zu analysieren, um IT-Unterstützungspotentiale zu ermitteln.

3.3 Ergebnisse

Fokussierung. Aus Sicht der Feuerwehr sollte Selbsthilfe besonders bei einfacheren Tätigkeiten stattfinden: „[...] gerade bei Starkregen wäre Selbsthilfe manchmal sehr hilfreich, wenn man weiß, dass man noch einige Sachen selbst im Keller besitzt, dann kann man 2 cm selbst wegputzen. [...] Anstatt hier anzurufen und zu schreiben, dass man 2 cm Wasser im Keller hat.“ (I05). Tätigkeiten, die selbst erledigt werden können, sollten nicht an offizielle Stellen gemeldet werden, um die Arbeit der Feuerwehr nicht durch die Meldung von Lappalien unnötig zu erschweren. „Es rufen ganz konkret Leute an: ‘Die Straße ist gesperrt, sagen Sie mir mal, wo ich herfahren muss’. Das überfordert uns natürlich auch. Wir können nicht das Denken für viele Menschen übernehmen“ (I06). Reale Selbsthilfe ist aus Sicht der Feuerwehr also gewünscht und

dient der Entlastung der offiziellen Hilfsorganisationen. Reale SHG werden auch deswegen als wertvoll angesehen, weil sie oftmals einen Informationsvorsprung gegenüber BOS vorweisen können. Dieser Vorsprung besteht aus einem detaillierteren Wissen über Schäden und die Umgebung: *„Es ist ja schon so, dass irgendeiner anruft, der das gesehen hat und wenn er dann mit dem Einsatzleiter oder Gruppenführer direkt vor Ort sprechen kann, dann fragen wir schon, ob er in der Nähe wohnt oder per Zufall vorbeigekommen ist oder ob er noch einen Gastank oder ähnliches hinterm Haus stehen hat bzw. was für eine Heizungsanlage derjenige hat.“* (I07). Mit digitalen Medien, wie Bildern, *„können Bürger nicht zur Beschreibung einer Situation beitragen“* (I03), dennoch nutzt die Polizei, die den Einsatz aus der Leitstelle führt *„Bilder, die von Reportern gemacht wurden um akkurate Informationen über die Situation“* (I04) zu erhalten.

Zusammenarbeit. Aus Sicht der Feuerwehr ist eine enge Zusammenarbeit zwischen BOS und SHG schwerlich umsetzbar, und zwar nicht aufgrund fehlenden Willens oder Geringschätzung, sondern vor allen Dingen aufgrund fehlender gesetzlicher Grundlagen: *„Alles, was wir tun, muss eine gesetzliche Grundlage haben. [...] Eine bürgerliche Selbsthilfegruppe ist keine Einheit der Gefahrenabwehr. Wir können die überhaupt nicht einsetzen, selbst wenn wir es wollten.“* (I06).

Veränderungen. Ein weitverbreitetes Phänomen im Zusammenhang mit Selbsthilfe wird von drei Feuerwehrkräften genannt (I07, I06, I09). Die gemeinschaftliche Selbsthilfe ist in kleinen Dörfern deutlich ausgeprägter als in großen Städten: *„Je mehr Sie ins Ländliche fahren, desto mehr unterstützen sich die Bürger, desto weniger rufen die auch nach dem Staat. Je dichter die Ballungszentren werden, je anonymmer, desto mehr schreien die nach dem Staat.“* (I09). Ein Interview beschreibt die Einstellung: *“Es muss ja irgendjemanden geben, der dafür zuständig ist, meinen Keller leer zu räumen und das idealerweise kostenlos, weil ich Steuerzahler bin“* (I06). Manchmal helfen Bürger aus rechtlichen Bedenken nicht: *“Wer haftet denn dafür, dass wenn ich den Baum zur Seite räume, und beschädige dabei ein anderes Fahrzeug?“* (I06). Anhand der Zitate lässt sich bereits erahnen, dass eine zu hohe Anspruchshaltung von Seiten der Bürger als negativ angesehen wird. Ebenso wird eine zeitliche Veränderung ausgemacht. Es wird ein Mentalitätswechsel der Bürger beklagt, die Verantwortung von sich auf staatliche Organisationen zu schieben: *„Da merkt man schon sehr deutlich, dass das Anspruchsdenken extrem zugenommen hat, in den letzten 15 bis 20 Jahren. Da kommt der Anruf, dass der Kanaldeckel zehn cm danebenliegt. Normal tritt man dagegen und dann ist der wieder drin. Nein, heute ruft man nach dem Staat.“* (I09). Es wird daher eine stärkere Sensibilität für Selbsthilfe empfohlen: *„In unserem Interesse läge es eher die Bevölkerung zu sensibilisieren, dass Selbsthilfemechanismen wieder greifen.“* (I05).

Zusammenfassung. Insgesamt lässt sich festhalten, dass die BOS die Tätigkeit von SHG durchaus schätzen und anerkennen. Sie beklagen allerdings eine besonders in Städten nachlassende Bereitschaft innerhalb der Bevölkerung, Selbsthilfe zu leisten und wünschen sich daher die Förderung von Selbsthilfemechanismen. Aufgrund der fehlenden gesetzlichen Grundlage ist zu vermuten, dass SHG in den formalen Plänen bei der Bekämpfung von Katastrophen zwar eine eher geringe Rolle spielen, auf In-

formationsebene jedoch ein Vorsprung seitens der Bürger erkannt wird und daher ein Informationsaustausch für beide Seiten hilfreich sein kann.

4 Kollaborationsunterstützung realer und virtueller SHG

Das Ziel dieses Beitrags ist es Pfade zur besseren Zusammenarbeit virtueller und realer SHG zu beschreiben. Die Twitter-Studie hat Entwicklungen im Aktivitätsfokus deutlich gemacht und verschiedene Nutzer-Typen in virtuellen SHG ergründet: die Helfer, die Reporter, die Retweeter und die Wiederholer. Insbesondere die Gruppe der Helfer ist aktiv in reale und virtuelle Selbsthilfeaktivitäten involviert und Hilfsaktivitäten finden besonders dann statt, wenn die primäre Gefahr vorüber ist. Die Interview-Studie hat eine Reihe von Feststellungen bezüglich der Wahrnehmung von realen SHG durch BOS verdeutlicht. Sowohl virtuelle als auch reale SHG entstehen, um die negativen Auswirkungen einer Katastrophe zu bekämpfen. Die Interviews deuten darauf hin, dass deren Existenz auch von BOS gewünscht und anerkannt wird.

Software wird von virtuellen SHG besonders dafür verwendet, Warnungen auszusprechen, vielfältige Informationen zu generieren und zu verteilen, Hilfsangebote zu verbreiten und allgemein zu kommunizieren. Reale SHG nutzen Social Media besonders für die Koordination ihrer Aktivitäten, indem sie beispielsweise Statusmeldungen über Twitter bekanntgeben. In der gegenwärtigen Arbeitsweise von virtuellen und realen SHG können anhand der Literatur- und empirischen Studien Defizite festgestellt werden, die Unterstützungspotenziale offenbaren: eine fehlende zentrale Anlaufstelle sowohl für virtuelle als auch für reale Gemeinschaften; eine fehlende auf reale Gemeinschaften zugeschnittene Koordinationsplattform; und die Wahrnehmung von rückgängiger Eigeninitiative im Bereich der realen Selbsthilfe.

Zur Adressierung der aufgeführten Herausforderungen ergeben sich die folgenden Gestaltungsaspekte und möglichen IT-Unterstützungsmöglichkeiten:

A: Integration in bestehende Netzwerke. Bereits 1984 wurde deutlich, dass die Existenz eines sozialen Netzwerks vor der eigentlichen Krise eine Bedingung für die Gründung von SHG darstellt [3]. Auf Social Media übertragen heißt dies, dass die in Diensten wie Twitter oder Facebook gebildeten Vernetzungen auch für SHG günstige Voraussetzungen darstellen. Dies wäre sowohl über die Nutzung eben dieser Dienste, jedoch auch über Schnittstellen oder die Realisierung spezifischer Anwendungen als eingebettete Facebook-App möglich. Wegen möglicher Akzeptanzprobleme sollten in jedem Fall keine neuen Interaktionsformen etabliert werden, sondern die auch von realen Helfern etablierte Nutzung unterstützt werden.

B: Entstehungsförderung. In unserer Interview-Studie und mancher Literatur [19] wird herausgestellt, dass bei SHG die positiven Effekte überwiegen; es wird jedoch auch deutlich, dass das Anspruchsdenken der Bürger in der Wahrnehmung der BOS in den letzten Jahren deutlich zugenommen hat und die Bereitschaft zur Selbsthilfe, zumindest bei kleinen Lagen, sinkt. Die Bedingungen, die die Gründung von SHG begünstigen [3], bieten Ansätze dafür, wie die Existenz gefördert und neue Helfer gefunden werden können. Die Twitter-Studie deutet an, dass das Funktionieren einer SHG davon abhängen könnte, wie das Ausmaß der Gefahr wahrgenommen wird

(Abbildung 2). Daher könnte die visuelle Darstellung steigender oder fallender Gefahr einen unterstützenden Effekt auf die Gründung von und Teilnahme an solchen Gruppierungen haben. Weiterhin könnten steigende Interaktionen virtueller SHG reale Hilfsaktivitäten fördern und eher passive Nutzer zur Teilnahme auffordern.

C: Verknüpfung virtueller und realer SHG. Virtuelle Gemeinschaften legen besonderen Wert auf den Bereich der Information, bei dem durch Zusammenarbeit kollektive Intelligenz entsteht. Reale SHG benötigen IT vor allem als Koordinationsplattform. Beide Bereiche können mittels unterschiedlicher Arten von Social Software unterstützt werden. Während besonders Wikis [23] auf den Aspekt der Information abzielen, können beispielsweise Gruppenmodule den Koordinationsproblemen Abhilfe schaffen. Um zwischen beiden Bereichen synergetische Effekte zu erzielen, müssen diese so eng miteinander verknüpft sein, dass zwangsläufig die Interaktion zwischen virtuellen und realen SHG entsteht. Virtuelle SHG könnten durch die Bereitstellung einer Vielzahl von relevanten Informationen sowohl betroffene Bürger als auch reale SHG direkt unterstützen. Akteure, die sich zunächst vorwiegend auf die virtuelle Selbsthilfe konzentrieren, könnten zur Beteiligung an einer realen SHG animiert werden. Daher sollten virtuelle und reale SHG zur Nutzung gleicher Systeme bewegt werden. Nutzer unterschiedlicher Rollen (basierend auf der Klassifikation der Twitter-Studie) könnten auf Basis des Nutzungsverhaltens optisch unterschiedlich dargestellt werden (spezielle Symbole). Somit wäre es auch möglich, Bedenken von BOS über die Reliabilität von Informationen anonymer Nutzern entgegenzuwirken.

D: Schnittstellen zum offiziellen Katastrophenmanagement. In der Twitter-Studie wurde herausgestellt, wie Nutzer mit einer Vielzahl unterschiedlicher Informationsarten umgehen, wie Bilder, Videos, Dokumente oder externen, medialen Informationen. Mechanismen zur Identifikation von wertvollen Informationen, z. B. basierend auf einer Bewertung durch virtuelle SHG (d. h. eher passive Nutzer, die nicht vor Ort helfen), könnten gerade für BOS hilfreich sein. Gemäß der Interviews ist momentan eine enge Zusammenarbeit zwischen BOS und realen SHG auf Koordinationsebene schwer umsetzbar, jedoch bietet sich auch dort ein Austausch auf Informationsebene aufgrund von Informationsvorsprüngen verschiedener Gruppen, wie beispielsweise der „Reporter“, an. Diese stellen teilweise Nachrichtenkanäle dar und können als Intermediäre handeln. Ein Monitoring der Aktivität spezifischer Rollen (z. B. vor Ort, aktiv), wie in der Twitter-Studie herausgestellt, könnte die Selektion vertrauenserweckender, und somit verwendbarer Informationen fördern, die auch zur Lagebeurteilung einbezogen werden könnten [24].

5 Fazit

Dieser Beitrag hat das Phänomen der in Katastrophen tätigen Selbsthilfegemeinschaften (SHG) mittels einer Analyse der Twitternutzung während einer Katastrophenlage und einer Analyse von Interviews mit deutschen BOS empirisch untersucht. Hierbei können virtuelle und reale SHG unterschieden werden. Erkannte Defizite in deren bisheriger Arbeitsweise sowohl innerhalb solcher Gemeinschaften als auch in Zusammenarbeit mit öffentlichen Gruppen haben Software-basierte Unterstützungspos-

tenziale offenbart und zur Zielformulierung einer Social Software zur Unterstützung von SHG geführt:

Indem Social Software sowohl virtuellen als auch realen Gemeinschaften als zentrale Anlaufstelle dient, könnten durch die Verknüpfung von virtuellen und realen SHG synergetische Effekte erzielt und durch die Schaffung von Schnittstellen Selbsthilfe im offiziellen Katastrophenmanagement etabliert werden. Für die Realisierung von IT zur Unterstützung von Selbsthilfe in Katastrophenlagen werden die (A) Nutzung existierender Netzwerke, die (B) Unterstützung der Entstehung dieser Gruppen durch das Schaffen von Awareness, (C) die zwingende Verknüpfung realer und virtueller Aktivitäten und damit verbunden die Animierung der realen Teilhabe bisher nur virtueller Teilnehmer sowie (D) die Integration in Systeme des offiziellen Krisenmanagements vorgeschlagen.

Limitationen. Die hier verwendeten empirischen Methoden unterliegen Grenzen: Die gesammelten Twitterdaten enthalten aufgrund der Suchmethode höchstwahrscheinlich nicht alle Tweets, die im gesuchten Zeitraum in Zusammenhang mit der Tornadokatastrophe getwittert wurden. Auch die dort verwendeten Analysemethoden haben nicht den genauen Inhalt sämtlicher Tweets erfassen können. Dennoch haben sie über die Literatur hinaus [8-10], [14-15] einige Schlussfolgerungen über virtuelle SHG ermöglicht. Die qualitativ ausgewerteten Interviews zielten zwar nicht nur auf das Thema der Selbsthilfe ab, konnten jedoch den bisherigen Kenntnisstand weiter differenzieren [1], [17-20]. Hierbei sind reale SHG nur aus der Perspektive von BOS betrachtet worden. Probleme, die bei der Zusammenarbeit von realen SHG auftauchen könnten, wurden in der Empirie nicht erfasst. Daher ist eine stärkere Einbeziehung von potenziellen Nutzern einer Social Software für SHG für eine weitere Erforschung ihrer Unterstützungspotenziale unerlässlich.

Danksagung: Diese Forschungsarbeiten wurden durch ein Stipendium der Research School Business & Economics ermöglicht und im Rahmen des Verbundprojektes 'infostrom' durch das Bundesministerium für Bildung und Forschung (Fö.-Kz. 13N10712) unterstützt.

Literatur

1. Stallings, R.A., Quarantelli, E.L.: Emergent Citizen Groups and Emergency Management. *Public Administration Review* 45, 93–100 (1985)
2. Reuter, C., Marx, A., Pipek, V.: Disaster 2.0: Einbeziehung von Bürgern in das Krisenmanagement. In: Eibl, M. (ed.): *Mensch & Computer 2011*, pp. 141–150. Oldenbourg, München (2011)
3. Quarantelli, E.L.: *Emergent Citizen Groups in Disaster Preparedness and Recovery Activities* (1984)
4. Helsloot, I., Ruitenber, A.: Citizen Response to Disasters: A Survey of Literature and Some Practical Implications. *Journal of Contingencies and Crisis Management* 12, 98–111 (2004)
5. Lowe, S., Fothergill, A.: A Need to Help: Emergent Volunteer Behavior after September 11th. *Beyond September 11th: An Account of Post-Disaster Research*, SP39, 293–314.

- Natural Hazards Research and Applications Information Center, University of Colorado, Boulder (2003)
6. Voorhees, W.R.: New Yorkers Respond to the World Trade Center Attack: An Anatomy of an Emergent Volunteer Organization. *Journal of Contingencies and Crisis Management* 16, 3–13 (2008)
 7. Palen, L., Liu, S.B.: Citizen communications in crisis: anticipating a future of ICT-supported public participation. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM Press (2007)
 8. Starbird, K., Palen, L.: Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pp. 1071–1080. ACM Press (2011)
 9. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pp. 1079–1088 (2010)
 10. Qu, Y., Huang, C., Zhang, P., Zhang, J.: Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake. In: *Proceedings CSCW*, pp. 25–34 (2011)
 11. White, C., Plotnick, L., Kushma, J., Hiltz, S.R., Turoff, M.: An online social network for emergency management. *International Journal of Emergency Management* 6, 369–382 (2009)
 12. Palen, L., Vieweg, S.: The emergence of online widescale interaction in unexpected events: assistance, alliance & retreat. In: *Proceedings CSCW*, pp. 117–126. ACM Press (2008)
 13. Hughes, A.L., Palen, L.: Twitter Adoption and Use in Mass Convergence and Emergency Events. In: *Proceedings ISCRAM*, Gothenburg (2009)
 14. Starbird, K., Palen, L., Hughes, A.L., Vieweg, S.: Chatter on The Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In: *Proceedings CSCW*, pp. 241–250. ACM Press (2010)
 15. Starbird, K., Palen, L.: Pass It On?: Retweeting in Mass Emergency. In: *Proceedings of the International ISCRAM Conference*, pp. 1–10., Seattle, USA (2010)
 16. Strauss, A.: *Qualitative Analysis for Social Scientists*. Cambridge press (1987)
 17. Dynes, R.R.: *Social Capital: Dealing with Community Emergencies*. *Homeland Security Affairs* 2 (2006)
 18. Lanzara, G.F.: Ephemeral Organisations in Extreme Environments: emergence, strategy, extinction. *Journal of Management Studies* 20, 71–95 (1983)
 19. Lorenzen, D.: Risikokommunikation bei Naturkatastrophen - Ausgewählte Ergebnisse der Befragung im Herbst 2004 (2005)
 20. Pfeil, J.: Maßnahmen des Katastrophenschutzes und Reaktionen der Bürger in Hochwassergebieten. Deutsches Komitee für Katastrophenvorsorge e.V. (DKKV) (2000)
 21. Ley, B., Pipek, V., Reuter, C., Wiedenhofer, T.: Supporting Improvisation Work in Inter-Organizational Crisis Management. In: *Proceedings CHI*, pp. 1529–1538. ACM, NY (2012)
 22. Reuter, C., Pohl, P., Pipek, V.: Umgang mit Terminologien in inter-organisationaler Krisenkooperation - eine explorative Empirie. In: Eibl, M. (ed.): *Mensch & Computer 2011*, pp. 171–180. Oldenbourg, München (2011)
 23. White, C., Plotnick, L., Addams-Moring, R., Turoff, M., Hiltz, S.R.: Leveraging a Wiki to Enhance Virtual Collaboration in the Emergency Domain. In: *Proceedings HICSS* (2008)
 24. Reuter, C., Ritzkatis, M.: Unterstützung mobiler Geo-Kollaboration zur Lageeinschätzung von Feuerwehr und Polizei. In: Alt, R., Franczyk, B. (eds.): *Proceedings Wirtschaftsinformatik*, Leipzig (2013)

Unterstützung mobiler Geo-Kollaboration zur Lagebeurteilung von Feuerwehr und Polizei

Christian Reuter und Michael Ritzkatis

Universität Siegen, Institut für Wirtschaftsinformatik, Siegen, Germany
christian.reuter@uni-siegen.de
michael.ritzkatis@student.uni-siegen.de

Abstract. Aufgrund komplexer und dringlicher Aufgaben steht die Zusammenarbeit über örtliche und organisationale Grenzen hinweg bei Behörden und Organisationen mit Sicherheitsaufgaben (BOS), wie Feuerwehr und Polizei, an der Tagesordnung. Ziel dieses Beitrags ist es zu untersuchen, wie die Kollaboration von Einsatzkräften vor Ort und jenen in der Leitstelle durch mobile Geokollaborationssysteme unterstützt werden kann. Nach einer Darstellung verwandter Arbeiten werden anhand einer qualitativen empirischen Studie die Informations- und Kommunikationspraktiken mobiler Einsatzkräfte vorgestellt. Hierauf aufbauend folgt die Konzeptionierung und Umsetzung eines mobilen Geokollaborationssystems, welches an ein bestehendes Krisenmanagementsystem und Geoinformationssystem (GIS) angebunden ist und als Android-App realisiert wurde. Abschließend werden die Evaluationsergebnisse dieses Systems im Anwendungsfeld vorgestellt.

Keywords: Mobile Applications, Geoinformationssysteme, Feuerwehr, Polizei, Kollaboration, CSCW

1 Einleitung

Behörden und Organisationen mit Sicherheitsaufgaben (BOS) sind häufig mit Situationen konfrontiert, in denen schnell Entscheidungen getroffen werden müssen. Ob Orkan, Amoklauf oder Verkehrsunfall: Der Einsatzleiter muss zügig seine Lagebeurteilung durchführen um die vorhandenen Ressourcen sinnvoll einzusetzen. Die dafür benötigten Informationen stehen häufig nicht zentral zur Verfügung [1], sondern müssen von der Leitstelle, speziellen Einsatzkräften oder Dritten angefordert werden [2]. Zur Lagebeurteilung spielen vor allem Geoinformationen, wie Standorte von Gefahren, Ressourcen oder Einsatzabschnitten eine große Rolle. Diese werden auf Lagekarten in der Leitstelle oder in der Abschnittsführung verwaltet. Mit Hilfe mobiler Geräte wäre es auch vor Ort denkbar die Standorte der einzelnen Einheiten zu betrachten und einen schnelleren Überblick über die Lage zu erhalten, als dies bei Standortmeldungen und Lageauskünften via Funk der Fall ist [3]. Trotz der stetigen Entwicklung der Informations- und Kommunikationstechnologien konnten nur verhältnismäßig wenige IT-basierte Anwendungen und mobile Systeme Einzug in den Krisenmanagementzyk-

lus erhalten [4]. Dieser Beitrag untersucht inwiefern mobile Geokollaborationssysteme die Zusammenarbeit zwischen Einsatzkräften vor Ort und jenen in der Leitstelle unterstützen kann. Dazu wird aufbauend auf verwandten Arbeiten (Kapitel 2) und den Ergebnissen einer qualitativen empirischen Studie (3) die Konzeptionierung, Umsetzung (4) und Evaluation (5) eines mobilen Geokollaborationssystems dargestellt.

2 Verwandte Arbeiten

Ansätze zur mobilen Nutzung von Geoinformationssystemen (GIS) – Systeme zur Erfassung, Verarbeitung und Nutzung ortsbezogener Informationen – lassen sich aus der Perspektive der Datenerfassung, Datenverarbeitung und Kollaboration betrachten.

Datenerfassung. Einige Ansätze, wie Kunze et al. [4] mit ihrem Sensornetzwerk zur Erfassung von Personalien bei Massenanfällen von Verletzten oder Nestler & Klinker [5] in der Nutzung von RFID-Chips, betonen die effektivere Erschließung der Schadenslage durch die Verwendung mobiler Sensordaten, bei denen jedoch auf gängigen Smartphones nicht vorhandene Sensortechnik notwendig ist, und somit derzeit nicht für zahlreiche Einsatzkräfte verwendbar ist. Laufs et al. [6] begrenzen sich auf gängige Sensoren wie Kamera, Mikrofon und GPS und stellen die Verwendung einer Zeitleiste dar, mit welcher der Zustand einer Information zu jedem Zeitpunkt wiederhergestellt werden kann, was als Dokumentationsgrundlage dienen kann. Fajardo & Oppus [7] betonen die Notwendigkeit der Ortung von Informationen. Die Erhebung von Informationen mithilfe Sensoren wird durchgängig als besonders hilfreich beschrieben. Die erhobenen Informationen müssen jedoch geeignet verarbeitet werden.

Datenverarbeitung. Liu & Palen [8] betonen die räumliche und zeitliche Repräsentation der Lage um diese in Gänze beurteilen zu können. Birkhäuser et al. [9] fokussieren die Informationsdarstellung auf digitalen Lagekarten um eine effizientere Lagebeurteilung zu ermöglichen. In der Darstellung sensibler Informationen ergeben sich jedoch auch Aspekte der Zugriffsrechte [10]. Convertino et al. [11] sehen dafür die Bereitstellung von zwei Karten vor, um so Informationen wahlweise für alle oder nur sich selbst anzuzeigen. Wu et al. [12] stellen in ihrem System CIVIL ebenfalls eine öffentliche und eine private Karte bereit. Andere Ansätze [14] beschreiben die Möglichkeit, den Informationsaustausch in einem mobilen Client und einem Webinterface abzubilden. Brinkhoff et al. [15] beschreiben die Erstellung eines Open-Source GIS für den mobilen Kontext, der die Möglichkeit einer Datenzwischenspeicherung bei Verbindungsausfall mit anschließender Synchronisation beinhaltet. Diese ist besonders in Katastrophenlagen wertvoll ist, da eine dauerhafte Verbindung nicht immer gewährleistet werden kann. Der Ansatz von Monares et al. [16] unterstreicht die Notwendigkeit der Reduzierung des Datenverkehrs im mobilen Kontext, gerade in schwachen Empfangsgebieten. In der Datenverarbeitung sind zusammengefasst neben räumlicher und zeitlicher Darstellung auch unterschiedliche Zugriffsrechte, mögliche Verbindungsausfälle und reduzierte Datenmengen zu berücksichtigen.

Kollaboration. Neben Erfassung und Darstellung von Daten ist vor allem eine Unterstützung der Zusammenarbeit unter den Einsatzkräften sinnvoll. Eine möglichst realitätsnahe Lagedarstellung möchten Marches & McNeill [17] erzielen, um durch

diese Nähe eine fundierte Beurteilung zu ermöglichen. Neuvel & Zlatanova [18] erweitern die Verwendung von GIS um die Erfassung und Aushandlung von Ressourcen. Stevens & Wulf [13] unterscheiden in der Gewährung von Zugriffsrechten zwischen den Konzepten Awareness, Protection und Negotiation. Schafer et al. [19] fokussieren die Individualisierung, die zur Vermeidung von Informationsüberflutungen jedem Benutzer seine eigenen Kategorien anzeigen lässt. Dem Problem unterschiedlicher Datenformate nehmen sich Schiller et al. [20] mit Schaffung eines einheitlichen Formats an. Cai et al. [21] stellen Client- und Server- System vor, welches besonders durch einen externen Server mit gesicherter Verbindung eine Weiterarbeit aller ermöglicht, wenn nur ein Gerät keine Verbindung zum Internet besitzt. Singh & Ableiter [22] betonen, dass eine Person sowohl Sender als auch Empfänger von Informationen sein sollte. Um in diesem Kontext Chaos zu vermeiden ermöglichen Catarci et al. [23] das Zuweisen von klaren Aufträgen an Einsatzkräfte. In der übergreifenden Lagebeurteilung müssen auch terminologische Herausforderungen beachtet werden [24]. Zusammengefasst spielen neben technischen Aspekten, wie Datenformaten und Ausfallsicherheit, auch organisationale Aspekte der Aushandlung eine wichtige Rolle.

Abgrenzung. Die aufgeführten Ansätze suggerieren, dass sinnvolle Verwendungsmöglichkeiten mobiler Geräte und GIS existiert. Im Bereich der Datenerfassung zeigen die Arbeiten, dass die Verwendung komplexer Sensortechnik zur Zeit noch fraglich ist [6-7], weshalb in dieser Arbeit eine Beschränkung auf die Sensoren typischer mobiler Geräte (Smartphones, Tablets) stattfinden wird. Einige verwandte Arbeiten zeigen bereits gute Ansätze zur Kollaboration [20-22], jedoch mit Schwerpunkt auf Datenformaten oder der Verteilung der Informationen und weniger dem Erfassen der Kollaborationspraktiken als solches. Im Folgenden wird daher der Schwerpunkt auf die empirische Ermittlung von Kollaborationspraktiken und deren technische Unterstützungsmöglichkeiten unter Verwendung eines GIS, mobiler Sensoren und unter Betrachtung möglicher Ausfälle des Datenverkehrs [15-16] gelegt.

3 Empirische Studie: Potentiale mobiler Geokollaborationssysteme für BOS

Verwandte Arbeiten konnten bereits das Potential von GIS und der Erstellung von Informationen unter Verwendung mobiler Sensoren aufzeigen. Ziel der empirischen Studie ist nun die Praxis der Akteure zu untersuchen, um weitere Erkenntnisse zu mobilen Einsatzpraktiken und technischen Potentialen zu generieren.

3.1 Methodik

Innerhalb der Empirie wurden die Informations- und Kommunikationspraktiken insb. von Polizei und Feuerwehr mit Fokus auf die mobile Lagebeurteilung analysiert. Die Studie fokussierte die aktuelle Praxis der einzelnen Organisationen bei der intra- und interorganisationalen Bewältigung eines großflächigen Stromausfalls [1]. Die Daten wurden im Jahr 2011 in zwei Kreisen Nordrhein-Westfalens erhoben, bei deren Aus-

wahl darauf geachtet wurde, dass sie geographisch und strukturell verschieden sind. Es wurde ein ländliches (Kreis A) und ein urbanes Gebiet (Kreis B) gewählt. Um die Praxis geeignet erfassen zu können, wurden qualitative Methoden wie vier Beobachtungen, fünf Gruppendiskussionen und 22 Interviews mit hauptamtlichen Kräften verschiedener Ebenen der Feuerwehr, Polizei, Kreisverwaltung und Energienetzbetreiber (~90 Minuten, Tabelle 1) angewendet. Einige hauptamtliche Mitarbeiter sind zusätzlich ehrenamtlich in Freiwilligen Feuerwehren ihres Wohnortes tätig.

Tabelle 1. Ausgewählte Interviews (Phase 1): Informations- und Kommunikationspraktiken

Nr.	Kreis	Organisation	Rolle
I02	A	Polizei	Leiter des Führungs- und Lagedienstes
I03	A	Polizei	Abschnittsführung
I04	A	Polizei	Polizeibeamter im Wach- und Wechseldienst
I06	A	Feuerwehr	Stv. Leitstellenleiter
I07	A	Feuerwehr	Mannschaftsmitglied
I09	B	Feuerwehr	Amtsleiter Bereich Feuerschutz und Rettungsdienst
I11	B	Feuerwehr	Einsatzleiter
I13	B	Feuerwehr	Abteilungsleiter, Brandinspektor

Bei der Auswertung des vorliegenden Datenmaterials wurde sich am Grounded Theory-Verfahren nach Strauss [25] orientiert. Dazu wurden die Transkripte offen kodiert und die Aussagen in Textbausteine und spätere Kategorien aufgeteilt. Das Vorwissen aus der Literaturstudie diente dabei als Erhöhung der „theoretischen Sensitivität“ [25]. Ein Bestandteil der Grounded Theory ist das ‚Theoretische Sampling‘, was bedeutet, dass die Auswahl der Untersuchungseinheiten durch die im Analyseprozess entstehende Theorie geleitet wird. Ein Ergebnis der ersten Phase war unter anderem, dass besonders bei Polizei und Feuerwehr eine Zusammenarbeit auf Basis von Lagekarten notwendig ist, um Entscheidungen zu treffen. Diese besitzen jedoch verschiedene Führungsstrukturen: Bei der Polizei werden die Einsätze direkt aus der Leitstelle heraus („von hinten“), bei der Feuerwehr hingegen durch den Einsatzleiter vor Ort („von vorne“) geführt, wobei die Leitstelle dem Einsatzleiter unterstützend dient. Zur genaueren Untersuchung mobiler Kollaborationspraktiken wurden fünf ergänzende teilstrukturierte Interviews im Jahr 2012 durchgeführt (~60 Minuten, Tabelle 2), mit denen die aktuelle Praxis der sowohl auf strategisch-taktischer-, als auch auf operativer Ebene agierender Akteure tiefgehend mit speziellem Fokus auf die Kollaboration mobiler Einsatzkräfte und Leitstellenmitarbeiter analysiert werden konnte.

Tabelle 2. Geführte Interviews (Phase 2): Mobile Kollaborationspraktiken

Nr.	Kreis	Organisation	Rolle
IM1	A	Polizei	Leiter des Führungs- und Lagedienstes
IM2	A	Feuerwehr	Leitstelle Datenpflege / Digital-Funk / Lagedienst
IM3	A	Feuerwehr	Administrator Kreisleitstelle
IM4	A	Polizei	Wachleiter
IM5	B	Feuerwehr	Abteilungsleiter Leitstelle

3.2 Ergebnisse

Durch die Auswertung der Interviews konnten die Aussagen zu Kategorien zusammengefasst werden, die im Folgenden dargestellt werden.

Dringlichkeit und Komplexität. Smartphones und Tablets sind bei den Interviewten häufig aus dem privaten Umfeld bekannt (IM4): *„Einige Kollegen und Kolleginnen haben über Smartphones auch einen Internetzugang, der häufig sinnvoll ist um sich zum Beispiel bei Google Maps mal ein Luftbild von einer Örtlichkeit anzuschauen“* (I04). Diese werden insbesondere aus zeitlichen Gründen genutzt: *„Wenn wir dafür extra ein zusätzliches Notebook oder ähnliches Gerät hochfahren müssen um damit Informationen zu holen, wird das [...] nicht benutzt werden“* (I04). Neben einer kurzen Startzeit war gefordert, dass *„die Informationen schon schnellstmöglich vorliegen“* (I11) sollten. Die Dringlichkeit wirkt sich auch auf die mögliche Komplexität der Anwendung aus: *„Dann sag ich dem klick mal hier drauf und dann haste die Kräfteübersicht. Je einfacher, je besser“* (IM4). Dabei muss jedoch eine große Bandbreite an Nutzern berücksichtigt werden: *„[...] Bei Freiwilligen Feuerwehren ist es oft so, dass da nicht unbedingt die Computermenschen vorsitzen, sondern die tippen auch schon einmal Schreibfehler ein. Es muss für die einfach sein“* (IM2).

Zielgruppe. Mobile Anwendungen suggerieren eine Einsatzmöglichkeit vor Ort. Dort sind jedoch mehrere Zielgruppen denkbar: *„So als Unterstützung für Einsatzkräfte ok. Da gibt es ja keine Grenzen, das kann ja von mir aus auch jeder Streifenwagen haben [...]. Dann halt ohne taktische Relevanz“* (IM4). Demnach sollten verschiedene Sichten möglich sein; *„Also ich als Benutzer melde mich an und [...] dann habe ich die und die Information des Systems schon vorinstalliert. [...] So etwas wäre super“* (I03). Sind für die Zielgruppe der unteren Hierarchieebenen wie z. B. Streifenwagen noch Smartphones ausreichend, gilt für Abschnittsleitung und Einsatzleitung jedoch: *„Wenn Sie eine wirkliche Erleichterung für die Einsatzkräfte und Abschnittsleiter haben wollen, sind die natürlich zu klein“* (IM4). Dort bietet sich eher der Einsatz von Tablets an.

Informationsbeschaffung. Als ein Hauptwunsch des Einsatzes mobiler Systeme wurde erläutert, dass *„[...] auch wenn ich nicht örtlich in der Leitstelle bin, [...] ich mir ggf., Informationen zukommen lassen kann [...]. Ich möchte das jetzt selber mal visualisiert haben, selber im O-Ton sehen“* (I13). Derzeit arbeiten Einsatzkräfte bevorzugt mit visualisierten Informationen auf Lagekarten (IM3), jedoch gibt es auch bei Lagekarten verschiedene Typen, z. B. Satellitendarstellung: *„Also was uns immer hilft ist ein Blick von oben drauf, also diese Luftbildgeschichte“* (IM1). Auf der anderen Seite gibt es Karten, die beispielsweise bei Vermissteneinsätzen benutzt werden: *„Das wäre eine super Unterstützung, wenn wir auch mal die topographischen Karten hätten“* (IM4). Neben der Karte an sich wäre ein Ziel die Darstellung unterschiedlicher Informationskategorien, beispielsweise ein Wetteroverlay: *„Du hast da eine Rauchwolke und kannst dann mal gucken. [...] Da und da geht die Wolke hin in der und der Schichthöhe, der und der Bereich ist gefährdet“* (IM3). Diese Informationen können zwar aktuell auch über eigene Smartphones abgerufen werden, jedoch nicht in einer integrierten Karte: *„[...] wenn es all diese Informationen in einem geben würde, wäre das schon toll“* (I07).

Awareness. Neben der individuellen Informationsbeschaffung bietet auch die Unterstützung von Kooperation und Awareness Potential. *„Auch wenn wir mit einem Funksystem arbeiten, dass aus dem vorletzten Jahrhundert stammt, ist es doch so, dass wir mit dem Funk alle versorgen“* (IM4). Auch bei einer Vermisstensuche ist es für die Polizei hilfreich zu wissen, in welchem Bereich die Feuerwehr agiert und umgekehrt: *„Wenn ich dann sehen würde: hier ist das Waldgebiet, hier ist die Feuerwehr, da bin ich. Dafür alleine würde es sich schon fast lohnen“* (IM4). Auch sinnvoll erscheint dies für ein gemeinsames Verständnis: *„Das wär jetzt ultimativ: Wenn [die Information] in Verbindung mit WLAN oder UMTS übertragen wird und die Leitstelle kann direkt mitgucken, dann sind auch die Schnittmengen zwischen der Einsatzleitung vor Ort und der Leitstelle größer. Sonst redet der über das Feuer und der versteht ein anderes Feuer“* (IM2). Neben dieser Darstellung in der Leitstelle sollten Einsatzkräfte vor Ort auch die Möglichkeit erhalten, für sie relevante Informationen der Leitstelle zu erhalten: *„Dann würde ich ja die Straßensperren sehen und würde auch sehen welche ich wegnehmen muss“* (IM4). Die Leitstelle könnte auch einen Plan der verfügbaren Hydrantenpläne auf ihre Lagekarte legen und die Einsatzkraft dann selbst den Weg bestimmen lassen (IM2).

Ortung. Aus dem Wunsch der georeferenzierten Darstellung von Einsatzkräften oder Fahrzeugen ergibt sich die Notwendigkeit der GPS-Ortung. Diese Möglichkeit wird aus zwei Sichten positiv betrachtet: *„[...] die Kollegen sehen das als Eigensicherungskomponente“* (I02), weil die Leitstelle im Bedarfsfall genau den Standort der Einsatzkraft bestimmen kann. Ebenso erhoffen sich das Leitstellenpersonal und die Abschnittsführer durch deren Ortung eine schnellere und genauere Kräfteübersicht. Denn *„wenn man sieht, wo Kräfte sind und wie kann man diese verteilen und sind die ggf. nah und wie weit sind die mit ihrem Einsatz“* (I03), dann können auf dieser Basis schneller Handlungsentscheidungen getroffen werden. Aus Datenschutzgründen sollte eine Einsatzkraft jedoch nicht automatisch überwacht und geortet werden können sondern durch eine spezielle Einstellung dies erst bestätigten (IM3, IM4).

Sensork. Neben der Nutzung der Ortung sind auch andere mobile Sensoren interessant: *„Für den Führungsstab wäre es natürlich schon interessant wenn der ein paar Fotos von der Einsatzörtlichkeit hat“* (I04). Neben Fotos sind auch andere Formate wichtig: *„Video, Foto, Tonaufnahmen, schnell, unkompliziert Texte schreiben können“* (IM3). Allerdings müssen diese Informationen mit Ortsangaben versehen werden, damit diese bewertet werden können (I04, IM1, IM2) und die *„[...] Bilder auch einen eindeutigen Bezugspunkt in der Karte erhalten“* (IM5). So erhoffen sich die Befragten, Missverständnisse bei der Lagedarstellung zu vermeiden. *„Wir erleben das ja immer wieder in den Einsatzleitungen, die ja nicht vor Ort sind, dass wir ein völlig anderes Bild von der Situation haben als die Kräfte draußen“* (IM1). Da sich der Zustand der Einsatzumgebung auch ändern kann, ist neben den Ortsangaben auch eine Zeitangabe von Nöten (IM1).

Annotationen. Neben der Einbindung von externen Informationen scheint es für die Einsatzkräfte hilfreich zu sein, Informationen auf der Karte selbst einzupflegen, darzustellen und gesondert markieren zu können: *„Wenn man da jetzt ein Textfeld hin bauen könnte, wo dann ein Funkrufname oder so stehen würde“* (IM1). Weiterhin war gewünscht *„dass wir die Bereiche selbst einzeichnen, damit ich mir da einen*

Überblick verschaffen kann“ (IM4). So kann beispielsweise ein Abschnittsleiter der Leitstelle durch eine Freihandzeichnung verschiedene Straßensperren oder verschiedene Standorte markieren.

Nachbereitung. Die im Einsatz gesammelten Informationen wären auch für die Dokumentation des Einsatzes interessant: „Einfach Uhrzeit und was passiert. Ich wage zu behaupten, dass es da noch kein Standard [...] gibt“ (IM3). Anhand dieser Dokumentation könnte eine Nachbesprechung des Einsatzes stattfinden: „Da haben wir nachher auch eine Nachbesprechung anhand von Internetvideos, von dem Fotografen und dem Pressemenschen gemacht“ (IM2). Es wird deutlich, dass auf externe Informationen zur Nachbesprechung zurückgegriffen werden musste, weil diese in der Dokumentation sonst nicht vorlagen. „Das kann man schön mit so einer Lagekarte verbinden, dann hat man irgendwann die Möglichkeit, alles aus einer Hand zu erhalten“ (IM3). Jedoch stellt sich die Frage ob eine Aufzeichnung aller Daten oder nur selektiv ausgewählter Daten sinnvoll ist.

Verbindungsausfall. Gerade im betrachteten Szenario eines Stromausfalls sollte eine Möglichkeit der Datensynchronisation bei Verbindungsausfall realisiert werden, damit keine Informationen verloren gehen: „Also diese Tablet PCs müssen gewisse Datensätze vorhalten, bzw. online darauf zugreifen können“ (IM5). „Also das eine Wichtige ist um jegliche Unterschiede zu vermeiden muss diese zu 100% synchron sein mit einer Lagekarte, die die Leitstelle führt“ (IM5). Für diesen Fall müssten auch Daten auf dem Gerät vorgehalten werden, damit eine Weiterarbeit ohne Verbindung ermöglicht wird (IM2).

Kollaboration. Ebenfalls wurden Situationen erkannt, in denen tatsächliche Kollaboration erforderlich ist: „Ich müsste auch in der Lage sein zu kommunizieren um eine gemeinsame Lagedarstellung hinzukriegen“ (IM4). Um Übersichtlichkeit zu ermöglichen, sollten auf einer Karte verschiedene Organisationen für ihre Einsatzkräfte unterschiedliche Farben zur Kennzeichnung erhalten (IM1). Eine Führungskraft ab einer gewissen Hierarchieebene ist nicht mehr in Sichtweite zu seinen Einsatzkräften: „Ich bin ja als Abschnittsleiter auch nicht vor Ort, in aller Regel nicht. [...] Ich soll ja führen, nicht agieren“ (IM4). Nur so könne die Führung der Einsatzkräfte ohne Ablenkung stattfinden. „Aber genau das [gemeint: mobile Kollaborationsanwendung] ist etwas, das den Abschnittsführer entlasten würde und ihn mit Informationen versorgen könnte“ (IM4). Durch die geschilderten Funktionen und Anforderungen erhoffen sich die Befragten für genau diese Führungskräfte eine möglichst schnelle und detaillierte Lageübersicht zur Entscheidungsfindung.

4 Entwicklung eines mobilen Geo-Kollaborationssystems

Ziel der Entwicklung war die prototypische Umsetzung eines mobilen Geo-Kollaborationssystems auf Basis der Ergebnisse der empirischen Studie. Die Kollaboration von Einsatzkräften in der Leitstelle und jenen vor Ort (d.h. in der Abschnittsführung oder am Einsatzort, nicht jedoch im Gebäude) wurde dort als besonders wünschenswert und hilfreich beschrieben. Daher wurde in dieser Arbeit eine mobile Kollaborationsanwendung mit Anbindung an ein bestehendes GIS [26], welches im glei-

chen Projekt entwickelt wurde, erstellt. Um ein möglichst robustes, zuverlässiges und schnell einsatzbereites Werkzeug mit breiter Zielgruppe zu konzipieren wurden als Geräte Smartphones und Tablets gewählt. Da die Anwendung auch zeitweise unabhängig von äußeren Faktoren, wie einer Verbindung im UMTS-Netz oder WLAN, lauffähig sein sollte, wurde eine native Android-Anwendung¹ umgesetzt, die zwar im Gegensatz zu einer mobilen Website mehr an die Technologie und das Betriebssystem gebunden ist, aber auch besser auf die Sensoren zugreifen kann.

4.1 Umgesetzte Aspekte der Kollaborationsunterstützung

Bei der Umsetzung der Funktionalitäten wurde auf ein bestehendes System zurückgegriffen, dessen Funktionen jedoch für mobile Geräte angepasst und um Besonderheiten (Display, Interaktion, Sensorik) erweitert wurden. Die resultierende Anwendung (Abbildung 1) beinhaltet eine Kartendarstellung auf Basis von GoogleMaps (1), eine flexible Informationsdarstellung (2), die Möglichkeit von Annotationen (3), einer an- und ausschaltbaren Ortung (4) und Navigation (5), der Erfassung von Bildern, Videos und Tonaufnahmen (6) sowie dem Kollaborationsmodus zum Teilen von Information (7).

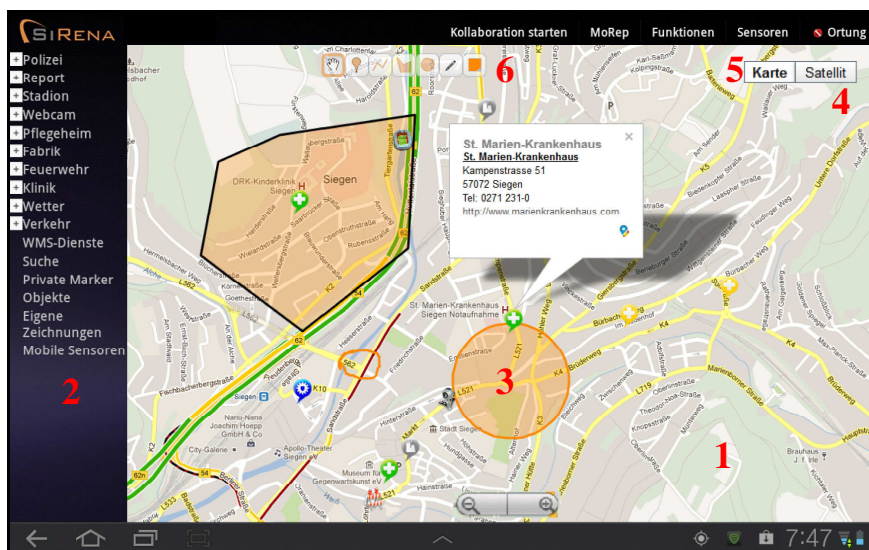


Abb. 1. Screenshot des mobilen Geo-Kollaborationssystems (Tablet)

(1) Kartenanwendung. Zur Umsetzung der Kartenfunktionalitäten wurde die Google Maps API in der Version 3² verwendet, die den Zugriff auf aktuelles Kartenmaterial ermöglicht und auch wichtige Funktionen zur Darstellung von Informationen auf der Karte, wie z. B. Schnittstellen zur Einbindung externer Informationsquellen, bereit-

¹ <https://developers.google.com/android/>

² <https://developers.google.com/maps/documentation/javascript/>

stellt. Nach Authentifizierung in der App werden die externen Daten im Hintergrund geladen um ein sofortiges Arbeiten auf der Karte zu ermöglichen. Durch die Einhaltung von Standards sollen mögliche Wechsel des Kartenmaterials ermöglicht werden.

(2) **Informationsdarstellung.** Externe Informationen, die in der Informationsleiste verwaltet werden, können mithilfe von Web Map Services (WMS) und der auf XML basierende Keyhole Markup Language (KML) als Kartenoverlays geladen werden. Die für verschiedene Einsatzlagen benötigten Kartentypen (z.B. Satellitenübersicht, topographische Karte, schematische Karte) werden bereitgestellt und bieten den Wechsel auch ohne Änderung der angezeigten Informationen an.

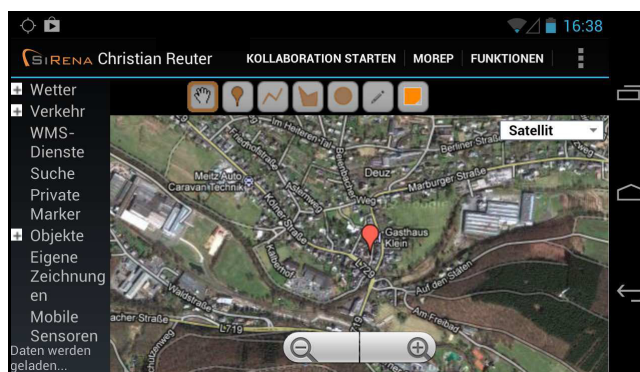


Abb. 2. Satellitenmodus des mobilen Geo-Kollaborationssystems (Smartphone)

(3) **Annotationen.** Im Informationsaustausch bietet es sich an, diese auf der Karte durch verschiedene Zeichnungen oder Texte markieren zu können. So kann einerseits für einen selber als auch für andere schnell ein besonderes Gebiet oder ein Umkreis, beispielsweise bei dem Austritt von Gefahrenstoffen zur Evakuierung, markiert werden. Über der Karte befindet sich die Toolbox, mit der Marker, Kreise, Linien und Polygone in verschiedenen Farben auf die Karte hinzugefügt werden können.

(4) **Ortung.** Durch den eingebauten GPS-Empfänger ist es technisch möglich zu jedem Zeitpunkt den Standort des Benutzers zu ermitteln. In der Anwendung selbst kann die GPS Ortung aus- bzw. eingeschaltet werden kann, um eine ungewünschte Ortung zu verhindern. Durch eine Zustandsüberprüfung innerhalb der Anwendung wird einer Funktion dadurch der Zugriff auf die Ortung gewährt oder verweigert.

(5) **Navigation.** Die Standortinformation kann neben der reinen Lokalisierung von Medien auch zur Navigation dienen. Nach dem Starten der Navigation wird der Benutzer aufgefordert auf den Standort auf Karte zu klicken, zu dem er navigieren möchte. Danach werden der aktuelle Standort als Startkoordinate und der Klickpunkt als Zielkoordinate an die Google Navigations-App übergeben und durch diese eine Routenanweisung erstellt, welche den Benutzer zum Ziel navigiert.

(6) **Mediensensorik.** Mithilfe der Kamera und des Blitzes ist es möglich, schnell und einfach Bilder, Videos und Audioaufnahmen des Einsatzgeschehens zu erstellen und auf der Karte mit dem entsprechenden Standort zu versehen. So können auch andere Einsatzkräfte das entsprechende Bildmaterial betrachten. Neben dem Bild werden automatisch Ort und Zeit erfasst. Unabhängig davon müssen die anderen Ein-

satzkräfte explizit die Darstellung dieser Informationen einstellen, um eine Informationsüberflutung zu vermeiden.

(7) Kollaboration. Die wohl wichtigste Funktion bildet der Austausch von Informationen in Echtzeit. Der Kollaborationsmodus ermöglicht, dass Änderungen an der aktuellen Karte alle anderen zu dieser Sitzung eingeladenen zeitgleich auf ihrem Gerät sehen. Dies kann genutzt werden, um eine Lage, auch interorganisational, verteilt zu besprechen. Um den Kollaborationsmodus zu nutzen, werden WebSockets verwendet, mit denen im Vergleich zu normalen HTTP-Verbindungen eine bidirektionale Verbindung möglich ist, so dass der Server auch ohne konkrete Anfrage des Clients Daten senden kann.

4.2 Mobile Besonderheiten der Kollaborationsunterstützung

Interaktion. Die Steuerung der Karte ist über mobile Wischgesten möglich, sodass bei der Benutzung aus dem mobilen Kontext bekannte Steuerungstechniken eines Touchscreens benutzt werden können.

Kollaboration. Im mobilen Kontext stellt die Kollaboration eine besondere Herausforderung dar, weil eine dauerhafte Verbindung zu einem Server aufrechterhalten werden muss. Unter Berücksichtigung dieser Besonderheit wurde anstatt der Übertragung komplexer Objekte der Austausch von Zeichenketten ermöglicht, so dass beispielsweise nicht ein ganzes Markerobjekt versendet wird, sondern nur die Eigenschaften des Markers (Standort, Name etc.). In der Zielanwendung findet dann erst eine Erstellung des Markerobjekts statt. So kann die Datenlast der Übertragung minimiert werden.

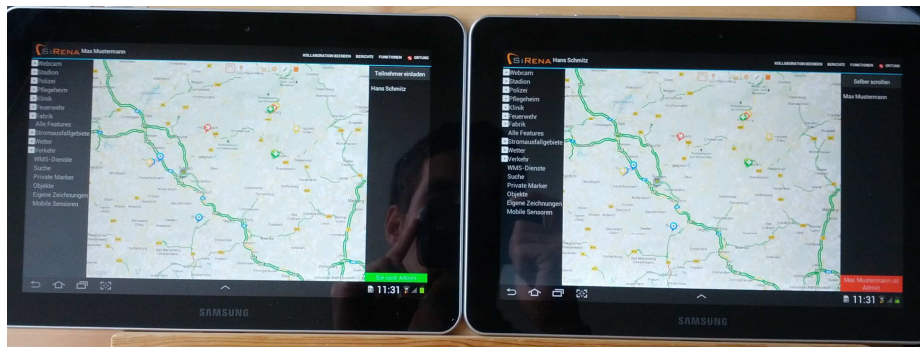


Abb. 3. Kollaborationsmodus des mobilen Geo-Kollaborationssystems (links mit Schreibrechten, rechts ohne)

Datenpersistenz. Um eine Weiterarbeit bei einem Verbindungsabbruch zu ermöglichen, werden die Informationen solange im Gerät zwischengespeichert, bis diese aktualisiert werden. Für einen Verbindungsabbruch stehen so zumindest die Informationen der letzten Aktualisierung zur Verfügung. Auch im Kollaborationsmodus wird der aktuelle Stand auf dem Gerät zwischengespeichert. Während eines Verbindungsverlustes eines Teilnehmers wird so die Weiterarbeit der anderen Teilnehmer ermög-

licht. Änderungen, die in der Zeit eines Verbindungsausfalls durch andere Nutzer veranlasst werden, werden nach Wiedererlangen der Verbindung synchronisiert. Die Kollaboration wird in einer Client-Server Umgebung etabliert werden, weil einerseits die Schreibrechte und andererseits die Daten für neue Mitglieder zentral verwaltet werden. Dieser ist bewusst nicht im mobilen Gerät, sondern auf einem externen Web-server etabliert.

5 Evaluation des mobilen Geo-Kollaborationssystems

5.1 Methodik

Um die empirisch ermittelten und in das System eingeflossenen Erkenntnisse und Anforderungen zu prüfen, wurde eine Evaluation des entwickelten Prototyps durchgeführt. Dazu wurde ein Szenario-basierter Walkthrough mit anschließendem Interview gewählt. Hierbei waren die Teilnehmer aufgefordert, wichtige Fragen und Anmerkungen zur dargestellten Anwendung direkt zu äußern („Thinking Aloud“). An den Evaluationen (je ~45 Minuten) konnten sechs Personen teilnehmen.

Tabelle 3. Durchgeführte Interviews (Phase 3): Evaluation Kollaborationsanwendung

Nr.	Kreis	Organisation	Rolle
IM6	A	Polizei	Leiter des Führungs- und Lagedienstes
IM7	A	Polizei	Abschnittsführer
IM8	A	Polizei	Abschnittsführer
IM9	A	Polizei	Leitungsstab
IM10	A	Polizei	Leitungsstab
IM11	A	Feuerwehr	Oberbrandmeister, Administrator



Abb. 4. Evaluation in der Feuerwehr-Leitstelle

5.2 Ergebnisse

Displaygröße. Gerade mobile Varianten typischer Computer-Anwendungen haben teilweise Einschränkungen der Displaygröße. Mit der Darstellungsgröße sowohl der

Smartphone- als auch der Tabletvariante waren die Befragten zufrieden: „Also ich bin ja der älteste und blindeste in der Truppe hier. Ich würde mir zutrauen, dass ich das noch ohne Brille hinbekomme“ (IM07). Allerdings „[...] bei einer Flächenlage kommen wir damit [Smartphonevariante] natürlich nicht hin, da sind wir uns im Klaren. Aber wenn ich einen Ausschnitt aus dem Innenstadtbereich oder sonst irgendetwas brauche“ (IM06) wurde der Einsatz eines Smartphones als ausreichend betrachtet. Vor allem bei der Smartphonevariante wurde jedoch der Wunsch geäußert, die Informationsleiste „[...] ein- und ausblendbar [zu machen] [...] um die Karte maximal darzustellen“ (IM11).

Kartenmaterial. Alle Befragten kannten die verwendeten Google Maps Karten bereits aus dem privaten Umfeld: „Jetzt etwas anderes hier einzubringen, macht meines Erachtens keinen Sinn, weil das kennt jeder, und kann man nutzen, ist auf dem aktuellen Stand“ (IM11). Neben der Aktualität wird auch der Satellitenmodus von Google Maps für wertvoll erachtet (IM06, IM11). Der Einsatz des Satellitenmodus wurde bereits in der Empirie als maßgeblich für die Lagebeurteilung bezeichnet.

Sensorik. Auch zu den einzelnen Funktionen äußerten sich die Befragten. Vor allem die Verwendung der mobilen Sensoren wurde von den Befragten als positiv betrachtet: „Ich muss die [Ortung] explizit hier drin noch einmal anschalten?“ (IM11). Dass die GPS Ortung zusätzlich in der Anwendung eingeschaltet werden muss, akzeptierten die Befragten. Dennoch machten sie sich die GPS Funktion zu nutzen, beispielsweise bei der Erstellung eines neuen Einsatzes vor Ort, indem sie „[...] klicken um zu sagen, hier bin ich jetzt“ (IM11). Bei der Informationserstellung mit den mobilen Sensoren (Kamera und Mikrofon) äußerten die Befragten, dass die Erfassung von Ort und Zeit der Erstellung besonders wichtig sei. Aber diese Information sei vor allem für Dokumentationszwecke interessant: „Das muss nicht mit angezeigt werden, das muss im Hintergrund sein, wo man sich [...] eine Chronologie ausdrucken kann“ (IM06).

Kollaboration. Besonderer Schwerpunkt in der Evaluation wurde auf die Demonstration des Informationsaustausches gelegt. „Ich finde das total klasse“ (IM06). Insbesondere in der Verwendung dieses Informationsaustausches in einem Einsatzfall sehen die Befragten eine Verbesserung ihrer Arbeitspraxis. „Ich kann ja in der Führungsgruppe auch noch andere Informationen gebrauchen, wenn ich mit mehreren Abschnitten arbeite, dass ich mir einfach Abschnittskarten ablege und die dann immer wieder aktualisiert bekomme“ (IM06). So fungieren die Abschnittsführer als Informationslieferant und in der Führungsgruppe laufen die Informationen dann automatisch zusammen. Insbesondere die Funktion, „schick mir die Karte was wir bis jetzt gemacht haben“ (IM07) wurde positiv bewertet.

Verbesserungspotentiale. Obwohl der Informationsaustausch insgesamt positiv bewertet wurde, sind einige Verbesserungsvorschläge geäußert worden. Vor allem das aktuelle Verhalten, dass immer nur Einer zu einem Zeitpunkt Änderungen an der geteilten Karte vornehmen kann, wurde in einigen Situationen als problematisch bezeichnet: „Jetzt bekommt der Mann vor Ort eine wichtige Information, die will er gerne in seiner Karte noch eintragen“ (IM07). Eine Person muss also auch Informationen nur für sich eintragen können. „Es würde eigentlich Sinn machen, dass eine zentrale Lagekarte durch den Stab geführt wird und Abschnittskarten dem Stab zur

Verfügung gestellt werden“ (IM06). „Das hätte immer den Vorteil: Der eine kann mit seiner Karte draußen immer weiter arbeiten in dem Abschnitt“ (IM06). Dies entspricht den Vorschlägen der Literatur [11]. Da trotzdem die Möglichkeit bestehen soll, die Informationen der Kollaboration zu erhalten, jedoch nicht zwangsläufig an ihr teilnehmen zu müssen, ist schließlich der Vorschlag nach der Einbindung der Kollaborationsinformationen ähnlich wie eine normale Informationsressource entstanden (IM06). Ebenso wurde bei den Elementen der Toolbox zum Markieren von Gebieten auf der Karte die Bitte gestellt, einen festen Radius eingeben zu können. Bisher ist nur die Veränderung des Kreisradius per Drag & Drop möglich. Exakte Radien sind in gewissen Einsatzsituationen, wie zum Beispiel einer Evakuierung, immens wichtig. „Wenn ich jetzt so ein Radiusymbol darauf lege, ob ich dann nicht eine Größe vorgeben kann für den Radius und dann eingeben“ (IM06).

6 Fazit

Dieser Beitrag konnte eine Möglichkeit aufzeigen, wie die Arbeit von Einsatzkräften von Feuerwehr und Polizei durch den Einsatz mobiler Kollaborationssysteme unterstützt werden kann. Nach der Diskussion verwandter Arbeiten hat die im Rahmen dieser Arbeit durchgeführte empirische Studie Facetten der Einsatzpraxis mobiler Einsatzkräfte erörtert und Anforderungen ermittelt. Neben Faktoren wie Dringlichkeit und Komplexität der Einsatzpraxis konnte eine breite potentielle Zielgruppe mobiler Anwendungen ermittelt werden, auch um die Informationsbeschaffung und Awareness sowohl in der Leitstelle als auch vor Ort zu verbessern. Mobile Sensoren der Ortung wurden als besonders hilfreich für die Lageübersicht empfunden, Fotos, Videos und Tonaufnahmen als mögliche Informationsartefakte. Wichtig ist, dass auch bei zeitweisem Verbindungsausfall eine Weiterarbeit möglich sein soll.

Auf Basis der Anforderungen wurde ein mobiles Geo-Kollaborationssystem mit Anbindung an ein GIS-System umgesetzt, welches mit Android auf ein aktuelles, weit verbreitetes und zuverlässiges Betriebssystem aufbaut. Die Realisierung des Kollaborationsmodus, zur gemeinsamen Betrachtung einer Lage, wurde mithilfe von WebSockets, einem auf TCP basierenden Netzwerkprotokoll, welches bidirektionale Verbindungen zwischen einer Webanwendung und einem WebSocket-Server herstellen kann, umgesetzt. Die Kombination dieser beiden Technologien machte die Erstellung im dargestellten Funktionsumfang für mobile Geräte bei möglichst geringen Kosten, durch die Verwendungsmöglichkeit gängiger und verbreiteter Smartphones, erstmalig mit einer relativ großen Nutzergruppe möglich, und ist somit auch für ehrenamtliche Einheiten interessant. Nicht zuletzt durch den Austausch von reinen Zeichenketten anstatt komplexer Objekte konnte der Kollaborationsmodus gerade im mobilen Kontext vorhandene Datenverbindungen effizient nutzen und so die Datenlast mindern. Der Prototyp zeichnet sich auch durch die Speicherung von Daten im Falle des Verbindungsverlustes und die Weitergabe von bereits erarbeiteten Informationen (wie Markern und Annotationen) an neue Kollaborationsmitglieder aus. Somit erhalten jegliche Kollaborationsteilnehmer alle Informationen des Geschehens vor

Ort. Lediglich die Karten sind momentan noch nicht dauerhaft auf dem Gerät gespeichert.

Zwar konnten die erhobenen Anforderungen technisch umgesetzt und somit durch Software unterstützt werden, jedoch ergaben sich aus der Evaluation Verbesserungspotenziale. Neben einigen Details ist insbesondere die Funktion, Karten zu persistieren und somit zur Dokumentation und Nachbesprechung der Krisensituationen zu verwenden, eine wichtige Verbesserungsmöglichkeit. Die selektive Teilung eigener Kartenlayer (bspw. Straßensperren der Polizei für die Feuerwehr) wurde als weitere sinnvolle Funktion angesprochen. Weiterhin fokussiert die Arbeit derzeit nicht den Einsatz z. B. in Gebäuden. Durch eine Umsetzung dieser Verbesserungsvorschläge kann es möglich werden, dass ein Mitarbeiter in der Leitstelle authentische Informationen über die Lage "durch Augen mehrerer Einsatzkräfte" erhält und so noch schneller effektivere Handlungsentscheidungen treffen kann. Damit wird die Unterstützung emergenter Kollaboration, deren Teilnehmer, Zeitpunkt und Ort im Vorhinein unbekannt sind, ermöglicht.

Danksagung: Diese Forschungsarbeiten wurden durch ein Stipendium der Research School Business & Economics ermöglicht und im Rahmen des Verbundprojektes 'infostrom' durch das Bundesministerium für Bildung und Forschung (Fö.-Kz. 13N10712) unterstützt.

Literatur

1. Ley, B., Pipek, V., Reuter, C., Wiedenhofer, T.: Supporting Improvisation Work in Inter-Organizational Crisis Management. In: Proc. CHI, pp. 1529-1538. ACM, NY (2012)
2. Reuter, C., Marx, A., Pipek, V.: Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations. International Journal of Information Systems for Crisis Response and Management (IJISCRAM) 4 (1), 1-16 (2012)
3. Meißner, A., Steinebach, M.: Neue IT-Infrastrukturen im Notfall- und Rettungswesen - Potential und Risiko. In: von Knop, J.: Netz- und Computersicherheit: Sind wir auf einen Angriff auf unsere Informationssysteme und Informations-Infrastrukturen vorbereitet? (2004)
4. Kunze, C., Rodriguez, D., Shammas, L., Chandra-Sekaran, A., Weber, B.: Nutzung von Sensornetzwerken und mobilen Informationsgeräten für die Situationserfassung und die Prozessunterstützung bei Massenanfällen von Verletzten. In: Informatik 2009 - Im Focus das Leben. LNI, Vol. P-154, pp. 1465-1478. GI, Bonn (2009)
5. Nestler, S., Klinker, G.: Mobile computing in mass casualty incidents (MCIs). In: Workshop Mobile Informationstechnologien Mobiles Computing in der Medizin MoCoMed 2009, pp. 1-15 (2009)
6. Laufs, U., Zibuschka, J., Roßnagel, H., Engelbach, W.: Entwurf eines Multi-touch-Systems für die organisationsübergreifende Zusammenarbeit in nicht-operativen Phasen des Notfallmanagement. In: Heiß, H.-U., et al. (eds.): Informatik 2011. LNI, Vol. P-192, pp. 342 ff. GI, Bonn (2011)
7. Fajardo, J.T.B., Oppus, C.M.: A Mobile Disaster Management System Using the Android Technology. WSEAS Transactions on Communications 9 (6), 343-353 (2009)
8. Liu, S.B., Palen, L.: Spatiotemporal Mashups: A Survey of Current Tools to Inform Next Generation Crisis Support. In: Proceedings ISCRAM, Gothenburg (2009)

9. Birkhäuser, B., Pottebaum, J., Koch, R.: Unterstützung von Einsatzentscheidungen der Feuerwehr auf Basis IT-unterstützter Kräftekoordination. In: Proceedings Informatik 2009 - Im Focus das Leben. LNI, Vol. P-154, pp. 1393–1396, GI, Bonn (2009)
10. Zlatanova, S., Fabbri, A.G.: Geo-ICT for Risk and Disaster Management. In: Scholten, H.J., Velde, R., van Manen, N. (eds.): Geospatial Technology and the Role of Location in Science Geospatial Technology and the Role of Location in Science. Springer (2009)
11. Convertino, G., Mentis, H.M., Slavkovic, A., Rosson, M.B., Carroll, J.M.: Supporting common ground and awareness in emergency management planning. *ACM Transactions on Computer-Human Interaction* 18, 1–34 (2011)
12. Wu, A., Zhang, X.L., Convertino, G., Carroll, J.M.: CIVIL : Support Geo-collaboration with Information Visualization. In: Proceedings Group, pp. 273–276 (2009)
13. Stevens, G., Wulf, V.: Computer-supported access control. *ACM Transactions on Computer-Human Interaction* 16, 1–26 (2009)
14. Puras, J., Iglesias, C.A.: Disasters 2.0. Application of Web 2.0 technologies in emergency situations. In: Proceedings ISCRAM, pp. 111–126 (2009)
15. Brinkhoff, T., Bertling, M., Biermann, J., Gervens, T., König, R., Kümper, D., Neis, P., Stollberg, B., Rolfs, C., Weiser, A., Weitkämper, J., Zipf, A.: Offenes Katastrophenmanagement mit freiem GIS Zur interoperablen Kopplung von Leitstellensystem, mobilen Clienten und GDI mit Prozessierungsdiensten. In: Symp. für angewandte Geo.-Inf. (2008)
16. Monares, Á., Ochoa, S.F., Pino, J.A., Herskovic, V., Rodriguez-Covili, J., Neyem, A.: Mobile computing in urban emergency situations: Improving the support to firefighters in the field. *Expert Systems with Applications* 38, 1255–1267 (2011)
17. Marchese, M., Mcneill, F.: Interaction models to support peer coordination in crisis management. In: Proceedings ISCRAM, pp. 230–241 (2008).
18. Neuvel, J., Zlatanova, S.: The void between risk prevention and crisis response. In: Proceedings UDMS, pp. 1–14 (2006)
19. Schafer, W., Ganoë, C.H., Carroll, J.M.: Supporting Community Emergency Management Planning through a Geocollaboration Software Architecture. *Journal Computer Supported Cooperative Work* 16, 501–537 (2007)
20. Schiller, F., Eichler, G., Erfurth, C., Späthe, S., Rossak, W.: Geolocated Communication Support in Rescue Management. In: Proceedings ISCRAM (2011)
21. Cai, G., MacEachren, A., Sharma, R.: Enabling GeoCollaborative crisis management through advanced geoinformation technologies. In: Proceedings National Conference on Digital Government Research, pp. 227–228 (2005)
22. Singh, G., Ableiter, D.: Twiddlenet: Smartphones as personal content servers for first responders. In: Löffler, J., Klann, M. (eds.): Mobile Response. LNCS, Vol. 5424, pp. 130–137. Springer (2009)
23. Catarci, T., Leoni, M. De, Marrella, A.: The WORKPAD Project Experience: Improving the Disaster Response through Process Management and Geo Collaboration. In: Proceedings ISCRAM (2010)
24. Reuter, C., Pohl, P., Pipek, V.: Umgang mit Terminologien in inter-organisationaler Krisenkooperation - eine explorative Empirie. In: Eibl, M. (ed.): Mensch & Computer 2011. Oldenbourg, München (2011)
25. Strauss, A.L.: Methodologische Grundlagen der Grounded Theory. In: Strübing, J., Schnettler, B. (eds.): Methodologie interpretativer Sozialforschung: Klassische Grundlagentexte. UVK, Konstanz (2004)
26. Ley, B., Pipek, V., Reuter, C., Wiedenhofer, T.: Supporting Inter-organizational Situation Assessment in Crisis Management. In: Proceedings ISCRAM, Vancouver, Canada (2012)

Entwicklung neuer Services in Lehre und Praxis – ein Abgleich von personellem Angebot und Nachfrage

Peer Benholz und Helena Preiß

Otto-Friedrich-Universität Bamberg, Lehrstuhl für Supply Chain Management, Bamberg,
Germany

peer.benholz@stud.uni-bamberg.de

helena.preiss@uni-bamberg.de

Abstract. In Zeiten der Evolution hin zu einer Dienstleistungsgesellschaft sehen sich Hochschulabsolventen neuen Berufsmöglichkeiten im Bereich der Serviceentwicklung gegenüber. Dies betrifft sowohl Services aus Sicht der Wirtschaftsinformatik als auch aus Sicht der Dienstleistungsbranche. Aber welche Anforderungen stellen die Unternehmen an ihre Bewerber und sind deutsche Universitäten in der Lage, ihren Absolventen die benötigten Kompetenzen zu vermitteln? Dieser Beitrag gibt mit Hilfe einer Stellenanzeigenanalyse und einer Analyse von Vorlesungsbeschreibungen eine Antwort und deckt auf, dass es vor allem im Bereich des Service Engineering, also der methodischen Entwicklung von Dienstleistungen, einen Nachholbedarf seitens der Universitäten gibt. Besonders im Bereich der praktischen Erfahrungen, der Anwendung von Methodenwissen und den sozial-kommunikativen Kompetenzen gibt es Lücken, welche zur Sicherung des Wissensvorsprungs und des Innovationserfolgs in Deutschland geschlossen werden müssen.

Keywords: Hochschulanalyse, Service, Dienstleistung, Service Engineering, Stellenanzeigenanalyse

1 Entwicklung innovativer Services als strategische Alternative

Unternehmen unterschiedlicher Branchen sehen sich veränderten Marktbedingungen gegenüber, die sie zu einer Anpassung ihrer Strategie zwingen. In Zeiten der Globalisierung, des Wandels hin zu einer Dienstleistungsgesellschaft, der Beschleunigung der Taktraten sowie der Verkürzung von Produktlebenszyklen [1], müssen Unternehmen entscheiden, wie sie sich in ihrem Marktumfeld behaupten möchten. Dies gilt besonders für Firmen, die sich bereits jetzt mit einer schwierigen Konkurrenzsituation konfrontiert sehen. Deren Märkte zeichnen sich durch Preisverfall, wachsende Konkurrenz durch ausländische Mitbewerber, Turbulenzen durch großangelegte Mergers & Acquisitions bei gleichzeitig wachsenden Kundenansprüchen hinsichtlich Qualität und Service aus [2-5].

Aber wie können Unternehmen unter solch dynamischen Umweltbedingungen erfolgreich bleiben? Porter (1980) schlägt hierfür zwei grundsätzliche Wege vor: Ent-

weder ein Unternehmen bietet seine standardisierten Produkte zum günstigsten Preis an und übernimmt so die Preisführerschaft oder es differenziert sich vom Wettbewerb durch einzigartige Angebote [6]. Insbesondere der zweite Weg bietet Vorteile. Dazu gehören höhere Gewinnmargen, eine langfristige Kundenbindung und der Aufbau von Wettbewerbsvorteilen, welche die Konkurrenz nur schwierig und mit Zeitverzug einholen kann [6]. In der heutigen Zeit verspricht vor allem das Angebot innovativer Services den erfolgreichen Aufbau einer Nischenstrategie [7-9]. Sie ermöglichen die Entwicklung von schwer zu imitierenden Angebotsbündeln und helfen dem Unternehmen, sich von der Konkurrenz abzuheben [2], [10-11].

Innovative Services sind eine Differenzierungsmöglichkeit für unterschiedliche Branchen und dementsprechend gibt es eine Vielzahl möglicher Ausgestaltungsformen. Allen gemeinsam sind vorausgehende Entwicklungsaktivitäten, welche die Unternehmen vor eine bisher unbekannte Aufgabe stellen. Diese Entwicklung kann mit Hilfe des Konzepts des Service Engineering geschehen, welches allerdings bestimmte Fähigkeiten und Kompetenzen seiner Anwender voraussetzt. Ziel dieser Arbeit ist es zu klären, ob Universitätsabsolventen während ihrer Ausbildung mit den nötigen Fähigkeiten ausgestattet werden, um als sogenannte Service Engineers an der Entwicklung neuer Dienstleistungen mitzuarbeiten. Zur Beantwortung dieser Frage werden nach der Motivation im zweiten Abschnitt die Begrifflichkeiten der Arbeit definiert und die Ergebnisse einer Literaturrecherche bezüglich der Anforderungen an einen Dienstleistungsentwickler vorgestellt. Im dritten Abschnitt wird die Forschungsmethodik präsentiert, anschließend gibt das vierte Kapitel die Ergebnisse hinsichtlich der Forschungsfrage wieder. Die Arbeit endet mit einem Fazit und Empfehlungen, wie die Defizite in der Ausbildung von Service Engineers geschlossen werden können.

2 Service Engineering zur Entwicklung neuer Dienstleistungen

Um ein einheitliches Verständnis zu schaffen, werden zu Beginn die verwendeten Termini definiert. Dies sind zum einen „Service“ aus betriebswirtschaftlicher Sicht und aus Sicht der Wirtschaftsinformatik und zum anderen „Service Engineering“.

Aus betriebswirtschaftlicher Sicht können Service und Dienstleistung als äquivalent angesehen werden [12]. Deshalb soll zur deutlichen Abgrenzung gegenüber dem wirtschaftsinformatischen der betriebswirtschaftliche Servicebegriff im Folgenden durch Dienstleistung ersetzt werden. Grundsätzlich gilt: Auch nach Jahrzehnten der Forschung kann sich die Wissenschaft nicht auf eine einheitliche Dienstleistungsdefinition einigen. Zur Auswahl stehen enumerative, negative und institutionelle Definitionen, wobei keine den wissenschaftlichen Ansprüchen genügt [13]. Deswegen wird sich meist mit einer Definition über konstitutive Merkmale beholfen, um Dienstleistungen zu charakterisieren. Als Beispiel kann hier der Ansatz nach Maleri und Frietzsche (2008) gewählt werden [14]:

Dienstleistungen werden im nachfolgenden als unter Einsatz externer Produktionsfaktoren für den fremden Bedarf produzierte immaterielle Wirtschaftsgüter verstanden.

Die Wirtschaftsinformatik hingegen versteht unter dem Begriff Service „[...] ein softwaretechnisch realisiertes Artefakt zum Anbieten von Funktionalitäten.“ [15]. Er beschreibt somit eine Software oder den Ausschnitt einer Software, welche ausgeführt werden können und dem Nutzer einen Dienst zur Verfügung stellen [15].

Unabhängig davon, welchem Serviceverständnis man folgt, beide müssen erst entwickelt werden, um sie auf dem Markt anzubieten. Werden Services im Sinne einer Software durch die Profession des Software Engineering entwickelt, hat sich für die Entwicklung von Dienstleistungen das Konzept des Service Engineering etabliert [16]. In ihm fließen Erkenntnisse aus den Ingenieurwissenschaften, der Informatik und der Betriebswirtschaftslehre zusammen [17]. Die Einflüsse aus ganz unterschiedlichen Professionen lassen naturgemäß eine Vielzahl von möglichen Definitionen für den Begriff „Service Engineering“ entstehen. Bei genauer Betrachtung der einzelnen Definitionsansätze in [18-22] lassen sich drei gemeinsame Merkmale herausarbeiten. So handelt es sich beim Service Engineering um:

- *eine Systematik zur Entwicklung von Dienstleistungen;*
- *ein Verfahren, welches geeignete Modelle, Methoden und Werkzeuge bereit stellt;*
- *ein Konzept, dass auf einer strukturierten Denkweise beruht.*

Treffend hat diese Bullinger (2003) in seiner Definition zusammengefasst, welche die Grundlage für die hier vorliegende Arbeit bildet. Er beschreibt Service Engineering als „[...] concerned with the systematic development and design of new services using suitable models, methods and tools.“[23].

Die Vorteile, die sich durch einen strukturierten Prozess bei der Entwicklung neuer Dienstleistungen ergeben, sind vielfältig, z. B. die Erhöhung der Qualität durch definierte Entwicklungsaktivitäten und -ergebnisse oder eine verbesserte Projektdokumentation [3], [24]. Diese Vorteile sind aber nur zu erzielen, wenn kompetente Mitarbeiter mit der Entwicklung betraut werden. Für dieses Berufsbild ist die Bezeichnung des Service Engineer bzw. des Dienstleistungsentwicklers entstanden. Dieser muss mit besonderen Kompetenzen ausgestattet sein, um erfolgreich zu sein [22], [25]. Welche dies aber im Einzelnen sind, wurde bisher nur unzureichend ermittelt. Den Autoren ist nur eine empirische Studie bekannt, welche die Anforderungen an Service Engineers aus Unternehmenssicht erhoben hat [26]. Hier fordern die Unternehmen vor allem Fachwissen. Eine tiefere Analyse der deutschsprachigen Literatur zum Service Engineering zeigt darüber hinaus, dass neben Kenntnissen aus Informatik, Ingenieurwissenschaften und der Betriebswirtschaftslehre auch „weiche“ oder persönliche Eigenschaften erwünscht sind. Dazu gehören eine unternehmerische Denkwelt [25], Kreativität, (Entwicklungs-)Erfahrung und Marketingkenntnisse [27]. Doch welche weiteren Ansprüche stellen Unternehmen an die Persönlichkeiten ihrer künftigen Service Engineers und sind deutsche Universitäten in der Lage, diese Kompetenzen zu vermitteln? Mit Hilfe einer Stellenanzeigenanalyse und einer Analyse relevanter Vorlesungen gibt der folgende Abschnitt eine Antwort auf diese Fragen.

3 Methodik zum Abgleich des Bildes des „Service Engineer“ in Lehre und Praxis

Aus der Literatur lässt sich entnehmen, dass der Ausbildung von Service Engineers in Deutschland grundsätzlich bisher wenig Beachtung geschenkt worden ist [25]. Die Arbeiten von Mertins/Spath [26] und von Fähnrich/Opitz [22] nehmen insofern eine Vorreiterrolle ein, als dass sie erstmals aufarbeiten, welche Anforderungen seitens der Unternehmen an einen Service Engineer gestellt werden. Tiefergehende empirische Erhebungen zu diesem Themenfeld sind den Autoren nicht bekannt; die vorliegende Arbeit soll diese Lücke schließen. Um den Status Quo hinsichtlich Angebot und Nachfrage an Hochschulabsolventen zu erheben und gegebenenfalls Verbesserungsvorschläge abzuleiten, kamen eine Untersuchung des universitären Lehrangebots und die sozialwissenschaftliche Methode der Stellenanzeigenanalyse zum Einsatz.

3.1 Die deutsche Universitätsausbildung im Service Engineering

Um das Lehrangebot an deutschen Universitäten zu den Themen Dienstleistungsentwicklung und Service Engineering zu ermitteln und die vermittelten Kompetenzen zu erheben, wurde folgendes, schrittweises Vorgehen gewählt.

1. Schritt: Auswahl der betrachteten Universitäten. In der vorliegenden Arbeit wurde sich bei der Auswahl der Universitäten an der Größe orientiert. Es wurden die, gemessen an der Anzahl der Studierenden im Wintersemester 2011/2012, 25 größten staatlich anerkannten, deutschen Universitäten untersucht [28]. Mit 849.118 eingeschriebenen Studierenden repräsentierten sie rund 55,1% der zu diesem Zeitpunkt eingeschriebenen Studenten [29]. Von den untersuchten Universitäten haben alle 25 eine Fakultät für Wirtschaftswissenschaften, eine Fakultät für Informatik bzw. für Mathematik und Informatik sowie zehn eine technische Fakultät. Die Beiträge [22], [25-27] zeigen, dass vor allem Fähigkeiten aus diesen drei Fachbereichen für angehende Service Engineers von Bedeutung sind und dort die fachlichen Kompetenzen vermittelt werden, die für die Arbeit als Dienstleistungsentwickler erwartet werden.

2. Schritt: Untersuchung des Lehrangebots im Bereich Service Engineering. Zur Untersuchung des Lehrangebots wurde auf die jeweiligen Vorlesungsverzeichnisse zurückgegriffen. Mittels Suchfunktion wurde das online verfügbare Lehrangebot des Sommersemesters 2012 nach den Begriffen „Dienstleistung*“ und „Service*“ durchsucht, wobei das „*“-Symbol als Platzhalter fungierte, welcher ein, kein oder mehrere Zeichen ersetzen kann. Auf diese Weise wurden bei der Suche alle Begriffe mit den Wortstämmen „Dienstleistung“ und „Service“ berücksichtigt. Da Universitäten ihre Veranstaltungen häufig zyklisch, also nur in Winter- oder Sommersemester, anbieten, wurde dieser Vorgang für das Wintersemester 2011/2012 oder, wie im Falle der FernUniversität Hagen, 2012/2013 wiederholt. Bei zwei Universitäten, der Technischen Universität Dresden und der Humboldt Universität zu Berlin, wurde kein frei

zugängliches, zentrales Vorlesungsverzeichnis gefunden, so dass manuell nach Veranstaltungen gesucht werden musste.

3. Schritt: Auswahl der Lehrveranstaltungen. Die Auswahl der Lehrveranstaltungen basierte auf ihrem Titel, um Veranstaltungen auszuschließen, die sich nicht schwerpunktmäßig mit dem Thema Dienstleistungen auseinandersetzen. Die Veranstaltungstitel wurden dahingehend untersucht, ob sie Rückschlüsse auf Inhalte aus dem Bereich der Dienstleistungsentwicklung zulassen. Insgesamt wurden auf diese Weise 27 Veranstaltungen gesammelt und festgehalten, von welchen Fakultäten und Lehrstühlen sie angeboten werden.

4. Schritt: Analyse der vermittelten Kompetenzen. Bei der Analyse der vermittelten Kompetenzen wurden die jeweiligen Syllabi dahingehend untersucht, ob sie für diese Arbeit relevante Kursinhalte (für eine Übersicht vgl. Tabelle 3) oder Lernziele definiert haben. Fehlten freizugängliche Beschreibungen der Veranstaltungen, so wurden die anbietenden Lehrstühle per Mail kontaktiert und um eine Inhaltsangabe gebeten. Es konnten insgesamt zwölf Veranstaltungen herausgefiltert werden, die Kompetenzen vermitteln, welche für die Ausbildung von Service Engineers relevant erscheinen. Anhand der Inhalte und Lernziele wurden dann die vermittelten Kompetenzen beschrieben, gruppiert und mit einer treffenden Bezeichnung überschrieben. Die Veranstaltung „Service Marketing“ der Goethe-Universität Frankfurt am Main wurde aufgrund fehlender Information ausgeschlossen.

3.2 Die Stellenanzeigenanalyse zur Untersuchung der Anforderungsprofile an Bewerber im Bereich der Dienstleistungsentwicklung

Zur Erhebung der Anforderungen, welche die Unternehmen im Bereich der Dienstleistungsentwicklung an ihre Bewerber stellen, wurde die Methode der Stellenanzeigenanalyse gewählt. Diese ist für die Analyse des Qualifikationsbedarfs prädestiniert [30] und hat sich in der Fachliteratur bereits bewährt [31-33]. Mithilfe von empirisch gewonnenen Daten zeigt die Stellenanzeigenanalyse momentane und zukünftige berufliche, personenbezogene und sozial-kommunikative Ansprüche von Unternehmen an spezifische Bewerbergruppen auf. Zudem kann sie den Qualifikationsbedarf verschiedener Berufs- und Tätigkeitsfelder abbilden [34]. Die Vorteile sind offensichtlich: Die verwendeten Stellenanzeigen sind öffentlich und es bestehen keine Zugangsprobleme zu den Informationen, wodurch die hohen Kosten anderer Forschungsmethoden entfallen. Außerdem haben die gewonnenen Informationen einen Prognosecharakter, da die Inserenten zukünftig gewünschte Qualifikationen frei nennen können und nicht etwa von vorgegebenen Qualifikationspotentialen abhängig sind, wie es beispielsweise bei internen Ausschreibungen der Fall ist [35]. Das systematische Vorgehen der Stellenanzeigenanalyse ist im Folgenden dargestellt.

1. Schritt: Auswahl der relevanten Stellenbörsen. Aufgrund des einfachen und schnellen Zugriffs auf eine große Menge von Stellenanzeigen wurden ausschließlich Internet-Stellenbörsen verwendet. Hierzu wurden die fünf Online-Jobbörsen mit der

stärksten Reichweite und den höchsten Besucherzahlen ausgewählt. Diese richten sich zudem an eine allgemeine Zielgruppe und nicht an eine spezifische Branche. Bei der Auswahl wurde sich an der Auflistung der Crosswater Systems Ltd. orientiert [36]. Folgende Internet-Stellenbörsen wurden für die Untersuchung ausgewählt: Arbeitsagentur, StepStone, Monster Deutschland, Gigajob und Jobscout24. Die Jobbörse von Kalaydo wurde nicht verwendet, da auf der Seite nicht ausschließlich Stellenanzeigen, sondern auch eine Vielzahl anderer Anzeigen geschaltet werden und die Besucherzahl der Stellenbörse somit verfälscht wird.

Tabelle 1. Auswahl der relevanten Stellenanzeigen

Suchbegriffe	Ausgewählte Jobbörsen					Σ
	Arbeits-agentur	StepStone	Monster	Gigajob	Jobscout24	
Service Engineering	14	9	13	0	2	38
Service Engineer	25	35	12	15	24	111
Dienstleistungsentwicklung	0	0	0	0	0	0
Dienstleistungsentwickler	0	0	0	0	0	0
Anzahl nach Schritt 2:	39	44	25	15	26	149
Anzahl nach Schritt 3:	3	6	4	2	2	17

2. Schritt: Suchmethodik und verwendete Suchbegriffe. Die ausgewählten Jobbörsen wurden zwischen dem 04.06.2012 und dem 06.07.2012 mehrfach nach passenden Stellenanzeigen durchsucht. In den Suchmasken wurden nach den folgenden Begriffen und Begriffskombinationen gesucht: „Service Engineering“, „Service Engineer“, „Dienstleistungsentwicklung“ und „Dienstleistungsentwickler“. Die Anführungszeichen wurden gesetzt, um sich ausschließlich Treffer mit den exakten Begriffen und Begriffskombinationen anzeigen zu lassen. Dies funktionierte nur bei den Jobbörsen von StepStone und Monster Deutschland. Folglich wurden bei den anderen drei Jobbörsen die hohen Trefferzahlen manuell um unpassende Stellenausschreibungen reduziert. Es wurden lediglich die Stellenanzeigen ausgewertet, welche die oben genannten Begriffe und Begriffskombinationen in ihrem Titel beinhalteten. Da bei diesen Jobbörsen unter dem Suchbegriff „Service Engineer“ häufig der Jobtitel „Service Ingenieur“ angezeigt wurde und sich darunter drei für diese Arbeit relevante Stellenausschreibungen befanden, wurde diese Begriffskombination nachträglich aufgenommen und berücksichtigt. Folglich wurde auf den Jobbörsen StepStone und Monster Deutschland noch einmal nach der Begriffskombination „Service Ingenieur“ gesucht. Um die Erhebung möglichst groß zu halten, wurden in den jeweiligen Sucheinstellungen keine zeitlichen Einschränkungen vorgenommen.

3. Schritt: Auswahl der relevanten Stellenanzeigen. Von den 149 gesammelten Stellenanzeigen wurden jene für relevant befunden, die ein abgeschlossenes Studium voraussetzen und die aktive Entwicklung von neuen Dienstleistungen als Aufgaben-

bereich angeben. Nachdem die Stellenanzeigen um irrelevante und doppelte Anzeigen bereinigt wurden, ergab sich eine finale Größe von 17. Die Zahl erscheint überraschend gering und lässt vermuten, dass die wachsende Bedeutung einer systematischen Dienstleistungsentwicklung bisher noch nicht von den Unternehmen wahrgenommen wird und sich demnach nicht in vakanten Stellen widerspiegelt. Weitere mögliche Gründe könnten zudem sein, dass solche Stellen intern besetzt und demnach nicht öffentlich ausgeschrieben werden oder dass die Dienstleistungsentwicklung nur ein Teilaufgabenbereich in anderen Berufsbezeichnungen ist. Das methodische Vorgehen sowie die Anzahl der ermittelten Stellenanzeigen sind in Tabelle 1 noch einmal übersichtlich dargestellt.

3.3 Der Kompetenzatlas von Heyse/Erpendeck

Um die in den Stellenanzeigen formulierten Anforderungen vergleichbar zu machen, wurden sie in ein Kompetenzschema eingeordnet. In der Literatur findet man zahlreiche Ansätze unterschiedlicher Autoren, z. B. [37-38]; für diese Arbeit wurde der Kompetenzatlas von Heyse/Erpendeck [39] gewählt. Dieser unterscheidet vier Basiskompetenzen: Personale Kompetenz, Aktivitäts- und Handlungskompetenz, Fach- und Methodenkompetenz und sozial-kommunikative Kompetenz. Trotz der Vielzahl unterschiedlicher Ansätze werden diese Basiskompetenzen praktisch von allen Kompetenzforschern akzeptiert [40]. Jede Basiskompetenz besteht aus 16 Kompetenzen, denen die unterschiedlichen Anforderungen aus den Stellenanzeigen zugeordnet werden konnten. Neben dieser hohen Anzahl an vordefinierten Kompetenzen bietet das Modell den Vorteil, dass es unternehmens- und branchenunabhängig eingesetzt werden kann und dass es Überschneidungen der Kompetenzarten berücksichtigt [40].

Auf diese Weise wurden beispielsweise in der Basiskompetenz „Fach- und Methodenwissen“ stellenspezifische Anforderungen, wie Kenntnisse der Kraftfahrzeugtechnik oder der klinischen Technologien, als „Fachwissen“ zusammengefasst. Dahingegen wurden geforderte Deutsch- oder Englischkenntnisse unter „Sprachkenntnisse“ gruppiert und der sozial-kommunikativen Basiskompetenz zugeordnet.

4 Ergebnisse der Stellenanzeigenanalyse und der Analyse des universitären Lehrangebots

Nachdem im vorangegangenen Abschnitt die verwendete Forschungsmethodik vorgestellt wurde, dient das vierte Kapitel der Präsentation der erzielten Ergebnisse. Dazu werden zuerst die Ausbildung von und die beruflichen Anforderungen an Service Engineers aus betriebswirtschaftlicher Sicht vorgestellt, bevor dann ein Exkurs das Thema Service Engineering aus Sicht der Wirtschaftsinformatik thematisiert.

4.1 Ausbildung von Service Engineers im betriebswirtschaftlichen Sinne

Betrachtet man nur die Titel der Veranstaltungen, stellt man fest, dass an den 25 größten Universitäten Deutschlands lediglich vier relevante Veranstaltungen mit dem Titel

„Service Engineering“ und keine einzige mit dem Titel „Dienstleistungsentwicklung“ angeboten werden. Wie bereits in Kapitel 3.1 erwähnt, wurden dennoch insgesamt 27 Veranstaltungen erfasst, die Kompetenzen zum Thema Dienstleistungsentwicklung vermitteln könnten. Tabelle 2 gibt einen Überblick über alle erfassten Veranstaltungen und die im Veranstaltungstitel vorkommenden Begriffskombinationen. Von den 27 Veranstaltungen vermitteln zwölf für diese Arbeit relevante Kompetenzen.

Tabelle 2. Relevante universitäre Veranstaltungen

Begriffskombinationen im Veranstaltungstitel	Anzahl	Davon mit relevanten Inhalten
Service Management	10	2
Service Engineering	5	4
Service Marketing	5	2
Service Innovation	4	2
Andere	3	2
Total	27	12

Betrachtet man ausschließlich diese zwölf Veranstaltungen, sieht die Aufteilung auf bestimmte Fakultäten wie folgt aus: Zehn Veranstaltungen werden an wirtschaftswissenschaftlichen und jeweils eine an einer informatischen und einer technischen Fakultäten angeboten. Anzumerken ist außerdem, dass zehn dieser zwölf Veranstaltungen einen englischen Titel haben. Zur tatsächlichen Unterrichtssprache wurden keine Angaben gefunden.

Um sich ein Bild über die Inhalte und Lernziele der zwölf relevanten Veranstaltungen machen zu können, dient Tabelle 3. Die relevanten Inhalte und Lernziele weichen innerhalb der betrachteten Veranstaltungen stark voneinander ab. So werden Themen aus „Dienstleistungsentwicklung“ und „Service Engineering“ in einigen Veranstaltungen nur angeschnitten, während andere diese als Schwerpunktthemen behandeln. Dennoch konnte aus den Inhalten und Lernzielen abgeleitet werden, welche Fähigkeiten man den Studierenden vermittelt. Da in den meisten Veranstaltungen lediglich Fachwissen und Methoden gelehrt werden, fallen nahezu alle abgeleiteten Kompetenzen in die übergeordnete Kompetenzart der Fach- und Methodenkompetenz nach Heyse/Erpendeck. Innerhalb der Fach- und Methodenkompetenz können vier Gruppen gebildet werden: Theoretische Grundlagen des Service Engineering, Interaktion mit dem Kunden, Anwendung der Grundlagen des Service Engineering sowie praktische Dienstleistungsentwicklung.

- *Theoretische Grundlagen des Service Engineering:* Inhalte, welche Basiswissen über Service Engineering und Dienstleistungsentwicklung vermitteln.
Beispiele: Gestaltung und Steuerung von Dienstleistungsprozessen sowie Vorgehensmodelle, Methoden und Werkzeuge der Dienstleistungsentwicklung.

- *Interaktion mit dem Kunden:* Inhalte, welche die Einbindung des Kunden in den Dienstleistungsentwicklungsprozess hervorheben.
Beispiele: Gestaltung der Integration des Kunden oder das Co-Creation-Konzept.
- *Anwendung der Grundlagen des Service Engineering:* Ziel ist die praktische Anwendung des theoretischen Basiswissens.
Beispiele: Anwendung von Vorgehensmodellen, Methoden und Werkzeugen des Service Engineering.
- *Praktische Dienstleistungsentwicklung:* Veranstaltungen, in denen die Studierenden unter Anwendung der gelernten Vorgehensmodelle, Methoden und Werkzeugen interaktiv eine neue Dienstleistung entwickeln oder eine alte Dienstleistung weiterentwickeln.
Beispiele: Selbstständigkeit oder Teamfähigkeit; hier werden neben Fach- und Methodenkompetenz auch Fähigkeiten aus den anderen Bereichen vermittelt.

Insgesamt zehn der zwölf Veranstaltungen (1, 2, 3, 4, 6, 7, 8, 9, 11 und 12) vermitteln Kompetenzen im Bereich der Grundlagen des Service Engineering, wobei die Veranstaltung 3 explizit Inhalte für die Entwicklung von e-Services vermittelt. Bei insgesamt sechs Veranstaltungen (1, 5, 6, 10, 11 und 12) wird die Interaktion mit dem Kunden im Dienstleistungsentwicklungsprozess hervorgehoben, wobei sich die Veranstaltungen 11 und 12 auf die Entwicklung von IT-Dienstleistungen beschränken. Vier Veranstaltungen (3, 4, 5 und 7) versuchen die praktische Anwendung der Grundlagen des Service Engineering zu lehren, wobei sich die Veranstaltung 3 wiederum auf Methoden für die Entwicklung von e-Services beschränkt. Lediglich in zwei Veranstaltungen (5 und 7) wird im Rahmen des Kurses versucht, unter Anwendung der gelernten Konzepte praktische Dienstleistungsentwicklung zu betreiben. In der Veranstaltung 7 liegt der Fokus allerdings nur auf der Entwicklung von IT-Dienstleistungen.

Tabelle 3. Gesamtübersicht über die relevanten Veranstaltungsinhalte und -ziele

Nr	Veranstaltungsinhalte	Lernziele
1	Gestaltung und Steuerung von Dienstleistungen Betrachtung der Gestaltung der Kundeninteraktion	
2	Prinzipien für die Gestaltung von Dienstleistungen	
3		Entwicklung von e-Service unter Anwendung von Vorgehensmodellen des Service Engineering Alternativen zur Einbindung des Kunden in e-Service-Entwicklung Ausgewählte Methoden zur Gestaltung
4		Fähigkeit, komplexe dienstleistungsrelevante Problematiken mit Hilfe des Service Engineering zu abstrahieren Lösung der Probleme unter Anwendung vermittelter Methoden und Werkzeuge
5		To understand and apply measures of the co-creation of customer value To understand and apply concepts of designing effective customer and employee oriented servicescapes
6	Vorgehensmodelle und Gestaltungsprinzipien des Service Engineering Planung, Gestaltung und Steuerung integrierter Leistungserstellung	
7		Introduce methods, showing planning, developing and controlling of services develop e-services interactively
8	Approaches, methods, tools and efforts to develop service innovations	
9	Wissensvermittlung über Serviceentwicklung	
10	Rolle von Kunden bei Service Innovationen	
11	Grundlagen, Definitionen, Vorgehensmodelle Verständnis für das planmäßige Entwickeln von IT-basierten Dienstleistungen	
12	Siehe 11	

4.2 Berufliche Anforderungen an Service Engineers im betriebswirtschaftlichen Sinne

128 der in den 17 ausgewählten Stellenanzeigen angegebenen Anforderungen konnten in das Kompetenzschema von Heyse/Erpenbeck eingeordnet werden. Auffällig ist, dass die sozial-kommunikativen sowie die Fach- und Methodenkompetenzen, mit 56 respektive 40 Nennungen, klar dominieren, während die personalen und die Aktivitäts- und Handlungskompetenzen lediglich 17 und 15 Nennungen aufweisen. Tabelle 4 gibt eine Übersicht über die am häufigsten geforderten Kompetenzen der Unternehmen innerhalb der jeweiligen übergeordneten Kompetenzart. Die Zahlen in Klammern repräsentieren die Anzahl der Nennungen der einzelnen Kompetenzarten bzw. Kompetenzen. Übergreifend sind die am häufigsten geforderten Kompetenzen dementsprechend die Sprachkenntnisse mit insgesamt 14 Nennungen, das Fachwissen mit 13 Nennungen, die Kundenorientierung und die Kommunikationsfähigkeit mit jeweils zehn Nennungen sowie die Eigenverantwortlichkeit und Selbstständigkeit mit neun Nennungen.

Die Berufserfahrung wurde gesondert behandelt, da sie einen ernst zu nehmenden Anspruch der Arbeitgeber darstellt, jedoch keine Berücksichtigung im Kompetenzatlas findet. Hier ergab sich folgendes Bild: In acht der 17 Stellenanzeigen wurde Berufserfahrung als Voraussetzung genannt. Fünfmal wurden keine genauen Angaben zur Dauer gemacht, in den anderen drei Fällen wurde eine Berufserfahrung von mindestens fünf Jahren gefordert. In sieben Stellenanzeigen wurde angegeben, dass Berufserfahrung wünschenswert sei, d. h. nicht zwingend erforderlich. In zwei Stellenanzeigen wurde keine Angabe zur Berufserfahrung gemacht.

Tabelle 4. Top-Anforderungen innerhalb der Kompetenzarten

Personale Kompetenz (17)	Aktivitäts- und Handlungskompetenz (15)	Fach- und Methodenkompetenz (40)	Sozial-kommunikative Kompetenz (56)
Selbstständigkeit und Eigenverantwortung (9)	Mobilität (6)	Fachwissen (13)	Sprachkenntnisse (14)
Einsatzbereitschaft (6)	Flexibilität (4)	EDV-Kenntnisse (7)	Kundenorientierung (10)
Zuverlässigkeit (2)	Initiative (2)	Projektmanagement (5)	Kommunikationsfähigkeit (10)
-	Innovationsfähigkeit (2)	analytisches Denkvermögen (4)	Teamfähigkeit (8)
-	Belastbarkeit (1)	Lernfähigkeit (3)	Souveränes Auftreten (6)

4.3 Exkurs: Service Engineering im Bereich der Wirtschaftsinformatik

Im Zuge der Auswertung der Stellenanzeigen wurden drei Anzeigen gefunden, die den in Kapitel 3.2 beschriebenen Anforderungen entsprechen, allerdings explizit die (Weiter-) Entwicklung von Software als Aufgabenbereich angeben. Somit wird hier Service nicht im Sinne einer Dienstleistung verstanden, sondern als Software oder als Kombination aus Software und Dienstleistung. Dies kann als ein Indikator gewertet werden, dass die Servicebegriffe aus betriebswirtschaftlicher und informatorischer Sicht konvergieren und sich nicht mehr so dichotom trennen lassen, wie in Abschnitt 2 definiert. Diese Beobachtung findet sich auch in der wissenschaftlichen Diskussion wieder. Unterschiedliche Autoren weisen darauf hin, dass Erstellung und Erbringung von Dienstleistungen ohne den Einsatz moderner Informations- und Kommunikationstechnologien und damit ohne die Hilfe von Softwareartefakten nicht mehr möglich sind [15]. Dies gilt besonders für innovative Dienstleistungsprodukte, mit denen sich ein Unternehmen von der Konkurrenz unterscheiden möchte. Hier bildet die Kombination aus Dienstleistung und unterstützender Software ein schwer zu imitierendes Ressourcenbündel, welches dem Unternehmen zu einem Wettbewerbsvorteil verhilft.

Bei den geforderten Kompetenzen im Bereich Service Engineering aus Sicht der Wirtschaftsinformatik ergibt sich ein ähnliches Bild wie bei den Dienstleistungsentwicklern: Mit elf Nennungen dominieren die sozial-kommunikativen Kompetenzen, gefolgt von den Fach- und Methodenkompetenzen sowie den Aktivitäts- und Handlungskompetenzen mit je sechs Nennungen. Die personalen Kompetenzen weisen lediglich zwei Nennungen auf. Da alle Kurse bis auf Nummer 4 an Fakultäten für Wirtschaftswissenschaften oder Informatik angeboten werden, lässt sich vermuten, dass sie Studenten der Wirtschaftsinformatik ebenfalls zugänglich sind. Bezieht man nun die Überlegung ein, dass zur erfolgreichen Entwicklung von Dienstleistungen die parallele Programmierung von Software nötig ist, erscheinen besonders die Veranstaltungen 3, 7, 11 und 12 interessant. Diese vermitteln speziell Kenntnisse an der Schnittstelle zwischen Informatik und Betriebswirtschaftslehre und fokussieren in ihren Inhalten größtenteils auf IT-Dienstleistungen bzw. e-Services.

5 Fazit und Ableitung von Handlungsvorschlägen

Als Fazit lässt sich zusammenfassen, dass sich die strategische Bedeutung neuer, innovativer Dienstleistungen bisher weder in der universitären Ausbildung noch in den vakanten Stellen von Unternehmen widerspiegelt. In den 25 betrachteten Universitäten lassen sich nur 12 relevante Veranstaltungen identifizieren und von den 149 potenziell interessanten Stellenanzeigen befassen sich nur 17 tatsächlich mit der Dienstleistungsentwicklung. Hier lässt sich ein Nachholbedarf auf beiden Seiten feststellen, welchen es zu beheben gilt, falls sich Deutschland als innovativer Servicestandort etablieren möchte.

Der Abgleich zwischen personellem Angebot und Nachfrage deckt weitere Defizite auf. So werden bis auf eine Ausnahme alle relevanten Vorlesungen an wirtschaftswissenschaftlichen Fakultäten gehalten, obwohl die Unternehmen in ihren Stellenanzeigen hauptsächlich nach ingenieurwissenschaftlichen Abschlüssen su-

chen. Desweiteren fordern die Unternehmen vor allem Fähigkeiten aus dem Bereich der sozial-kommunikativen Kompetenzen und erst an zweiter Stelle Fach- und Methodenwissen. Dies steht im Gegensatz zu den vermittelten Kompetenzen an den Universitäten; die Inhalte der analysierten Veranstaltungen umfassen vor allem die Grundlagen und die Modelle und Methoden des Service Engineering. Als dritter Punkt lässt sich zusammenfassen, dass auf beiden Seiten noch Uneinigkeit über die Inhalte und die Aufgaben eines Service Engineers bestehen. So werden auch in den Stellen, die mit Service Engineering überschrieben sind, nicht immer vorrangig Arbeitnehmer zur Dienstleistungsentwicklung gesucht. Meistens handelt es sich dabei nur um eine Tätigkeit neben vielen anderen. Häufig ist mit Service zudem eigentlich ein Softwareartefakt gemeint und in der Tätigkeitbeschreibung geht es um dessen Entwicklung und nicht um Dienstleistungen.

Zur Überwindung dieser Diskrepanzen lässt sich sowohl an Universitäten als auch an Unternehmen eine Reihe von Handlungsempfehlungen formulieren. Für die Universitäten gilt, das Angebot an Veranstaltungen zum Thema Dienstleistungsentwicklung insgesamt zu erweitern und vor allem technisch-naturwissenschaftliche Studenten mit diesem Themenkomplex vertraut zu machen. Für diese Absolventen ergeben sich durch die zunehmende Bedeutung von Dienstleistungen neue, attraktive Betätigungsfelder, die bisher kaum bekannt sind. Zudem ist den Universitäten anzuraten, in ihren Lehrveranstaltungen vermehrt auf die Vermittlung von sozial-kommunikativen Kompetenzen einzugehen, ohne dabei die Fachausbildung zu vernachlässigen. Ein Aufruf, der auch für andere Fachrichtungen gelten kann und sich nicht exklusiv auf das Service Engineering bezieht. Die praktische Einübung des gelernten theoretischen Wissens sollte eine größere Bedeutung bekommen, um die Studenten auf ihre Tätigkeiten in der Wirtschaft vorzubereiten. Für die Unternehmen gilt, dass sie ihre Anforderungsprofile auf Studenten aus betriebswirtschaftlichen Studiengängen ausweiten sollten, da die Fähigkeiten zur Dienstleistungsentwicklung an diesen Fakultäten gelehrt werden. Zusätzlich sollten sie daran arbeiten, das Berufsbild des Service Engineers bzw. des Dienstleistungsentwicklers zu schärfen, um Missverständnissen vorzubeugen.

Abschließend sind kurz die Einschränkungen der hier durchgeführten Analyse zu nennen. So wurden weder alle deutschen Universitäten noch alle Online-Stellenbörsen analysiert; Hochschulen und andere Medien zur Veröffentlichung von Stellenanzeigen wurden nicht betrachtet. Dennoch konnte ein erster, umfassender Einblick in das personelle Angebot und die zugehörige Nachfrage im Bereich der Dienstleistungsentwicklung gewonnen werden. In Zukunft gilt es, diese Erkenntnisse durch eine breitere empirische Basis zu detaillieren und zu erweitern. Ein mögliches Vorgehen wäre hier die Durchführung von strukturierten Interviews mit Hochschul- und Unternehmensvertretern.

Literatur

1. Klaus, P., Hartmann, E., Kille, C.: Top 100 in European Transport and Logistics Services - 2009/2010. DVV, Hamburg (2010)
2. Wagner, S.M.: Innovation Management in the German Transportation Industry. *Journal of Business Logistics* 29 (2), 215–231 (2008)
3. Hoffrichter, M.: Service Engineering - Dienstleistungen systematisch entwickeln. *Information Management & Consulting* 13, 26–30 (1998)
4. Luczak, H., Reichwald, R., Spath, D.: Einleitung. In: Luczak, H., Reichwald, R., Spath, D. (eds.): *Service Engineering in Wissenschaft und Praxis. Die ganzheitliche Entwicklung von Dienstleistungen*. DUV, Wiesbaden (2004)
5. Spath, D., Ganz, W., Tombeil, A.-S.: Introduction. In: Spath, D., Ganz, W. (eds.): *Taking the pulse of economic development*. Hanser, München (2011)
6. Porter, M.E.: *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Fress Press, New York (1980)
7. Cooper, R.G., Edgett, S.J.: *Product development for the service sector. Lessons from market leaders*. Perseus Books, Cambridge (1999)
8. Siegfried, P.: *Angewandtes Service Engineering für KMU*. WHL, Lahr (2010)
9. Bienzeisler, B., Freitag, M., Hofmann, H., Hübener, M.: *Service Engineering internationaler Dienstleistungen*. Fraunhofer, Stuttgart (2010)
10. Drahvszky, A.: Kundenbindung durch innovative Dienstleistungen. In: Bullinger, H.-J. (eds.): *Dienstleistungen - Innovation für Wachstum und Beschäftigung. Herausforderungen des internationalen Wettbewerbs*. Gabler, Wiesbaden (1999)
11. Meiren, T.: R&D Management for Services. In: Spath, D., Ganz, W. (eds.): *Taking the pulse of economic development*. Hanser, München (2011)
12. Bullinger, H.-J., Schreiner, P.: *Service Engineering: Ein Rahmenkonzept für die systematische Entwicklung von Dienstleistungen*. In: Bullinger, H.-J., Scheer, A.-W. (eds.): *Service Engineering. Entwicklung und Gestaltung innovativer Dienstleistungen*. Springer, Berlin (2006)
13. Corsten, H.: *Dienstleistungsmanagement*. Oldenbourg, München (2007)
14. Maleri, R., Fritzsche, U.: *Grundlagen der Dienstleistungsproduktion*. Springer, Berlin (2008)
15. Buhl, H.U., Heinrich, B., Henneberger, M., Krammer, A.: *Service Science. Wirtschaftsinformatik* 1, 60–65 (2008)
16. Meyer, K., Böttcher, M.: *Entwicklungspfad Service Engineering 2.0. Neue Perspektiven für die Dienstleistungsentwicklung*. LUV, Leipzig (2011)
17. Nüttgens, M., Heckmann, M., Luzius, M.: *Service Engineering Rahmenkonzept*. *Information Management & Consulting* 13, 14–19 (1998)
18. DIN e. V.: *Service Engineering. Fachbericht Nr. 75*. Beuth, Berlin (1998)
19. Luczak, H., Sontow, K., Kuster, J., Reddemann, A.: *Service-Engineering. Der systematische Weg von der Idee zum Leistungsangebot. TCW-Report*, München (2000)
20. Gill, C.: *Architektur für das Service Engineering zur Entwicklung von technischen Dienstleistungen*. Shaker, Aachen (2004)
21. Meiren, T., Barth, T.: *Service Engineering in Unternehmen umsetzen. Leitfaden für die Entwicklung von Dienstleistungen*. Fraunhofer, Stuttgart (2002)

22. Fähnrich, K.-P., Opitz, M.: Service Engineering - Entwicklungspfad und Bild einer jungen Disziplin. In: Bullinger, H.-J., Scheer, A.-W. (eds.): Service Engineering. Entwicklung und Gestaltung innovativer Dienstleistungen. Springer, Berlin (2006)
23. Bullinger, H.-J., Fähnrich, K.-P., Meiren, T.: Service Engineering. Methodical Development of New Service Products. *Journal of Production Economics* 85 (3), 275–287 (2003)
24. Burr, W.: Service-Engineering bei technischen Dienstleistungen. DUV, Wiesbaden (2002)
25. Fähnrich, K.-P.: Service Engineering - Perspektive einer noch jungen Fachdisziplin. *Information Management & Consulting* 13, 37–39 (1998)
26. Mertins, K., Spath, D. (eds.): IT-Services. Neue Wege zur professionellen Dienstleistungsentwicklung. Fraunhofer, Stuttgart (2004)
27. Jaschinski, C.: Qualitätsorientiertes Redesign von Dienstleistungen. Shaker, Aachen (1998)
28. Hochschulkompass: <http://www.hochschulkompass.de/hochschulen/download.html>
29. Statistisches Bundesamt: Studierende insgesamt nach Hochschulen, <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Hochschulen/Tabellen/StudierendeInsgesamtHochschulart.html>
30. Alex, L.: Stand der Überlegungen zur Früherkennung der Qualifikationsentwicklung. In: Alex, L., Bau, H. (eds.): Wandel beruflicher Anforderungen. Der Beitrag des BIBB zum Aufbau eines Früherkennungssystems Qualifikationsentwicklung. W. Bertelsmann, Bielefeld (1999)
31. Grob, H.L., Lange, W.: Zum Wandel des Berufsbild bei Wirtschaftsinformatikern – Eine empirische Analyse auf Basis von Stellenanzeigen. *Wirtschaftsinformatik* 38 (2), 236–241 (1996)
32. Gallivan, M., Truex, D., Kvasny, L.: Changing patterns in IT skill sets 1988-2003: a content analysis of classified advertising. *DATA BASE for Advances in Information Systems* 35 (3), 64–87 (2004)
33. Riedl, R., Zwettler, E.-M.: Anforderungen an IT-Personal. *HMD – Praxis der Wirtschaftsinformatik* 272, 81–90 (2010)
34. Säiler, M.: Anforderungsprofile und akademischer Arbeitsmarkt. Die Stellenanzeigenanalyse als Methode der empirischen Bildungs- und Qualifikationsforschung. In: Konrad, F.-M. (eds.): Erziehung & Bildung – Eichstätter Studien. Waxmann, Münster (2009)
35. Dietzen, A., Kloas, P.-W.: Stellenanzeigenanalyse - eine effektive Methode zur Früherkennung des Qualifikationsbedarfs. In: Alex, L., Bau, H. (eds.): Wandel beruflicher Anforderungen. Der Beitrag des BIBB zum Aufbau eines Früherkennungssystems Qualifikationsentwicklung. Bertelsmann, Bielefeld (1999)
36. Crosswater Systems Ltd.: Reichweitenstarke Jobbörsen nach Alexa Ranking, http://www.crosswater-job-guide.com/php_jobmarkt_rang/Jobb_rsen_nach_Reichweite_list.php?orderby=aAlexa_Traffic_Rang_Worldwide
37. Kirbach, C., Montel, C., Oenning, S., Wottawa, H.: Recruiting und Assessment im Internet. Werkzeuge für eine optimierte Personalentwicklung und Potenzialerkennung. Vandenhoeck & Ruprecht, Goettingen (2005)
38. Nieke, W., Lehmann, G.: Kompetenzmodell, www.bildung-mv.de/de/search.html?page=2&query=ziertes (2005)
39. Heyse, V.: Strategien – Kompetenzanforderungen – Potenzialanalysen. In: Heyse, V., Erpenbeck, J. (eds.): Kompetenzmanagement. Methoden, Vorgehen, KODE® und KODE®X im Praxistest. Waxmann, Münster (2007)
40. Heyse, V., Erpenbeck, J.: Kompetenztraining. Schaeffer-Poeschl, Stuttgart (2004)