Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2019 Proceedings

IS Research Methods, Theorizing and Philosophy of Science

DISKNET – A Platform for the Systematic Accumulation of Knowledge in IS Research

David Dann
Karlsruhe Institute of Technology (KIT), david.dann@kit.edu

Alexander Maedche
Karlsruhe Institute of Technology (KIT), alexander.maedche@kit.edu

Timm Teubner
TU Berlin, teubner@tu-berlin.de

Benjamin Mueller
University of Lausanne, benjamin.mueller@unil.ch

Christian Meske
Freie Universität Berlin, christian.meske@fu-berlin.de

See next page for additional authors

Presenter Information

David Dann, Alexander Maedche, Timm Teubner, Benjamin Mueller, Christian Meske, and Burkhardt Funk

DISKNET – A Platform for the Systematic Accumulation of Knowledge in IS Research

Short Paper

David Dann
Karlsruhe Institute of Technology
Karlsruhe, Germany
david.dann@kit.edu

Alexander Maedche
Karlsruhe Institute of Technology
Karlsruhe, Germany
alexander.maedche@kit.edu

Timm Teubner
TU Berlin
Berlin, Germany
teubner@tu-berlin.de

Benjamin Mueller
University of Lausanne
Lausanne, Switzerland
benjamin.mueller@unil.ch

Christian Meske
Freie Universität Berlin
Berlin, Germany
christian.meske@fu-berlin.de

Burkhardt Funk
Leuphana University
Lüneburg, Germany
funk@uni.leuphana.de

Abstract

The accumulation of knowledge is key for any discipline, IS being no exception. With the number of publications, theoretical constructs, and empirical findings growing, surging demand for structuring and meta-analysis is foreseeable. We introduce DISKNET, an online platform that enables the extraction, exploration, and aggregation of construct's definitions, semantic relations, and analytical relations. While these aspects exhibit a rather standardized structure in theory, their practical documentation is non-uniform, highly dispersed, and tricky to seize technically. This has impeded the efficiency and effectiveness of review and meta-analytical processes, and resulted in a fragmented theoretical superstructure. We suggest that tool support for systematic knowledge accumulation is a central step to counteract these issues and to build to a consistent body of knowledge within the IS discipline. The current prototype of DISKNET draws on a large sample of SEM-based studies to demonstrate relevant design principles for a platform for systematic accumulation of knowledge.

Keywords: knowledge repository, construct identity, meta-analysis, nomological network

Introduction

Accumulation of knowledge is seen as key for any scientific discipline, including Information Systems (IS) (Gregor 2006). One important concept to facilitate documenting and communicating this accumulated knowledge systematically is that of *constructs*, which help scientific research to identify, label, and distinguish phenomena encountered in the world and provide vessels to link explanations and predictions (Mueller and Urbach 2017). Typically, such constructs are shared through the process of scientific publishing, which makes them accessible for further scrutiny and refinement as part of the scientific discourse. In this context, the process of reviewing existing literature is a prerequisite for discovering constructs that already exist and what their connections can tell us about the phenomena we seek to study.

Such a review is essential to determine opportunities for theoretical advancements and to create new knowledge. However, constructs are typically embedded in unstructured, non-standardized documents such as manuscripts in conference proceedings and journals or books. While these documents may be archived in electronic libraries in a structured manner, the findings therein are not. At the same time, the number of scientific publications is growing rapidly (Bornmann and Mutz 2015). In IS, several factors exacerbate the situation, including the high degree of multi-disciplinarity and diversity in terms of applied paradigms, reference disciplines, and methods (Gregor 2006). As a result, the IS literature increasingly recognizes the problem of an ever more fragmented theoretical superstructure (e.g. Hovorka et al. 2013; Larsen and Bong 2016; Mueller and Urbach 2017; Rai 2018).

In light of these developments, surprisingly little research addresses the problem that searching for, identifying, and aggregating existing knowledge becomes increasingly inefficient or ineffective: either the (limited) human efforts to manually review a continually growing set of constructs and relations increase, and hence the inefficiency of the review process, or efforts remain the same, consequently decreasing the effectiveness due to a lower share of processed literature. This substantially affects the scientific progress as a whole, limiting its pace and quality. Yet, while the IS discipline pays attention to the methodological aspects and pitfalls of reviewing literature in general (e.g. Boell and Cecez-Kecmanovic 2015; Tate et al. 2015), we observe a scarcity with regard to dedicated approaches and tools for enabling the systematic extraction and analysis of knowledge (in the form of constructs and their relations) from existing literature.

Theories provide "truth" and artifacts provide "utility," both informing each other (Hevner et al. 2004). Given the described circumstances, we argue that without supporting tools, the process of structured knowledge accumulation will become increasingly cumbersome in IS research. Against this backdrop, we report on an ongoing design science project in which our objective is to develop *a platform for systematic accumulation of knowledge within IS research* and explore *requirements and design principles guiding the development of such a platform*. Focusing on constructs, we emphasize their definitions, semantic relationships (i.e., links to other constructs), and analytical relationships (i.e., indicators used to operationalize them). These are supposed to be captured and made accessible in a structured database.

The contribution of our project is twofold. First, we leverage our requirements and design principles to develop a design theory for the class of knowledge accumulation systems. Derived from that theory and principles, we provide a publicly accessible prototype – DISKNET (Digital Scientific Knowledge Network; `http://disknet.org`) – which eventually should be maintained by the IS community itself, as in Wikipedia, to ensure a sustainable accumulation of constructs and the knowledge they document. Our artifact supports scholars when exploring previous findings on the relationship between two or more constructs, therefore increasing research rigor as it will become easier to systematically and comprehensively include established knowledge. Beyond facilitating research activities, this project will enable boundary spanning meta-research on constructs and relations, but also investigations on the operationalization, use, and association of corresponding variables over time. With this, the project may become a mean to counteract the issue of a fragmented theoretical superstructure in IS research we reasoned for earlier.

Conceptual Foundations and Related Work

Core challenges to theoretical work

One of the key problems regarding theoretical work is that of *construct identity*. When studying construct identity, theorists encounter two core fallacies: the jingle fallacy and the jangle fallacy. The *jingle fallacy* describes a situation where two constructs employ the same name but for different conceptual entities (Thorndike 1904). In IS research, for example, Larsen and Bong (2016) find that 'perceived usefulness' is used to represent a variety of phenomena which, upon closer inspection, mean very different things. In their analysis of construct identities, they find that 'perceived usefulness' is equally used to describe users' perceived importance of skill proficiency on job performance as well as for users' belief that a system can enhance job performance. The *jangle fallacy* (Kelley 1927) describes situations in which a specific conceptual entity is referred to by different names. For instance, different research domains might call constructs by different names, even though referring to the same phenomenon. In IS, Larsen and Bong (2016) identify 'switching benefits' to be virtually identical (based on an analysis of the constructs' analytical relations) to some extant conceptualizations of 'perceived usefulness' and conclude that these constructs "[…] are similar enough that researchers using one construct should be aware of the other" (p.531).

Both fallacies obscure conceptually meaningful similarities and differences in nomological networks and thus complicate the discovery of overlaps between such networks. This leads to a set of challenges for IS researchers. First, researchers' ability to identify *boundary spanning constructs* is impeded. Hovorka et al. (2013), for instance, identify several constructs that are used across different research communities within IS. If present, the jangle fallacy makes it challenging to identify boundary spanners because different communities use different names for the same phenomenon. Only meticulous conceptual analysis will reveal whether these different names actually refer to the same construct or not. A jingle fallacy might suggest a whole new set of antecedents, measurements, or consequents for a supposed boundary spanning construct, even though the two constructs mean different things and no actual boundary spanning is present. Second, and continuing this strand of thinking, also formal approaches to knowledge aggregation such as *meta-analysis* (e.g. as in medical research; Borenstein et al. 2009) are complicated. Jingle-related problems make it impossible to merely aggregate insights that seem to revolve around the same construct, consequently increasing the effort required to do such important work. Jangle-related problems may cause errors of omission, which bias and challenge the results of the meta-review. Third, on a higher level, *synthesis and consolidation* to avoid fragmentation of domain knowledge remain challenging. Over time, different studies might investigate a construct embedded in different nomological nets (i.e., semantic relations) or using different indicators (i.e., analytical relations). In settings where the jingle and jangle fallacies create construct identity problems, such discovery and investigation become increasingly difficult and may jeopardize the needed intension, extension, and consolidation of knowledge.

To safeguard against any form of construct identity problem, researchers need to carefully analyze (1) construct definitions and scope conditions, (2) semantic relations (i.e., referential connections to surrounding constructs in the respective nomological nets), and (3) analytical relations (i.e., the manifest indicators used to operationalize the constructs) (Mueller and Urbach 2017; Suddaby 2010). This is a tremendous effort – both quantitatively and qualitatively – because the relevant information is not easily accessible in unstructured manuscripts. Since this imposes a high workload on researchers, we ask how tools can support and improve this process and make achieving the necessary erudition more efficient.

Past approaches

Recognizing these problems – in IS and beyond – scholars have suggested a set of different approaches to make extant theoretical knowledge more easily accessible and comparable. *Repositories* aim at ensuring the efficiency of conceptualization and operationalization in research projects. In Marketing, for example, handbooks with established and standardized scales for essential core constructs are available (e.g. Bearden and Netemeyer 2011). In IS, similar effort has thus far remained on the level of theories as a whole rather than breaking it down to constructs and their measurements. For instance, Dwivedi et al. (2011) provide an edited collection that provides an overview of many of the core theories often used in IS. The same goes for the *Theories used in IS Research* wiki sponsored by the AIS (Larsen and Eargle 2015). A notable exception to this is the work done as part of the Human Behavior Project. It aims to integrate the behavioral sciences through natural language processing algorithms. One of the project's most immediate outcomes is the Inter-Nomological Network (INN; Larsen 2013), a web-based, searchable database of constructs along with their definitions and indicators. This research effort has also produced a first tool for automated construct identity detection based on natural language processing (Larsen and Bong 2016). Their merits notwithstanding, existing repositories face limitations. Those repositories that focus on theories look at these as often isolated nomological networks rendering the analysis of constructs' semantic relations beyond individual theories or even papers difficult. Repositories such as the INN that focus on analytical relations, on the other hand, make it difficult to understand and explore the semantic context of a construct. While good in the respective strengths, we propose that a platform for the systematic accumulation of knowledge in IS has to combine both to facilitate IS research.

Strategies based on *metadata* approach the problem differently. Projects in this vein attempt to overcome the challenges discussed above by providing ontologies and creating some kind of knowledge graph (Auer et al. 2018; de Waard and Kircz 2003). At the core of such knowledge graphs are the creation of models that establish a common understanding of information. While metadata-based approaches can provide abstract, overarching classification systems – for example, a structured representation of problems, methods, and results like in Auer et al. (2018) – they do not leverage the semantic relations among constructs thus far. Attempting to subsume virtually *any* type of scientific knowledge (e.g. results, ontologies, mathematical proofs) may be a stretch for what a single tool can handle.

We therefore propose to focus on knowledge encoded within constructs and the semantic and analytical relations between them. Further, we propose that there are three specific requirements that are currently not addressed by established approaches. First, most repositories and databases focus on either (rudimentary) semantic *or* analytical relationships. Moreover, these repositories do not provide insights into observed relationships between constructs. Here, we propose that a platform exploring, and aggregating knowledge needs to capture *both* the semantic and analytical perspective, such that a construct's full set of relations becomes accessible for analysis. Therefore, a UI-guided extraction process is needed. Second, constructs must not be strictly linked to any one nomological network they have been used in in the past, but reuse of particular constructs needs to be searchable and discoverable. Accordingly, a stand-alone recording of individual constructs and their relations is more important than a full capture of theories as a whole. Finally, scholars must be able to condense knowledge by collating several publications into a single research project first, then aggregating respective constructs using their domain knowledge.

Our project draws on empirical studies based on structural equation modeling (SEM) as a substantive context for the prototype we are developing in the first design cycle. SEM has become a quasi-standard approach in many fields as it allows researchers to test theories and concepts with relative ease (Hair et al. 2012). In fact, about 20 percent of all publications in the major IS outlets employ structural equation modeling (Urbach and Ahlemann 2010). Most important for our efforts, we propose that SEM-based research is characterized by relatively standardized ways of reporting results. This renders SEM-based work particularly suitable for the early stages of our design project because it allows easy access to (1) construct definitions and scope conditions, (2) semantic relations, and (3) analytical relations. Its utility for our purpose notwithstanding, we acknowledge that SEM-based work is but one type of studies that engage with constructs in IS research and that the discipline's theoretical work is much broader and richer. Nonetheless, the high degree of standardization in the means of representation employed by SEM-based work allows convenient access to the three aspect of construct clarity. This helps us to focus on building a viable prototype first and then expanding it for other types of research used in IS in later design cycles.

The Design Science Research Project

We follow a design science research (DSR) process as described by Kuechler and Vaishnavi (2008). Following a DSR paradigm is a promising approach because we not only attempt to understand and describe the challenges that researchers face in knowledge accumulation, but aim to provide a solution supporting scholars in actual knowledge accumulation. Beyond DISKNET, the actual solution (i.e., prototype), we also leverage our requirements and design principles to develop a design theory (Gregor and Jones 2007) for the class of knowledge exploration and accumulation systems. Overall, our DSR project consists of three related cycles, depicted in Figure 1. Within this short paper, we describe the intermediate results achieved during the first cycle (grey boxes).

Cycle 1: In the first design cycle, our goal is to establish foundational capabilities of the platform in order to enable extraction, exploration, and aggregation of knowledge. The foundation of Cycle 1 is a problem exploration rooted in our team's experience as authors and editors. Across the team, we represent a mix of backgrounds, research fields, and career stages and have explored our experiences with respect to finding, analyzing, and working with constructs as a starting point for our design work. The initial design principles that result from this work were and continue to be subject to a continuous formative evaluation with different stakeholders (e.g. editorial colleagues, doctoral students) and we plan to run a summative evaluation at the end of this cycle. The result of this cycle is making the DISKNET platform available for the IS community in a beta version. We are not aware of any large-scale repository of constructs used in and results obtained through SEM-based research. To date, this data remains dispersed across a plethora of documents. Even within single documents, the relevant information is often fragmented across the text, tables, and diagrams (which often are not machine-readable). We hence argue that this matter's relevance, accessibility, and the existence of an apparent gap provide a promising starting point for our endeavor.

Cycle 2: The second cycle focuses on running a pilot knowledge accumulation study using our platform. Focus will be on running a meta-study and exploring further specific needs and requirements within the knowledge aggregation step. We plan to run a comparison study where independent researchers perform meta-studies using traditional means vs. our platform. This cycle will thus place emphasis on out tool's exploration and aggregation aspects. What results is a revised prototype of DISKNET with a comprehensive evaluation of all of its core design principles, also employing naturalistic evaluation approaches.

Cycle 3: In cycle 3, we emphasize automation and requirements important to adoption. This far, focusing on SEM-based research as a starting point for our DSR project is driven by the accessibility of the relevant data. Yet, our manual approach to identifying relevant publications and extracting knowledge is limited in terms of scalability. We will hence work toward automating the identification and extraction procedures. Using labeled data from our platform, we plan to employ and evaluate machine learning techniques for identifying relevant publications as well as constructs and relations included therein; related efforts are ongoing and have produced first prototypes (Dann et al. 2017; Ludwig et al. 2020). While full automation may not be (reliably) feasible in the near future, we propose that a synergetic collaboration between researchers and automated procedures will be fruitful. Beyond automation, we propose that from Cycle 3 onwards issues related to the adoption and continued use of our work by the community need to be addressed. We will place great emphasis on additional requirements regarding the governance structure behind our platform because questions pertaining to maintenance, expansion, and quality assurance are central to our work's ability to have an impact.

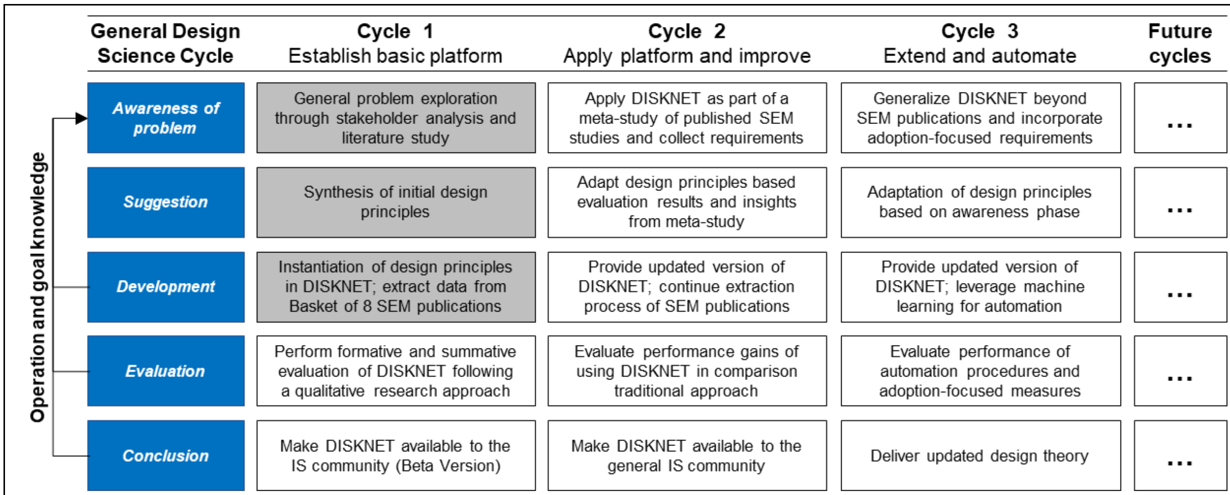

Figure 1: DSR project; adopted from Kuechler and Vaishnavi (2008)

Requirements and Design Principles

First, especially in the case of SEM-based research, it is essential to not only identify existing and related publications but also understand the existing knowledge graphs (key constructs and relationships) investigated in the corresponding studies. This typically requires performing a literature review or a meta-study. There exist several challenges. For instance, Abrami et al. (1988) identify specifying inclusion criteria, locating studies, coding study features, calculating individual study outcomes, and data analysis. Second, in order to actually perform a SEM-based study, key constructs and relationships need to be identified, and a research model needs to be defined. From a conceptual point of view, we can classify the various challenges in processing existing knowledge into three phases: extraction, exploration, and aggregation. We articulate a set of key requirements based on the challenges discussed above (Table 1).

A foundational requirement (R1a/1b) for supporting the extraction process is to enforce data quality by implementing a structured and guided workflow to capture semantic and analytical relations within a central data store. To scale the extraction process, it should be possible to enable multiple users to perform work on individually assigned papers. Based on the extracted knowledge graphs, a subsequent step is to explore the collected data and examine the use of constructs across a set of literature. Thus, the platform should provide a simple keyword-based search (i.e., within publications' title/abstract/keywords) with a focus on discovering publications and the contained knowledge graphs (R2). Finally, as part of the aggregation step, the platform should empower scholars to assign explored publications into aggregation projects (R3a) and allow performing aggregation-specific activities (R3b). Domain experts may now leverage their knowledge to merge multiple constructs (considering potential jingle/jangle issues) into aggregated meta-constructs. Thereby, metrics such as the number of reported relations, mean measured effects, or observed contrary effects become apparent and may open paths for future research.

Table 1: Requirements (EXTR = Extraction, EXPL = Exploration, AGGR = Aggregation)		
Requirements		**Description**
EXTR	R1a	The platform should support the entire workflow of extracting knowledge graphs from publications in a multi-user environment with a security and logging concept building on a central data store.
	R1b	The platform should support a UI-guided process for entering the data enforcing data quality through graph-based constraints.
EXPL	R2	The platform should allow simple keyword-based search and present results as search results (incl. visualization) for exploration purposes.
AGGR	R3a	The platform should enable creating dedicated aggregation projects using a subset of the extracted publications.
	R3b	The platform should support aggregation project-specific activities.

Building on these identified requirements, we derive an initial set of design principles for the platform, listed in Table 2. We also map these design principles to the corresponding requirements.

Table 2: Initial design principles (DP)		
Design Principles	**Description**	**Requirements**
DP1	Provide an extraction backend to support the manual process of knowledge extraction efficiently and effectively.	R1a, R1b
DP2	Provide a simple search interface to support exploration of existing publications.	R2
DP3	Provide specific front- and backend functionalities for aggregation projects to allow for further processing.	R3a, R3b

Development and Software Implementation

The initial design principles introduced above have been instantiated through a web-based platform. Here, we shortly introduce the underlying technologies leveraged for implementing the platform. Subsequently, we explain the core functionalities along the activities of extraction, exploration, and aggregation.

The platform utilizes Django, a Python-based open-source framework for web applications. It provides a variety of pre-implemented methods and functions for creating generic views for data acquisition within a dedicated backend with restricted access. Specific tasks and views are written in custom Python, HTML, and JavaScript. Using the framework's object-relational mapper, data is stored in a MySQL database. The platform operates on Apache2 virtual hosts on a Unix-based server architecture. Collaborative source code development and review is managed through a version control system (Git).

We instantiate DP1 by providing an extraction backend, only accessible by registered users. Publications are uploaded and described with metadata. The knowledge graph(s) are added as images (e.g. graphical research model, result tables). Given SEM-based research's high degree of standardization, we provide dedicated fields for entering, for instance, path coefficients, significance, and number of observations.

Figure 2 depicts mockups of the user interface to support the aggregation process. We use an intuitive front-end that allows users to search within publications' title, abstract, and keywords with ease. This allows for the exploration of constructs and relations within a certain theme (e.g. research on "flow" or "Taobao.com"). Users will be presented with a list of all constructs used within the retrieved results. As depicted in Figure 3, they are then able to aggregate constructs into *meta-constructs* using an intuitive drag-and-drop logic (Cheung 2015). Users can drag available constructs to an aggregation-box to create and name meta-constructs for the subsequent analysis (e.g. aggregating "Booking Intention", "Switching Intention", and "Intention to Rent"). This manual process is supported by, for instance, showing construct definitions and items on demand (e.g. through mouse hovering).

Moreover, based on construct similarity scores, candidates for merging can be proposed by the system. Importantly, ultimate merging decisions remain with the human user because different research contexts may warrant different aggregation. For instance, constructs such as customers' "intention to purchase" and "intention to rent" may be merged when investigating the impact of "sales personnel training" on firms' success (i.e., focus on intention as a general outcome) but not when the research aims to investigate the different drivers of buying and renting decisions explicitly. Furthermore, a typical case refers to *similar but converse* constructs (e.g. "intention to use" vs. "intention to evade"). It may be useful to aggregate these into a common meta-construct, where the researcher will need to flag one as "reverse." Once meta-constructs are defined and populated, the platform creates a meta-graph (requires two or more underlying effects). Given that SEM-based papers usually report sample size, standardized path coefficients, and significance levels (and consequently boundary values for standard errors), we will be able to provide meta-analytical assessments, for instance, on weighted average effect sizes, or effect controversialness.

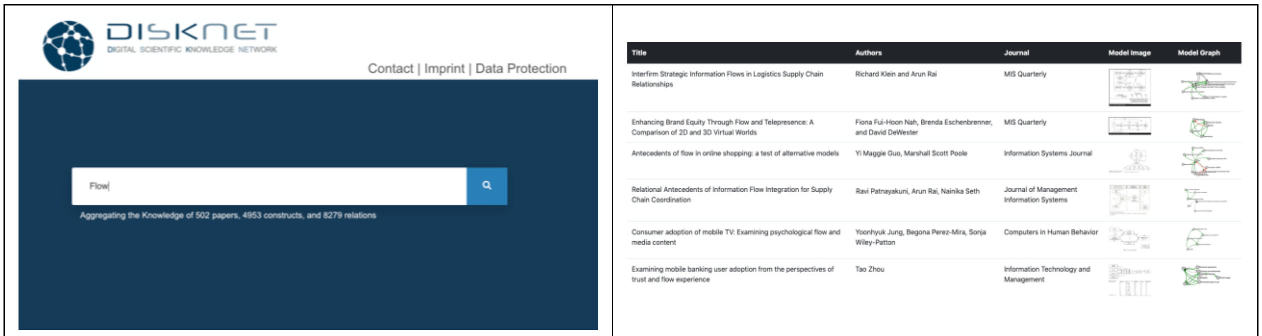

Figure 2: Construct search functionality provided by DISKNET

Thus far, we have populated the platform with an initial set of publications from the last 20 years. At the time of publication, we have extracted more than 500 publications from the basket of eight (ISR, MISQ, JMIS, JAIS, JSIS, ISJ, JIT, EJIS), yielding about 4,000 constructs and 6,600 relations. This set of publications, constructs, and relations will be extended continuously.

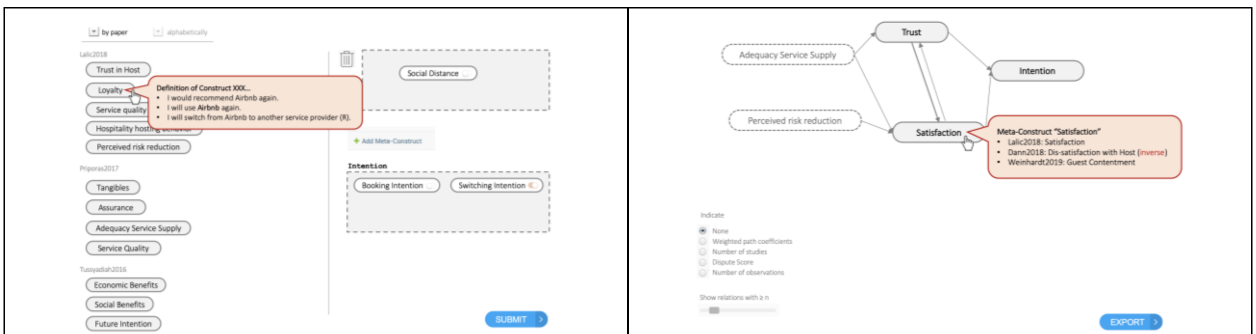

Figure 3: Exemplary exploration and aggregation functionality provided by DISKNET

Discussion and Future Directions

In this paper, we report on our ongoing efforts to build tool support for the effective and efficient systematic extraction, exploration, and aggregation of knowledge in IS research. More specifically, we aim to develop a foundational infrastructure that will allow the IS discipline to advance and make existing knowledge (as documented in its theories, constructs, and relations) accessible. Through this effort, we hope to provide capabilities that will make due diligence processes, a necessary prerequisite for high-quality research, less cumbersome (e.g. by providing a concise list of constructs used in IS research along with their semantic and analytical relations). This will not only allow for a comprehensive description of antecedents and consequences of investigated constructs but also the different operationalizations of the latter.

In terms the jingle and jangle fallacies discussed earlier, we are confident that effective use of our tool will help future researchers to become aware of and overcome any inherent construct identity problems. Using our tool would alert researchers to differences in the definitions and would allow them to see that the constructs differ in their semantic and analytical relations. Recognizing and overcoming the jangle fallacy will rely strongly on the analytical relations captured in our tool. While Larsen and Bong (2016) had to engage in a careful extraction and analysis of the respective measurement items, researchers using our tool would be easily alerted to the differences in the measurement items employed.

In the long run, the means our project provides will make it easier for researchers in IS to scrutinize and synthesize the existing theoretical superstructure underlying the various IS subdomains. In turn, this will safeguard the quality and improve the contribution of findings in research projects. We particularly envision that our database and tool will be used to explore boundary spanning constructs, perform meta-reviews of pertinent IS research, and improve construct identity in IS research projects.

While our work has great potential to support IS scholars and the progress of our discipline, our research and development are still in progress and will extend into further design cycles beyond the three we discuss here. Parallel to the development of the design principles and eventual release of DISKNET for ICIS 2019 (concluding *Cycle 1*), we currently work intensively on expanding the underlying dataset. At the time of submission, we have processed over 500 SEM-based publications from the eight leading IS journals; a sample we will be extended by adding further journals and conferences. Similarly, we are currently investigating analytical features to help with the exploration and manipulation of the data captured as well as refined data visualizations to facilitate working with the results of our artifact.

Beyond introducing our work, ICIS 2019 also provides a platform to elicit the necessary community support for the steps of *Cycle 2* (apply platform and improve). While capturing and classifying SEM-based studies from the basket of eight journals provided us with an accessible and convenient starting point, long-term efforts must be more comprehensive. This applies to both the expansion of the SEM-based sample beyond the basket, as well as to non-SEM-based quantitative work and qualitative work. We anticipate that respective requirements to be incorporated in future design cycles will lead to differences in the system's interface and data models. Moreover, we foresee that the very practice of capturing and working with the data will change as well. Nonetheless, we are confident that the three core aspects of construct clarity we focus on (i.e., construct definitions and scope conditions, semantic relations, and analytical relations) apply to these types of work just as much as they apply to our SEM-based starting set.

In *Cycle 3* and onwards, we also expect that aspects regarding the use and further development of the platform will become increasingly important. Based on feedback received thus far, we believe that this discussion will particularly involve the design of a governance scheme that supports our tool's disciplinary adoption and ensures ongoing and sustainable use. This step must also include experiences gathered with past approaches that have been discontinued since (Newsted et al. 1998). One exemplary aspect in this is how to ensure contributions (e.g. original data, review of existing data). Possible strands of thinking can include incorporating our tool into the submission process at journals and conferences, such that authors are requested to enter their papers once accepted for publication. Whether such policies have the desired effect and are met with acceptance will have to be investigated in future design cycles. The same applies to rules and responsibilities should discrepancies and contradictions materialize in our tool's data. We propose that steering committees could be one viable option, as would be an automated "open issues list" created by the system's machine-learning-based features highlighting where additional research is needed.

Complementing these efforts, we are developing automated techniques for identification, extraction, and assessment of relevant papers. We have started to leverage machine learning techniques to support the discovery of relevant papers and automatically extract constructs and their relations (Dann et al. 2017). Next, extending Larsen and Bong (2016), we employ state-of-the-art unsupervised learning techniques to tackle the jingle/jangle problem (Ludwig et al. 2020). In sum, the fallacies discussed in this paper, along with the increasing publication volume, underline the need for taking stock and synthesizing knowledge in IS research. For us, the provisioning of tool support for systematic knowledge accumulation is a central step to counteract the issue of a fragmented theoretical superstructure in IS research. We are convinced that our ongoing efforts to build tool support contribute to research rigor and a more comprehensive and consistent body of knowledge within the IS discipline. In this vein, we see our work responding to Keen's (1980) call – shared at the very first ICIS – to build on each other's work in order to help the field build a comprehensive whole rather than just individual fragments.

Acknowledgements

The authors gratefully acknowledge financial support from the ForDigital Research Alliance.

References

Abrami, P. C., Cohen, P. A., and D'Apollonia, S. 1988. "Implementation problems in meta-analysis," *Review of Educational Research*, (58:2), pp. 151–179.

Auer, S., Kovtun, V., Prinz, M., Kasprzik, A., Stocker, M., and Vidal, M. E. 2018. "Towards a Knowledge Graph for Science," in *WIMS 2018 Proceedings*, pp. 1–6.

Bearden, W. O., and Netemeyer, R. G. 2011. *Handbook of Marketing Scales*, Sage Publications.

Boell, S. K., and Cecez-Kecmanovic, D. 2015. "On being 'systematic' in literature reviews," *Journal of Information Technology*, (30:2), pp. 161–173.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. 2009. *Introduction to Meta-Analysis*, John Wiley & Sons Ltd, London.

Bornmann, L., and Mutz, R. 2015. "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the American Society for Information Science*, (66:11), pp. 2215–2222.

Cheung, M. W.-L. 2015. *Meta-Analysis: A strcutural equation modeling approach*, Wiley, London, UK.

Dann, D., Hauser, M., and Hanke, J. 2017. "Reconstructing the giant: Automating the categorization of scientific articles with deep learning techniques," in *WI 2017 Proceedings*, pp. 1538–1549.

Dwivedi, Y. K., Wade, M., and Schneberger, S. L. 2011. *Information Systems Theory: Explaining and Predicting Our Digital Society*, New York, NY, USA: Springer.

Gregor, S. 2006. "The nature of theory in Information Systems," *MIS Quarterly*, (30:3), pp. 611–642.

Gregor, S., and Jones, D. 2007. "The anatomy of a Design Theory," *Journal of the Association for Information Systems*, (8:5), pp. 312–335.

Hair, J. F., Sarstedt, M., Ringle, C. M., and Mena, J. A. 2012. "An assessment of the use of partial least squares structural equation modeling in marketing research," *Journal of the Academy of Marketing Science*, (40:3), pp. 414–433.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly*, (28:1), pp. 75–105.

Hovorka, D. S., Larsen, K. R., Birt, J., and Finnie, G. 2013. "A meta-theoretic approach to theory integration in information systems," in *HICSS 2013 Proceedings*, pp. 4656–4665.

Keen, P. G. W. 1980. "MIS Research: Reference disciplines and a cumulative tradition," in *ICIS 1980 Proceedings*, pp. 9–18.

Kelley, T. L. 1927. *Interpretation of Educational Measurements*, World Book Company; Oxford, UK.

Kuechler, B., and Vaishnavi, V. 2008. "On theory development in design science research: Anatomy of a research project," *European Journal of Information Systems*, (17:5), pp. 489–504.

Larsen, K. R. 2013. "Inter-Nomological Network," (available at inn.theorizeit.org; retrieved April 17, 2019).

Larsen, K. R., and Bong, C. H. 2016. "A tool for addressing construct identity in literature reviews and meta-analyses," *MIS Quarterly*, (40:3), pp. 529–551.

Larsen, K. R., and Eargle, D. 2015. "Theories used in IS Research Wiki," (available at https://is.theorizeit.org/wiki/Main_Page; retrieved April 9, 2019).

Ludwig, S., Funk, B., and Mueller, B. 2020. "Using natural language processing techniques to tackle the construct identity problem in Information Systems Research," in *HICSS 2020 Proceedings (in press)*.

Mueller, B., and Urbach, N. 2017. "Understanding the why, what, and how of theories in IS research," *Communications of the AIS*, (41:17), pp. 349–388.

Newsted, P. R., Huff, S. L., and Munro, M. C. 1998. "Survey instruments in information systems," *MIS Quarterly*, (22:4), pp. 553–554.

Rai, A. 2018. "Editor's Comment: The first few pages," *MIS Quarterly*, (42:2), pp. iii–vi.

Suddaby, R. 2010. "Editor's comments: Construct clarity in theories of management and organization," *Academy of Management Review*, (35:3), pp. 346–357.

Tate, M., Furtmueller, E., Evermann, J., and Bandara, W. 2015. "Introduction to the special issue: The literature review in information systems," *Communications of the AIS*, (37:5), pp. 103–111.

Thorndike, E. L. 1904. *An Introduction to the Theory of Mental and Social Measurements*, Science Press; New York, NY, USA.

Urbach, N., and Ahlemann, F. 2010. "Structural equation modeling in information systems research using partial least squares," *Journal of Information Technology Theory and Application*, (11:2), pp. 5–40.

de Waard, A., and Kircz, J. G. 2003. "Metadata in science publishing," in *CI 2003 Proceedings*, pp. 73–84.