

# Decision Support Systems for Financial Market Surveillance

Dissertation

zur Erlangung des wirtschaftswissenschaftlichen Doktorgrades der Wirtschaftswissenschaftlichen Fakultät der Georg-August-Universität Göttingen

vorgelegt von  
Dipl. Inf. Irina Alić  
aus Banja Luka

Göttingen, 2016

First Examiner:  
Second Examiner:  
Third Examiner:

Prof. Dr. Jan Muntermann  
Prof. Dr. Matthias Schumann  
Prof. Dr. Lutz M. Kolbe

*To my parents*

## Preface

This Thesis was written at the Chair of Electronic Finance and Digital Markets at Georg-August University of Göttingen.

The present research received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) within the context of the Project FIRST, large scale information extraction and integration infrastructure for supporting financial decision making, under grant agreement no. 257928.

I would like to thank to my supervisor Prof. Dr. Jan Muntermann, who supported me and provided many helpful suggestions. I am also thankful for the opportunity to be a part the European Project FIRST. Furthermore, I thank all of the members of the FIRST project consortium for their contributions to the IT Artifacts developments described in this thesis.

Furthermore, I would like to express heart-felt thanks to my friends and colleagues Michael Siering, Mirko Jan Zülch, Griselda Sinanaj, Timo Weinrich, Matthias Eickhoff, Runhild Wieneke and Tobias Langkau for going through all the phases of our exciting scientific research with me.

## Abstract

Financial decision support systems have become an important research topic and are also of great interest to practitioners. On one hand, the financial regulatory authorities are faced with tracking an increasing amount of online information published on financial blogs and news in order to maintain financial market surveillance. On the other hand, two rapidly emerging trends, the increasing amount of available data and the evolution of data mining methods, pose challenges for researchers. Financial decision support systems can aid decision making by providing the authorities with relevant information in a timely fashion. This thesis presents IT artifacts that can support financial market surveillance decision making. Furthermore, an explanatory design theory is provided that addresses the requirements posed by the regulatory authorities and compliance officers in financial institutions.

## Zusammenfassung

Entscheidungsunterstützungssysteme in der Finanzwirtschaft sind nicht nur für die Wissenschaft, sondern auch für die Praxis von großem Interesse. Um die Finanzmarktüberwachung zu gewährleisten, sehen sich die Finanzaufsichtsbehörden auf der einen Seite, mit der steigenden Anzahl von onlineverfügbaren Informationen, wie z.B. den Finanz-Blogs und -Nachrichten konfrontiert. Auf der anderen Seite stellen schnell aufkommende Trends, wie z.B. die stetig wachsende Menge an online verfügbaren Daten sowie die Entwicklung von Data-Mining-Methoden, Herausforderungen für die Wissenschaft dar. Entscheidungsunterstützungssysteme in der Finanzwirtschaft bieten die Möglichkeit rechtzeitig relevante Informationen für Finanzaufsichtsbehörden und Compliance-Beauftragte von Finanzinstituten zur Verfügung zu stellen. In dieser Arbeit werden IT-Artefakte vorgestellt, welche die Entscheidungsfindung der Finanzmarktüberwachung unterstützen. Darüber hinaus wird eine erklärende Designtheorie vorgestellt, welche die Anforderungen der Regulierungsbehörden und der Compliance-Beauftragten in Finanzinstituten aufgreift.

# Table of Contents

Preface .....	II
Abstract.....	III
Table of Contents .....	IV
List of Tables .....	VII
List of Figures.....	VIII
SECTION A: Introduction and Foundations .....	1
1 Motivation and Research Questions.....	2
2 Structure of the Thesis .....	5
3 Research Background.....	9
3.1 Design Science Research .....	9
3.1.1 IT Artifact .....	11
3.1.2 Design Theories for Information Systems .....	11
3.2 Decision Support Systems .....	12
3.3 Financial Market Manipulation.....	13
3.4 Information Systems for Financial Market Surveillance .....	15
4 Study Setup .....	17
SECTION B: Studies.....	21
5 Paper 1: State of the Art in financial DSS and Problem Statement .....	23
5.1 Introduction.....	24
5.2 Methodology .....	25
5.2.1 Analytical Framework .....	25
5.2.2 Identification Process.....	28
5.2.3 Selection Process .....	28
5.2.4 Classification and Analysis Process .....	28
5.3 Results of Literature Review .....	29
5.3.1 Problem Domain and related Requirements .....	33
5.3.2 Components .....	34
5.3.3 Evaluation Methods .....	36
5.4 Implications for Future Research and Conclusion.....	36

6	Paper 2: A Qualitative Multi-Attribute Model for Financial Market Surveillance	39
6.1	Introduction	40
6.2	Related Work	41
6.3	Methodology	42
6.3.1	Design Science Research Approach	42
6.3.2	Qualitative Multi-Attribute Modelling Methodology	43
6.3.3	Proposed Research Design	44
6.4	Problem Definition	45
6.5	Decision Model Suggestion	46
6.6	Qualitative Multi-Attribute Model Development	46
6.6.1	Attribute Identification	46
6.6.2	Attribute Structure	48
6.6.3	Attribute Scales	49
6.6.4	Aggregation Rules	51
6.7	Experimental Evaluation of the Qualitative Multi-Attribute Model	52
6.8	Conclusion	53
7	Paper 3: IT Artifact Instantiation and Evaluation	55
7.1	Introduction	56
7.2	Background	57
7.3	Instantiated IT Artifact	59
7.3.1	Data	59
7.3.2	Model Description	60
7.3.3	User Interface	65
7.4	IT Artifact Evaluation	67
7.4.1	Descriptive Analysis	67
7.4.2	Manipulation Examination	68
7.5	Conclusion	70
8	Paper 4: Explanatory Design Theory for Financial Market Surveillance Decision Support	73
8.1	Introduction	74
8.2	Research Background	75
8.2.1	Market Manipulation	75
8.2.2	Design Theories for Information Systems	75
8.3	Study Design	76
8.3.1	Action Design Research	76
8.3.2	Research Stages	76
8.4	An Explanatory Design Theory for Market Surveillance Decision Support	80
8.4.1	General Requirements	80
8.4.2	General Components	82
8.5	Conclusion	85

---

9	Executive Summary of Papers .....	87
	Paper 1: State of the Art in financial DSS and Problem Statement .....	87
	Paper 2: A Qualitative-Multi-Attribute Model for Financial Market Surveillance.....	88
	Paper 3: IT Artifact Instantiation and Evaluation .....	89
	Paper 4: Explanatory Design Theory for Financial Market Surveillance Decision Support.....	90
	SECTION C: Integration of Findings.....	93
10	Contribution and Conclusion .....	94
10.1	Contribution to Theory .....	94
10.2	Contribution to Practice .....	95
10.3	Conclusion .....	95
11	Discussion .....	97
11.1	Limitations .....	97
11.2	Recommendations for Further Research Work.....	97
11.3	Implications for Policy.....	98
	References .....	100
	Overview of author contribution in included publications.....	110



## List of Tables

Table 1: Fact sheet of publication nb. 1.....	23
Table 2: Coding scheme, together with examples from our analysis.....	29
Table 3: Classification of articles .....	30
Table 4: Classification of articles by data .....	34
Table 5: Classification of articles by methods .....	35
Table 6: Fact sheet of publication nb. 2.....	39
Table 7: Interview structure for problem definition .....	46
Table 8: Fact sheet of publication nb. 3.....	55
Table 9: Calculation of average values of the input variables.....	63
Table 10: Examination of the highest and lowest daily price changes.....	68
Table 11: Evaluation results based on market price value changes for the lowest and highest daily prices.....	69
Table 12: Fact sheet of publication nb. 4.....	73
Table 13: ADR Stages based on Sein et al., (2011) .....	79
Table 14: Design theory for financial market surveillance DSS .....	85
Table 15: Paper contributions.....	91
Table 16: Author contribution .....	110

## List of Figures

Figure 1: Mapping of central research questions to individual papers.....	5
Figure 2: Structure of the Thesis .....	7
Figure 3: Adapted DSR Framework based on (Vaishnavi & Kuechler, 2008) and (Hevner et al., 2004).....	18
Figure 4: Awareness of the problem cycle .....	18
Figure 5: Cycles of Suggestion, Development and Evaluation .....	20
Figure 6: Design theory according to Baskerville and Pries-Heje (2010).....	25
Figure 7: An expanded DSS framework based on Power (2004) .....	26
Figure 8: The key components of information mining based on Gopal et al. (2011) ....	26
Figure 9: Analytical framework for literature review .....	27
Figure 10: Research approach based on (Vaishnavi & Kuechler, 2008) and (Bohanec, 2003) .....	44
Figure 11: Structured attributes for assessment of P&D cases.....	48
Figure 12: Model of attribute structure.....	49
Figure 13: Decision rules for P&D attribute .....	52
Figure 14: Evaluation Examples.....	53
Figure 15: The hierarchical tree of attributes (Alić et al., 2013) .....	61
Figure 16: The attribute scales (Alić et al., 2013) .....	61
Figure 17: Calculation of artificial jumps as decision rules calibration .....	64
Figure 18: P&D alert output aggregation .....	64
Figure 19: FMS-DSS prototype.....	65

# SECTION A: Introduction and Foundations

# 1 Motivation and Research Questions

Over the last two decades, the information available online regarding financial instruments has attracted much attention from potential investors, compliance officers and regulatory authorities. Exhaustive online information sources, such as financial blogs and tweets, seduce nonprofessional investors into investing in potentially suspicious financial instruments. Most of such investors struggle with their involvement in faulty investments (SEC, 2012). The potential impact of such information available online on the decision making of investors makes decision support systems for the financial industry critical.

A substantial amount of research has been conducted on decision support systems to identify market abuse (Braeuer, Heddier, Knackstedt, & Platz, 2014; Lee, Qiu & Whinston, 2014). Thereby the Information Systems (IS) for financial market surveillance include the detection of notable market abuse patterns in structured data (Eren & Ozsoylev, 2006). However, research that integrates both structured and unstructured user-generated content data with the information provided by the regulatory authority in a system that supports financial institutions in their surveillance tasks is critical.

With the growing number of market manipulation cases experienced in recent years (Aggarwal & Wu, 2006), financial market surveillance has attracted increased attention from the regulatory authorities (SEC, 2012) as well as academia and the industry. Generally, markets with few regulations, such as over-the-counter (OTC) markets, are not considered a safe investment option. Since such markets lack proper monitoring and controls, a risk of fraud prevails. Traditionally, market manipulations determine market regulations and efficiency. The SEC (U.S. Securities and Exchange Commission) was created primarily to curtail such manipulative activities in financial markets (Zaki, 2013). The SEC advises or warns investors about market vulnerability and frauds through its online portal sanctioning the fraudsters by trading suspensions (SEC, 2012). Thus, markets with low supervisory requirements and low priced financial instruments are the main target of manipulators, posing a challenge for both investors and regulators.

Although manipulations appear to be declining in regulated markets, non-regulated markets such as OTC products are still affected (Aggarwal & Wu, 2003; Zaki, 2013).

The regulated markets have adopted an expensive transparent surveillance mechanism to ensure market protection (Goldberg, Kirkland, Lee, Shyr, & Thakker, 2003). This system focuses on structured data given as time series. Unstructured data, such as social net user-generated information, remain un-assessed. Since there is very little research focused on OTC markets, (Zaki, 2013), this thesis sought to construct a decision tree for detecting market manipulation based on user-generated content and data provided by regulatory authority, and develop a related design theory. The developed decision tree as a qualitative multi-attribute model is actually an early prototype as an IT artifact of a model-driven decision support system for market monitoring that helps to automatically process, analyze and interpret relevant data from online resources and triggers an alert when manipulation is detected. Generally, model-driven decision support systems emphasize access to and manipulation of a model in order to solve special problems by providing decision suggestions to the decision makers. This thesis presents a model-driven decision support system since the constituting elements of developed IT artifacts are based on qualitative multi-attribute models.

Easily accessible online information sources, such as financial blogs and tweets, allow scammers to promote stocks they already hold, spreading false positive information in an enthusiastic and persuasive manner (Dunham, 2007). Such information-based manipulative recommendations also known as pump and dump manipulation, seduce non-professional investors into spending their capital on potentially suspicious financial instruments. The problem of market manipulations is very interesting because it requires the scientist to find a solution for fraud detection. The practical relevance is due to the expectations that market participants act in a fair and secure market.

The amount of digital and online data available structured as time-series as well as unstructured data, such as tweets, news or blogs, doubles every two years; only two percent of the data are not digital. Furthermore, the amount of digital data is expected to grow to 10 Zettabytes by 2020 (Mayer-Schönberger & Cukier, 2013). The proper involvement of hybrid/big data as well as the knowledge management techniques, decision rules, algorithms and other data-mining techniques required to analyze such real-time big data available online are becoming increasingly important in academia and practice (Abbasi, Sarker, & Chiang, 2016). Thus, this dissertation was intended to discover ways to establish financial market surveillance decision support so as to provide a

design theory that can serve as a guideline for the development of IS to detect information-based market abuse on large real time online data.

The motivation for this dissertation arose from the increasing need to protect financial markets from abuse. This dissertation scrutinized the decision support system for information-based market manipulation in non-regulated markets with the design science research lens. The main objectives that guided this research are as follows:

- I. To determine the requirements and components of information systems for information-based market manipulations.
- II. To develop an IT Artifact for the detection of financial market manipulation following the design science research paradigm.
- III. To present an explanatory design theory for information systems for the detection of information-based market manipulation.
- IV. To evaluate the effectiveness of the developed IT Artifact for fraud detection using a set of real market data to explain the system's performance.
- V. To pinpoint shortcomings of previous research on financial markets since to our knowledge, no one has ever suggested employing different data sources (structured and unstructured data as well as the information suggested by the regulatory authority) to identify fraud in near real-time.

To reveal the relevant key indicators in a market surveillance context required to tackle the problem of information-based market manipulation, the thesis employed two central research questions and four specific sub-questions. The first two central research questions are formulated as follows:

**RQ1:** How can a model for the identification of information-based manipulation be developed or structured?

**RQ2:** What are design foundations for related IT architectures?

Going along with the two central questions, the dissertation theoretically supported and empirically explored the role of information-based manipulation factors in supporting compliance officers in decision-making. The specific sub research questions addressed in the different papers are presented in Section B.

## 2 Structure of the Thesis

This cumulative dissertation consists of four papers that together address the central research questions including the underlining sub-research questions depicted in Figure 1.

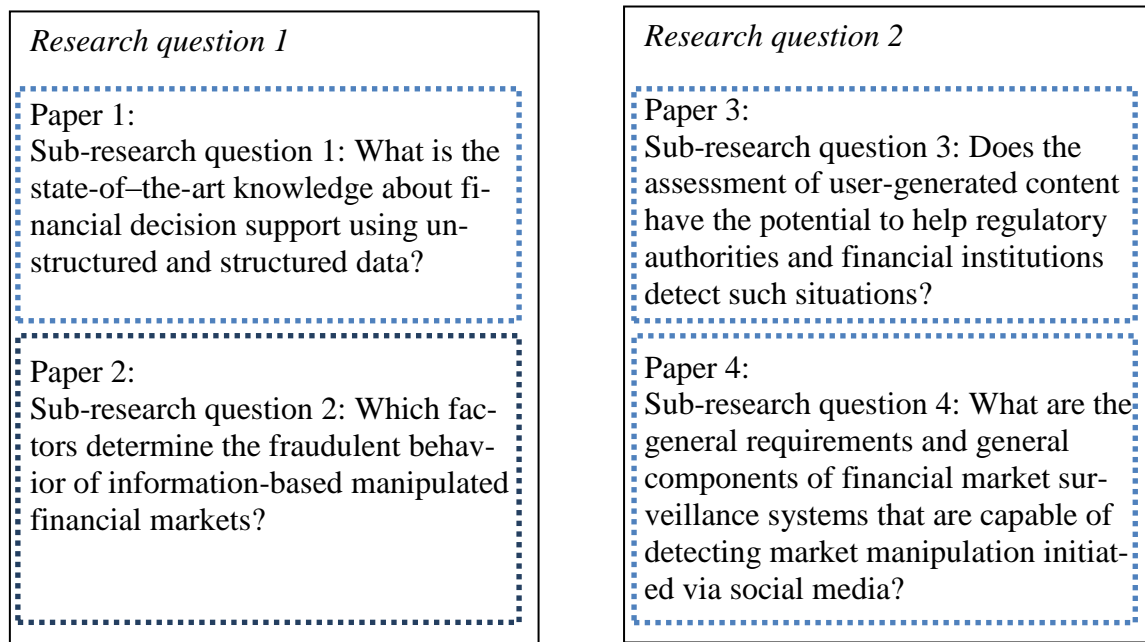


Figure 1: Mapping of central research questions to individual papers

Paper 1 aims to review the relevant existing research. To structure the literature review, a framework that integrates elements of Decision Support Systems (DSS), design theory and information mining is proposed. Our analysis reveals that more research is needed in the domain of fraud detection. We found very few IS studies despite the importance of these topics in the increasingly turbulent worldwide financial markets. Thus, Paper 1, the first empirical article in this dissertation, analyzes the current state-of-the-art financial DSS by developing and applying an analytical framework that may also help future researchers in the financial arena structure their investigations.

The next empirical part of this dissertation includes qualitative research results. Based on interviews with problem owners, paper 2 extends the argument established in paper 1. Thus, it deals with knowledge transfer and the development of key attributes for the

detection of information-based manipulation and addresses the first research question of this dissertation. The relevant studies to date in the literature primarily investigate the requirement aspect of financial DSS without the explicit involvement of the problem owners/target user. It appears that the target user is primarily mentioned in the present studies, whereas the requirement and problem statements are derived from the literature. This paper demonstrates how the requirements are generated, and the early prototype is developed. Based on interviews with domain experts, we first identified the factors used to detect suspicious situations and then applied a qualitative multi-attribute modelling technique. Thereby, our resulting model builds upon the valuable knowledge of domain experts and provides a means to address the challenge of information-based market manipulation.

The subsequent evaluation paper 3 builds upon the research results of paper 2. This evaluation paper draws on large-scale data to address the evaluation of an IT Artifact instantiation in the form of a model-based Decision Support System to support decision making in the field of market surveillance. With a data architecture developed in the context of a collaborative EU-funded research project, an evaluation of the implemented system is presented that is based on voluminous and heterogeneous data, including social media data. Thus, this artifact utilizes a qualitative decision model developed previously and presented in paper 2 to identify situations in which the prices of single stocks are affected by fraudsters who aggressively advertise the stock. The empirical results indicate that the developed artifact can help identify such situations and provide practical decision support in market surveillance.

Paper 4 brings the previous results together. It builds upon the research results from papers 1-3 and provides a formal representation of the developed design theory by addressing the second research question of this thesis. Its main objective is to explain how to design efficient market surveillance systems by providing insights into the needs of compliance officers. It analyzes the general components and requirements by creating an innovative surveillance system. Thereby, it provides advice on how to develop corresponding systems. The research utilizes the appropriate IS theory based on (Baskerville & Pries-Heje, 2010) to inform the research.

The four papers appear in chronological order as they were published in conference proceedings. The papers are listed below:



1. Alic, Irina; Muntermann, Jan; Gregory Robert W. 2012. State of the Art of Financial Decision Support Systems based on Problem, Requirement, Component and Evaluation Categories. In: Proceedings of the 25th Bled eConference eDependability: Reliable and Trustworthy eStructures, eProcesses, eOperations and eServices for the Future; Bled, Slovenia.
2. Alic, Irina; Siering, Michael; Bohanec, Marko. 2013. Hot Stock or Not? A Qualitative Multi-Attribute Model to Detect Financial Market Manipulation. In: Proceedings of the 26th Bled eConference eInnovation: Impacts of Individuals, Organizations and Society; Winner of the best paper award; Bled, Slovenia.
3. Alic, Irina. 2015. Supporting Financial Market Surveillance: An IT Artifact Evaluation. In: Proceedings of the 28th Bled eConference, Bled, Slovenia.
4. Alic, Irina. 2015. Financial Market Surveillance Decision Support: An Explanatory Design Theory. In: Proceedings of the 28th Bled eConference; Bled, Slovenia.

Regarding the guiding structure, this dissertation is divided into three parts: *Section A*, *Section B* and *Section C* Figure 2.

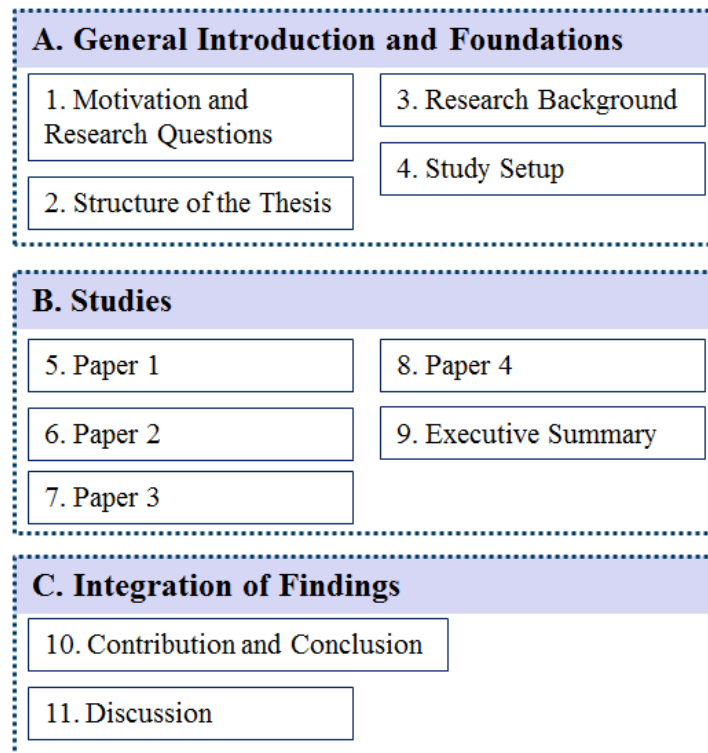


Figure 2: Structure of the Thesis

Section A proceeds as follows:

- The section highlights background research in market manipulation and discusses previous studies on design science research, decision support systems, financial market manipulation and design theories for information systems.
- The section describes the research methodology employed in the study.

Section B proceeds as follows:

- The section provides the results from the four papers included in this dissertation. The objectives, research analyses, results, major methodologies and conclusions offered in the papers are presented.

Section C proceeds as follows:

- The section provides the study conclusions.
- The section provides the major limitations and recommendations for further research.

## 3 Research Background

In this subsection, the background research on design science research, IT artifacts, decision support systems and design theories for information systems financial market manipulation is provided.

### 3.1 Design Science Research

There are two paradigms characterizing extensive research in information system fields: design science and behavioral science (Arnott & Pervan, 2012, Chatterjee, 2015). The behavioral science approach involves developing and verifying theories that predict organizational or human behavior (Winter, 2008), whereas design science research attempts to expand the boundaries of organizational and human capabilities through the creation of new artifacts (Baskerville & Pries-Heje, 2010) as well as the development of new design theory (Markus, Majchrzak & Grasser, 2002; Kuechler & Vaishnavi, 2008). This dissertation applies the Design Science Research lens to derive innovative IT artifacts and to develop a new design theory that can be used in academia as well as in practice.

The Design Science Research (DSR) paradigm has been applied in Information Systems for more than four decades, beginning with Morton's (1971) Management Decision Systems study publication. Since then several studies have been published. The Peffers, Tuunanen, Rothenberger, and Chatterjee (2007) study provides a comprehensive discussion of the design science research approach application in information systems. Due to the absence of methodology serving as an acceptable model for design science research, the authors sought to present a design science approach combining practices, procedures and principles for conducting research in information systems. The model developed intended to meet three major goals: consistency with previous literature studies, provision of a nominal framework for conducting design science research and provision of mental frameworks for design science research presentation and evaluation in information science. In the study, the authors described six main steps for design science research. The steps were identification and motivation of the problem, objective definition, design and development, communication, evaluation and demonstration (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

Iivari (2015) differentiates and contrasts two major strategies in design science research application in information systems. The first approach involves a researcher constructing an information technology meta-artifact based on the general solution concept while addressing categorized problems. The second approach entails a researcher constructing a concrete information technology artifact to solve a specific client problem in the same context. The researcher also uses prescriptive knowledge of the general solution concept while addressing categorical problems. The author discusses these two approaches using sixteen measures that represent process, outcomes, resource requirements and context (Iivari, 2015). Particularly relevant for this thesis is the design science research that incorporates the general design cycle as proposed by (Vaishnavi & Kuechler, 2008). The general design cycle contains Awareness of Problem, Suggestion, Development, Evaluation and Conclusion. The cycles utilized iteratively will guide this thesis in so far as it provides the steps required to conduct design science research.

Although design research has been extensively used in information systems, it is still in a state of conceptual confusion regarding the type of contribution (Baiyere, Hevner, Gregor, & Rossi, 2015). The author argues that in design science research, there are two fractions. One fraction justifies the design of the artifact that provides particular and innovative solutions to problems, whereas another supports the theoretical formulation of the design (Gregor & Hevner, 2013).

On one hand, the problem solution fraction research indicates that design science research deals with the successful construction of artifacts (Peppers, et al., 2007; Iivari, 2015). On the other hand (Markus, Majcharzak & Grasser, 2002; Kuechler & Vaishnavi, 2008), the design science research approach in information system discipline aims to design useful and innovative artifacts including contribution to the theory. This dissertation encompasses both aspects of design. Thus the dissertation seeks to bridge this gap and provide an innovative practical solution as well as a design theory.

In summary, the design science approach aims to create innovative IS products and theories for the effective achievement of theory development. In this context, the design science approach is the best starting point for constructing a theory on information systems. Hence, this dissertation focuses on the development of a practically relevant artifact by providing a theoretical value.

### 3.1.1 IT Artifact

The goal of Design Science Research is to make contributions to the knowledge base through developed IT Artifacts (Hevner, March, Park, & Ram, 2004). According to Hevner and Chatterjee (2010) the following IT artifact types exist: constructs, models, methods and instantiations, as well as the design theories. *Constructs* provide a language for the definition and communication of the problem and its solution. One construct applied in this thesis is sentiment. The task of sentiment analysis is to automatically extract and recognize the opinion in text (Bloom 2011). Generally, sentiments represent opinions based on emotions. In a financial context, the market sentiment represents either a positive, negative or neutral opinion of the market actors regarding the financial instruments (Zülch, Alic & Muntermann, 2011). In this thesis, sentiment is one important component of the developed artifact. *Models* represent the relationships between the constructs. In this thesis, a model in the form of a hierarchical tree decomposed into various components will be presented. *Methods* represent procedures used to perform specific tasks. *Instantiations* are based on constructs, models, and methods expressing the implementation in working systems. Design Theories in DSR draw upon guidance from practitioners (Markus et al. 2002; Walls, Widmeyer, & El Sawy, 1992) to solve problems (instantiation) for given user requirements (models). Research (Goldkuhl, 2004) suggests that design should be considered as product and process. From the product point of view, a given design theory is composed of some elements that provide descriptions of goals applicable to the theory. From the process perspective, the design method has components that describe the procedures for constructing artifacts. The kernel theories guide the whole design process, while design process hypotheses determine the consistency between the results from the design methods and the artifacts designed to meet user and business requirements (de Moor, 2007).

### 3.1.2 Design Theories for Information Systems

Several studies on theory-building approaches in Design Science Research have been published in recent years (Peffer et al. 2007; Gregor & Hevener, 2013). Hevner, March, Park and Ram (2004) describe design science research performance in information systems using a clear conceptual model and guidelines that underline the comprehension, implementation and assessment of research. The authors discovered that in design science research, awareness and comprehension of the problem domain as well as the dis-

covery of solutions are accomplished by building and applying the artifact designed. They also proposed seven guidelines to assist design science researchers in both contributing to IS theory and creating and evaluating as-of-yet unknown and innovative information technology (IT) artifacts. The guidelines include problem relevance, design as artifact, design evaluation, research rigor, research contributions, communication of research and the search process (Hevner, March, Park, & Ram, 2004).

De Moor (2007) defines a design theory as a prescriptive theory using theoretical foundations to illustrate the process of conducting design processes feasibly and effectively. The theory helps in describing, optimizing and analyzing design methods (Goldkuhl, 2004). The framework provided by Hevner, March, Park and Ram (2004) is helpful in the evaluation and presentation of information system research during theory development. Hevner's model illustrates the correlation underlying design research processes and behavior. The knowledge base, composed of methodologies and conceptual underpinnings, as well as the environment, composed of organizations, technologies and people, inform the design research processes. The analytical model helps scholars design new representations for information system solution process, problems and solutions (de Moor, 2007).

Particularly relevant to this dissertation is the recently developed information system design theory proposed by Baskerville and Pries-Heje (2010), which is an explanatory model of design IT artifacts. The theory distinguishes between general components and general requirements where the components are justified by the requirements. The explanatory design theory explains why a set of requirements is satisfied by a set of components. Hence, only two essential parts are needed for a complete explanatory design theory: general requirements and general solution components (Baskerville & Pries-Heje, 2010).

### 3.2 Decision Support Systems

Early definitions of Decision Support Systems refer to systems that assist managerial decision makers in decision situations (Morton, 1971; Ginzberg & Stohr, 1982; Sprague, 1980). These systems extend the decision maker's capabilities without replacing his/her assumption. DSS are broadly explained as interactive systems involving de-

cision makers and computers to solve unstructured or semi-structured problems, the type of problems that cannot be specified as an algorithm to be computed.

DSS can be first categorized according to their dominant component driver, resulting in model-driven, data-driven, document-driven, communication-driven and knowledge-driven DSS (Power, 2004).

Model-driven DSS maintain for instance financial models providing accounting functionalities to the user, whereas data-driven DSS maintain time-series data. Document-driven DSS maintain different types of documents in order to support the assessment and analysis of these documents, whereas communication-driven DSS support user communication and collaboration. Finally, knowledge-driven DSS incorporate special problem solving capabilities in the form of artificial intelligence by providing decision suggestions to the decision makers. This thesis focuses on model-driven DSS since the constituting elements of developed IT artifacts are based on qualitative multi-attribute models (paper 2).

Design science research on DSS plays a crucial role in improvements regarding DSS relevance and quality since it can engage the profession and industry in crucial projects. The latest evaluations of DSS research have demonstrated gaps, demonstrating the need to improve DSS relevance and quality (Arnott & Pervan, 2012). The study by Arnott and Pervan (2012) evaluated DSS design science research papers, particularly information systems. The results highlight vital DSS research concerns such as relevance, evaluation, theorizing, research design and strategic focus (Arnott & Pervan, 2012).

### 3.3 Financial Market Manipulation

Manipulation generally refers to the intention to influence someone's opinion (Barnhill 2014), causing this individual succumb. Accordingly, manipulators try to put other people in a certain mood (Sunstein, 2015). The process of market manipulation requires the following two main elements: the manipulator and a manipulative action (Aggarwal & Wu, 2006; Barnes, 2009). The manipulators can be divided into insiders who have the access to the internal information of the company and investors and traders/brokers who have no access to the internal information sources (Aggarwal & Wu, 2006).

Financial Market Manipulation refers to how the price of a financial instrument can be influenced. Within the financial field, there are diverse schemes of financial manipulation in the market. Allen and Gale (1992) provide several classes of market manipulation schemes, including trade-based, information-based and action-based. Trade-based manipulation refers to strategies for simply buying and selling without taking any actions to change the value of the company (Aggarwal & Wu, 2003). Action-based manipulation comprises actions undertaken by management in order to affect the value of their company (Putniņš, 2012). Lastly, information-based manipulation involves the utilization of falsified information. Here false and misleading information is published in order to manipulate prices.

According to the US regulatory authorities, one of the most widespread information-based fraud schemes is “pump and dump” market manipulation (Securities and Exchange Commission [SEC], 2012). To push share prices to an artificial level, the fraudsters manipulate share prices by first buying a specific stock and then spreading untrue positive information about the company over the internet. Finally, profit is made by selling the stock at an abnormally high price (Cumming, Zhan, & Aitken, 2012).

Market manipulation related to the illegal disclosure of untrue information by the sender via unstructured data has been explored by (van Bommel, 2003). The author used a framework with diffusion of private information in determining reasons for spreading stock hints. The study examined an investor with limited investment abilities who spread incorrect rumors to followers, who in turn moved their prices based on the rumors. Incorrect rumors resulted in positive occurrences of price overshoots, giving rumormonger advantage. The study revealed that rumors are very informative at equilibrium since the rumormongers gain increased profits compared with liquidity traders without any form of information (van Bommel, 2003).

Aggarwal and Wu (2006) present theoretical and empirical evidence concerning the manipulation of stock prices in the USA. The authors examined the effect of manipulator trading with other traders seeking information about stock value. Using data from actions by the SEC, the authors found that there are high rates of manipulation among informed parties including brokers, corporate insiders, market makers, large shareholders and underwriters. The results suggested that increased manipulation occurs in liquid stock and increases the volatility of stocks. The authors demonstrated that prices in-



crease when manipulators buy and decrease when manipulators sell (Aggarwal & Wu, 2006).

The study by Mei, Wu and Zhou (2004) provides an incidence that indicates that manipulators leverage behavioral biases by investors and price manipulation processes geared towards profit generation. The authors considered behavior-driven investors and manipulators capable of influencing price of assets and arbitrageurs. The study illustrates that because of limited arbitrage and behavioral biases by investors, manipulators are likely to profit from pump and dump trading methods via accumulating speculative assets, increasing asset prices and selling assets at higher prices. Some of the anomalies resulting from asset pricing as derived in the study include momentum, excess volatility and reversal (Mei, Wu & Zhou, 2004).

Due to transparency deficits and illiquidity of the securities in the non-regulated markets, investors often lack proper investment, struggling with their involvement in faulty investments. Hence, it is correct to state that for the identification of market abuse, the information systems for financial market surveillance must include the detection of notable market abuse patterns in structured and unstructured data.

### 3.4 Information Systems for Financial Market Surveillance

There are several scientific studies relating to Information Systems for market surveillance, including the research of Mangkorntong and Rabhi (2007) in which the architecture for automated trading patterns based on electronically available market data is introduced. In assessing the performance of the financial market surveillance system, the authors note that the tested system's main limitation is the lack of performance due to a missing API required to connect to other systems. More recent research (Diaz, Zaki, Theodoulidis, & Sampaio, 2011) presents an architecture for market monitoring where data mining techniques are utilized to detect abusive patterns in spam emails. In another study, Heping (2006) introduced a Multilevel Stochastic Dynamic Process (MSDP) framework for modelling time series for financial market analysis and surveillance, thereby focusing attention on signaling e.g., market crashes or trend accelerations. Other research (Huang, Liang, & Nguyen, 2009) tackles a visualization approach for fraud detection problems in financial markets. The system is based on pattern recognition by which an unusual pattern is matched to the similar pattern in the database. The financial

market systems for detecting abuse, such as the Securities Observation, News Analysis and Regulation Systems (SONAR) (Goldberg, Kirkland, Lee, Shyr, & Thakker, 2003), aim to monitor the stock market. The system applies data mining, text mining, statistical regression and rule-based detection to recognize both abuse patterns in the structured data and unusual trading following publication of the news.

In summary, to detect the various types of market manipulation, a surveillance system needs to handle traditional data (e.g., time series) as well as non-traditional data (e.g., news, blogs and twitter platforms).

## 4 Study Setup

To contribute to the knowledge base and answer the research questions, this dissertation thesis followed the DSR approach as proposed by (Hevner et al., 2004) and (Vaishnavi & Kuechler, 2008). The detailed specific methods addressed in different papers are presented in Section B.

Design science research encompasses the development and evaluation of constructs, frameworks, models, methods, instantiations or theories by which identified business needs or real problems can be addressed (Hevner et al., 2004). The set of artifacts presented in this thesis includes an analytical framework for literature review, a model that describes the relationships between components, an innovative instantiation of IT artifacts that solves a practical problem and a deductively developed theory that enhances the knowledge base. Together, they represent a new approach to understanding abusive behavior in the financial market, thus enabling regulatory authorities to counteract criminal activities more effectively. Thus, is this thesis a DSR framework to guide the development of the mentioned artifacts as proposed by (Hevner et al., 2004) is incorporated. Here the DRS cycles of building and evaluation closely interact with the knowledge base and the environment ensuring rigor and relevance of scientific work.

Additionally, in this thesis, the generalized process cycles as proposed by (Vaishnavi & Kuechler, 2008) are used. The authors suggest a series of iterative rounds to conduct design-oriented research projects. The recommended cycles are as follows:

- I. The initial “Awareness of the problem” aims to identify theoretical knowledge and practical user needs regarding the specific problem.
- II. The “Suggestion” aims to situate the requirements and components.
- III. The “Development” aims to develop artifacts.
- IV. The “Evaluation” explores the functionalities and performance.
- V. The “Conclusion” reflects on functionalities and performance.

Hence, the approach offers a composed general model that builds on the approaches of (Vaishnavi & Kuechler, 2008) and (Hevner et al., 2004) as presented in Figure 3.

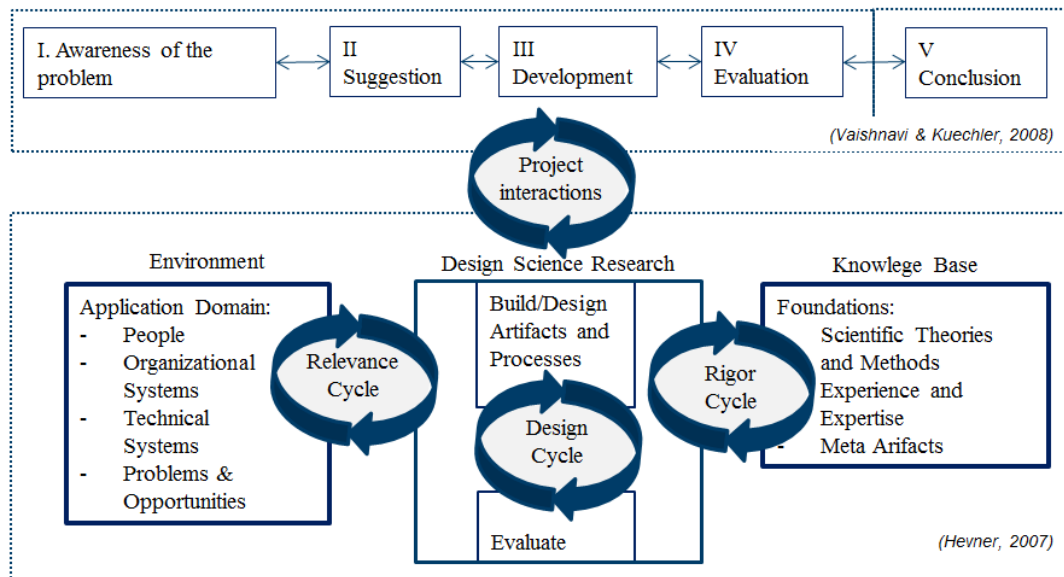


Figure 3: Adapted DSR Framework based on (Vaishnavi & Kuechler, 2008) and (Hevner et al., 2004)

**Awareness of the Problem** (Figure 4): The objective of the first cycle is to gain knowledge and to understand the problem. The result of this first cycle is **paper 1**.

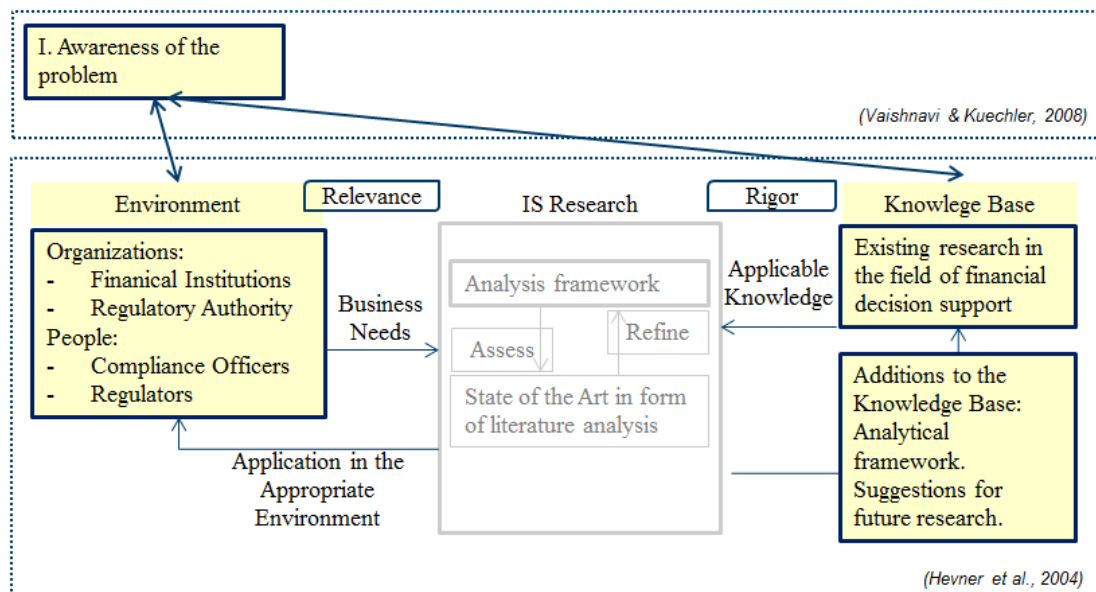


Figure 4: Awareness of the problem cycle

This paper 1 presents a structured analysis of the existing research in the field of financial decision support systems (DSS). Based on the literature reviewed, the paper presents an overview of existing literature, deriving areas for future research. The literature review is structured using a model combining design theory, decision support system

and information mining components; future research can be structured likewise. The analysis from the current studies suggests three key classes: financial analysis, fraud detection and risk management. Hence, it provides an overview of the existing relevant literature and derives problem definition and problem diagnosis by providing suggestions for future research.

**Suggestion, Development and Evaluation** (Figure 5): The second cycle builds on the results of the first cycle. Further iterative steps comprise interviews with domain experts. The knowledge gained in these interviews was mapped into design requirements for a qualitative multi-attribute model. To enhance the relevance of the model, several meetings with domain experts were conducted. Furthermore, this stage is also supported by the evaluation cycle. The final model experienced diverse small refinements until its completion and evaluation in **paper 2**. The developmental phase is accompanied by the evaluation cycle. In this phase an artifact is created to address the challenging problem of detecting fraudulent behavior in financial markets (**paper 3**). This paper examines a detection strategy of pump and dump manipulation to thwart fraudsters from unlawful profit techniques. Hence, an IT artifact instantiation in the form of a model-based decision support system that supports decision making in the field of financial market surveillance is presented. This artifact utilizes a qualitative decision model to identify situations in which prices of single stocks are affected by fraudsters who aggressively advertise the stock. An evaluation of the implemented system based on voluminous and heterogeneous data including user-generated content data is provided.

Finally, to ensure the rigor of the results, in several iterations, the designed artifacts and evaluation are abstracted to conceive an explanatory design theory (**paper 4**). The objective of this paper is to provide design suggestions that enable effective development of Financial Market Surveillance Decision Support Systems (FMS-DSS) for financial institutions. The study is guided by research questions aimed at determining general components and general requirements for financial market surveillance systems with the ability to detect a variety of market manipulations originating from social media usage.

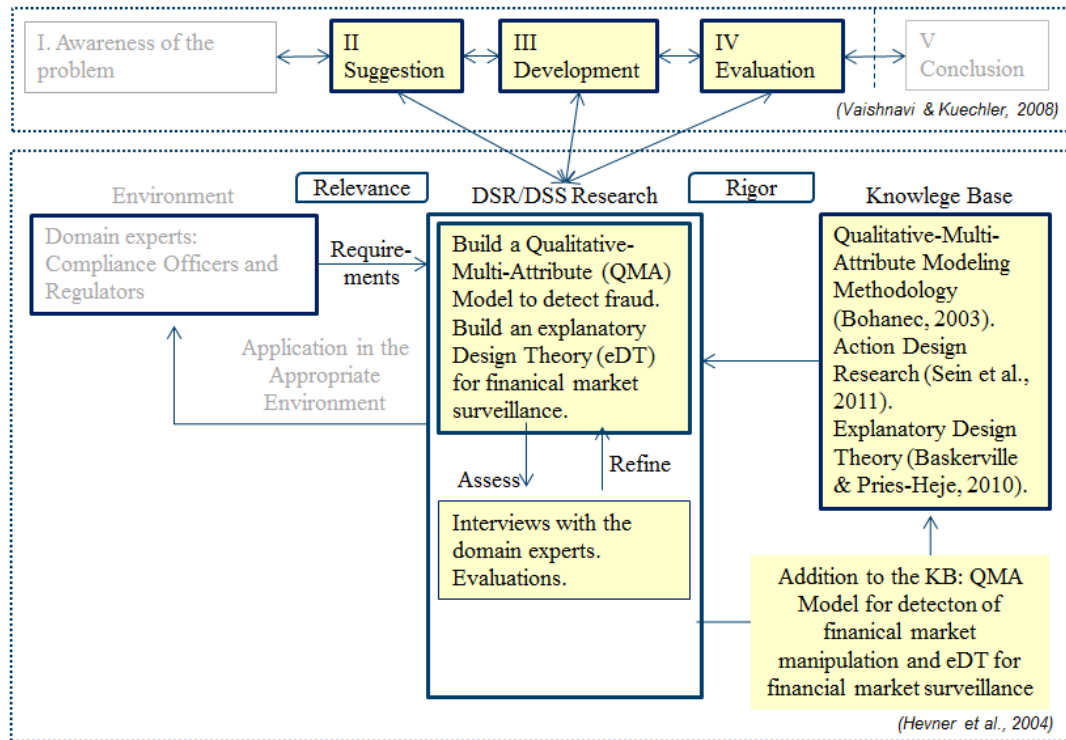


Figure 5: Cycles of Suggestion, Development and Evaluation

**Conclusion:** Lastly, the research project concludes with lessons learned from the project as presented in Section C.

## SECTION B: Studies





## 5 Paper 1: State of the Art in financial DSS and Problem Statement

<b>Title of Article</b>	<b>State of the Art of Financial Decision Support Systems based on Problem, Requirement, Component and Evaluation Categories</b>
<b>Author</b>	Irina Alić, Jan Muntermann and Robert W. Gregory Georg-August-University of Göttingen
<b>Published</b>	25th Bled eConference eDependability: Reliable and Trustworthy eStructures, eProcesses, eOperations and eServices for the Future June 17, 2012 – June 20, 2012; Bled, Slovenia
<b>Abstract</b>	Financial decision support has become an important information systems research topic and is also of highest interest to practitioners. Two rapidly emerging trends, the increasing amount of available data and the evolution of data mining methods, pose challenges for researchers. Thus, a review of existing research with the goal to guide future research efforts in this domain is timely. To structure our literature review and future research in this area, we propose a framework in the paper that integrates elements of decision support systems, design theory, and information mining. The framework is then applied in the paper. Our analysis reveals that the focus of existing research can be grouped into three major domain categories. More research is needed in two of the categories for which we found very few IS studies despite the high relevance of these topics due to increased turbulences in worldwide financial markets. Furthermore, we discuss the opportunities to make stronger use of heterogeneous data and of combined data mining techniques and to build upon the rich set of available evaluation methods.
<b>Keywords</b>	Literature Review, Financial Decision Support Systems, Structured Data, Unstructured Data, Heterogeneous Data, Information Mining, Text Mining, Data Mining.

Table 1: Fact sheet of publication nb. 1

## 5.1 Introduction

The financial services industry belongs to one of the most knowledge- and information intensive industries. As a result, there are massive amounts of data available, steadily increasing, which may be used as a basis for financial decision making. For example, nowadays a financial investor can draw upon multiple data sources, including news and rating agencies, the trading venues or newer sources of data such as financial twitter feeds, blogs and other social media content. The challenge is to make effective use of this data to improve financial decision making, for which in practice often a combination of different data types is required. Such heterogeneous data includes both unstructured textual data and structured data such as time series with a structure described in a schema (Arasu & Garcia-Molina, 2003). However, both the amounts of data and analytical challenges overwhelm practitioners, motivating further research.

We explore the contribution that information systems (IS) can make to the domain of financial decision making through the lens of Decision Support Systems (DSS), which represents one of the major research streams in IS research (Banker & Kauffman, 2004). Power (2001) defines DSS as an interactive computer-based system developed to support decision makers to identify and solve problems and make decisions. DSS are needed to cope with the massive amount of available data and enable financial decision making. Therefore, the topic of decision support in the financial domain is of highest practical relevance (Manyika, et al., 2011).

From a scientific perspective, a decent amount of research has been published over the last one to two decades that is directed towards understanding how to design effective decision support systems to support financial decision making. Therefore, we argue that it is time to conduct a systematic review of prior research in this important domain and thereby provide guidance for future research. Our research question is: What is the state of the art of knowledge about financial decision support using unstructured and structured data? Despite the high practical relevance of this topic, there are still important gaps in the literature and a synthesis of prior research is needed to guide further research.

The rest of this paper is organized as follows. In the next section we present our research methodology, including the theoretical framework that we developed. The fol-

lowing discussion of results from our literature review is structured according to this framework. The final section of our paper provides suggestions for future research.

## 5.2 Methodology

In this section, we first present our analytical framework. The approach of employing such a framework to structure and guide the literature review is an established approach (e.g., Dibbern, Goles, & Hirschheim, 2004). Thereafter, we explain our process of literature identification, selection, and analysis.

### 5.2.1 Analytical Framework

As explained by Markus et al. (2002), DSS represent one of the most prominent types of design theories that has driven an entire research stream in IS research. The concept of DSS originates from the work of Scott Morton (1971). While several definitions of IS design theory exist in the literature (Walls, Widmeyer, & El Sawy, 1992), we use the definition of explanatory design theory provided by Baskerville and Pries-Heje (2010), which defines IS design theory as a set of general components that are related to a set of general requirements with the overall goal of solving a class of problems (see next Figure).

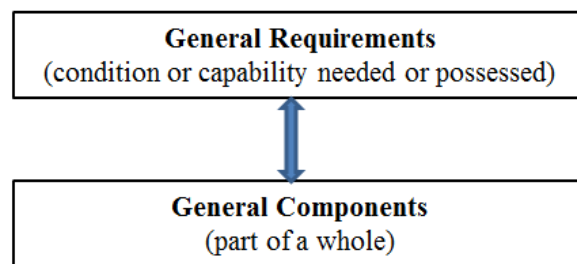


Figure 6: Design theory according to Baskerville and Pries-Heje (2010)

To identify general requirements and components of interest, we utilize the DSS classification framework of Power (2004). We selected Power's extended framework because this is one of the established frameworks to classify DSS Systems. According to this framework, DSS can be first categorized according to their dominant component driver, resulting in five different types of DSS: data-driven, model-driven, knowledge-driven,

document-driven, and communication-driven DSS (Power, 2004). In addition to this dominant component driver, there are three additional components in the extended framework: target user (for example individuals, groups and/or departments), the purpose of DSS (for example purpose that helps to support the targeted users) and the enabling technology for the construction of a DSS (Figure 7).

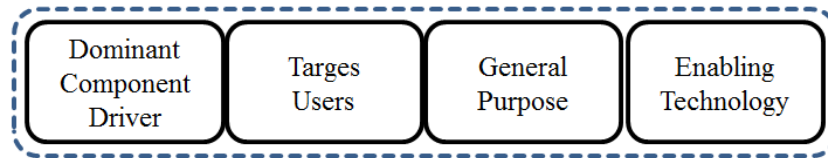


Figure 7: An expanded DSS framework based on Power (2004)

Furthermore, we selected the information mining framework by Gopal et al. (2011), because it presents the state of the art of information mining today. Accordingly, information mining is defined as “the organization and analysis of structured or unstructured data that can be qualitative, textual, and / or pictorial in nature with any set of techniques or methods.” (Gopal et al., 2011, p. 728). The framework consists of the following components: data type (for example textual, numerical or graphical data), application area (which could be software engineering, financial engineering, marketing, or other), techniques (for example SVM, neural networks or other data mining techniques), tasks (for example pattern matching and classification) and it consists of the final objective as the output component (e.g., diagnosis, profit) (Figure 8).

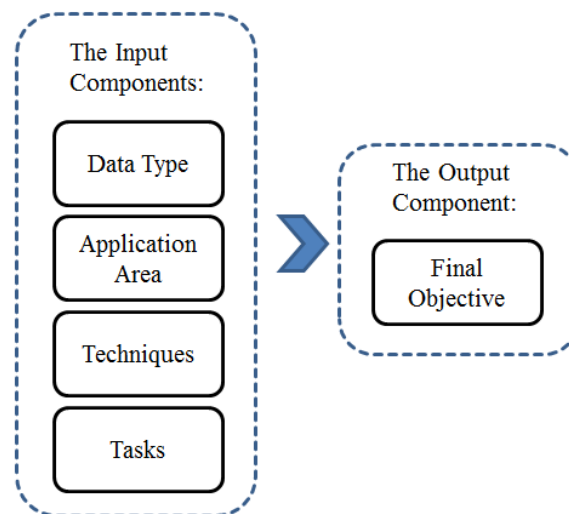


Figure 8: The key components of information mining based on Gopal et al. (2011)

We combined these frameworks (Figure 7 und Figure 8) into a holistic framework by employing the lens of design theory (Figure 6). Accordingly, we structure the components into four categories. First, the problem category (P), which specifies the problem in a domain area and the target user(s), where domain area is taken from the framework of Gopal et al. (2011) and target user is taken from the framework of Power (2004). Second, the requirements category (R), which specifies the purpose(s) according to Power's (2004) framework and the task(s) according to Gopal et al.'s (2011) framework. Finally, the components category (C), which is specified by data and methods according to Gopal et al.'s (2011) framework and by technologies according to Power's (2004) framework. In addition, we complement our framework by an evaluation category (E), based on design science literature that states the importance of evaluating design artefacts and theory (Hevner et al., 2004). The following figure summarizes the holistic framework.

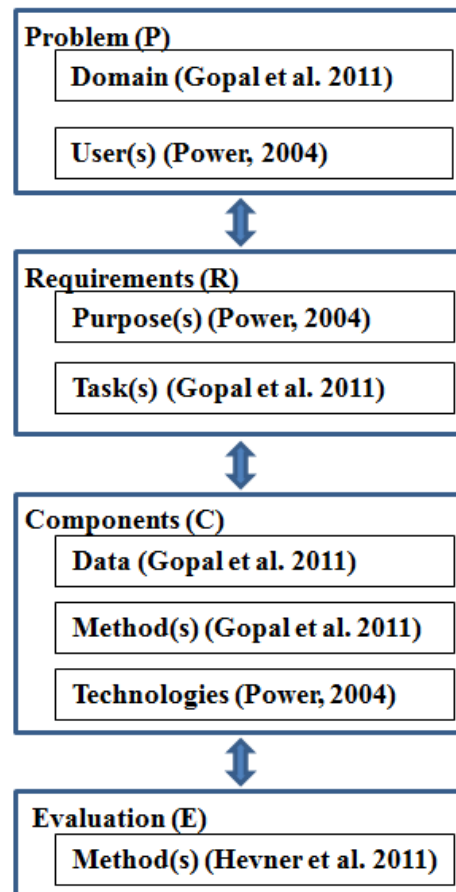


Figure 9: Analytical framework for literature review

Next, we explain our process of literature identification, selection, and analysis.

### 5.2.2 Identification Process

As a first step, we identified potentially relevant literature with a structured keyword search in a selection of scientific databases (Brocke et al., 2009; Ferber, 2003). Thereby, we limited our search to international peer-reviewed scientific literature. To retrieve articles from our keyword search that are potentially relevant for the topic of this literature review, we constructed a Boolean search string based on the components of our framework: ‘decision support and financ\* and mining’. This search string was used to search in the following databases: EBSCOHOST, ScienceDirect, JSTOR, IEEE Xplore, ACM Digital Library, and AIS Electronic Library. The search yielded 176 articles.

### 5.2.3 Selection Process

In the next step of our research process, we reduced the number of articles from 176 to 18. First, we carefully read the titles, abstracts, and selectively the introductory and conclusions sections of the 176 identified papers and removed papers from our list that did not deal with the defined topic, resulting in 17 articles after two iterations. Second, we carefully reviewed the remaining studies and conducted a forward and backward search based on Webster and Watson’s (2002) recommendations. The database Web of Science was used for the forward search. The final sample consisted of 18 articles.

### 5.2.4 Classification and Analysis Process

To apply and use our framework (Figure 9) for our literature review, we first selectively coded each article. The coding scheme derived from our framework, together with sample quotations from our data, is summarized in the following table.

Table 2: Coding scheme, together with examples from our analysis

<b>Code derived from the framework</b>	<b>Sample quotation from literature analysis</b>
P_user	P_user: individual investors "...where individual investors represent the target user group of the system..." (Muntermann, 2009, page 83)
P_domain	P_domain: corporate credit rating "Company credit ratings are typically very costly..." (Huang, Hsinchun, Hsu, Chen, & Wu, 2004, page 543)
R_purpose	R_purpose: portfolio selection "We formulate the winner and loser portfolio selection as two binary classification problems." (Huang, Lai, & Tai, 2011, page 20:7)
R_task	R_task: time series forecasting "...provides another promising tool in financial time series forecasting..." (Tay & Cao, 2001, page 340)
C_data	C_data: unstructured "...based on empirical dataset that comprises 425 company announcements..." (Muntermann, 2009, page 84)
C_method	C_method: single and multiple SVM "For each region, one SVM expert is constructed." (Tay & Cao, 2001, page 349)
C_technology	C_technology: "mobile devices and messaging services provide the enabling technology that provide flexible information supply and decision support on the basis of wireless communication technologies." (Muntermann, 2009, page 84)
Evaluation	Evaluation: Evaluation metric - "We conduct an experiment to measure the performance of our approach ... the rate of degeneration is slow, and the total overall accuracy drops gradually from 89.09% to 71.34%." (Chan & Franklin, 2011, page 8)

Coding reliability was achieved through a control of the first author's coding by the coauthors, following by intensive group discussions.

### 5.3 Results of Literature Review

In this section, we present the results of our literature review. Table 3 summarizes our coding of the literature. We discuss our analysis results according to the elements of our framework. For example, from our problem domain coding we identified three generic problem domains, which are financial analysis (Table 3, reference number 1 to 12), risk management (Table 3, reference number 13 to 17), and fraud detection (Table 3, reference number 18). Since the problem domain is strongly related to requirements category, we discuss these two together.

Table 3: Classification of articles

Reference	Problem	Requirements	Components	Evaluation
1. Wüthrich, Leung, Permuntillek e, Sankaran, Zhang, & Lam (1998)	P_Domain: Finan- cial analysis P_User: Investors	R_Task(s): Predic- tion of stock mar- ket daily move- ments of five stock indices R_Purpose(s): Support invest- ment decision	C_Data: Unstruc- tured (articles downloaded from Web) C_Method(s): Probabilistic rules	Evaluation metric: Accuracy between 40-46.7%. (for periods of 3 months) and over 60% (for few weeks)
2. Tay & Cao (2001)	P_Domain: Finan- cial analysis P_User: Investor	R_Task(s): Predic- tion (currency exchange rates) R_Purpose(s): Support invest- ment decision	C_Data: Struc- tured financial time series (stock index futures, 10/30-year gov- ernment bonds , given as daily closing prices) C_Method(s): Multiple SVM, single SVM	Evaluation metrics and statistical analysis: Compari- son between mul- tiple SVM and single SVM. The multiple SMV method outper- forms the single SVM
3. Gidófalvi & Elkan (2003)	P_Domain: Finan- cial analysis P_User: Investors	R_Task(s): Pre- diction of stock price R_Purpose(s): (Performance) portfolio manage- ment	C_Data: Unstruc- tured C_Method(s): Naïve Baye	Simulation and domain evaluation metric: Average profit per trade
4. Peramunetille ke & Wong (2001)	P_Domain: Finan- cial analysis P_User: Currency traders	R_Task(s): Predic- tion of intraday currency exchange rate movements R_Purpose(s): Buying of one currency and sell- ing of anotherde- cision	C_Data: Unstruc- tured (market news headlines) C_Method(s): Rulebased algo- rithm (based on 400 keyword de- livered by domain experts)	Simulation and evaluation metric: Accuracy 53% for DEM/US and 3 hours
5. Huang, Nakamori, & Wang (2004)	P_Domain: Finan- cial analysis P_User: Investors	R_Task(s): Predic- tion of Index quote R_Purpose(s): Supporting in- vestment decision	C_Data: Struc- tured (NIKKEI 225 Index) C_Method(s): SVM combined with other meth- ods	Evaluation metric: Hit ratio of com- bined model 75%
6. Pui, Fung, Yu, & Lu (2005)	P_Domain: Finan- cial analysis P_User: Investors	R_Task(s): Predic- tion of stock price movement R_Purpose(s): Supporting in- vestment decision	C_Data: Unstruc- tured and struc- tured (intraday stock prices and news stories) C_Method(s): SVM	Simulation: Buy and sell decision based on trend forecast Correct prediction if m=5 days is 65.4%
7. Brandl & Keber	P_Domain: Finan- cial analysis in FX	R_Task(s): Predic- tion of EUR/USD- exchange rates	C_Data: Struc- tured C_Method(s):	Simulation: Out- performs a defined



(2006)	market P_User: Foreign exchange market brokers	R_Purpose(s): Supporting investment decision	Genetic algorithm	benchmark
8. Muntermann (2009)	P_Domain: Financial analysis P_User: Private investor	R_Task(s): Prediction of stock price movement R_Purpose(s): Supporting investment decision	Unstructured (425 company announcements and corresponding intraday stock prices) C_Method(s): OLS regression, machine learning proposed	Simulation and evaluation metric: Statistical tests to compare supported trader with a random trader
9. Schumaker & Chen (2009)	P_Domain: Financial analysis P_User: Trading professionals	R_Task(s): Prediction of stock price movement R_Purpose(s): Supporting investment decision	C_Data: Unstructured (financial news) C_Method(s): Support vector regression (SVR)	Simulation and statistical analysis: F-measure 85%
10. Tsai & Hsiao (2010)	P_Domain: Financial analysis P_User: Investors	R_Task(s): Prediction of stock price movement R_Purpose(s): Supporting investment decision	C_Data: Structured (financial and macroeconomic variables from the Taiwan Economic Journal Database) C_Method(s): Genetic Algorithm, Decision Tree and Neural Net	Simulation and evaluation metrics: Average accuracy 75.34%
11. Huang, Lai, & Tai (2011)	P_Domain: Financial analysis P_User: Investors, analysts	R_Task(s): Prediction of stock price movement R_Purpose(s): Supporting investment decision	C_Data: Structured (historical stock prices of leading U.S. companies of S&P100 Index) C_Method(s): SVM, AdaBoost	Simulation and evaluation metrics: Accuracy 66.41%
12. Chan & Franklin (2011)	P_Domain: Financial analysis P_User: Investors	R_Task(s): Prediction of financial trend or behaviour R_Purpose(s): Supporting investment decision	C_Data: Unstructured (2000 financial reports) C_Method(s): Decision Tree	Evaluation metrics: Accuracy between 71.34% and 89.1%
13. Huang, Hsinchun, Hsu, Chen, & Wu (2004)	P_Domain: Credit rating P_User: Bond raters	R_Task(s): Prediction of bond rating R_Purpose(s): Supporting Credit rating decision	C_Data: Structured (bond-rating data sets from the US and Taiwan markets) C_Method(s): SVM and Neural Net	Simulation and evaluation metric: Accuracy ~80%
14. Sinha & Zhao	P_Domain: Credit rating	R_Task(s): Credit rating classifica-	C_Data: Structured (given as years of previous	Simulation, comparison of methods: Neural net

(2008)	P_User: Domain experts	tion R_Purpose(s): Performance comparison of data mining classification methods incorporating domain experts knowledge	residence, monthly income and payments) C_Method(s): Naïve Bayes, Decision Tree, Neural Net, k-Nearest Neighbour, SVM	84.8%, kNN 77.8%, SVM 78.2%
15. Ruggieri, Pedreschi, & Turini (2010)	P_Domain: Credit scoring P_User: Manager	R_Task(s): Classification of potential discriminatory risks R_Purpose(s): Discover and measure discrimination in credit scoring model	C_Data: Structured (transactions representing the good/bad credit class of bank account holders and beneficiary demographics) C_Method(s): Rulebased algorithms	Simulation: Comparison of two inference models on the basis of historical data
16. Groth & Muntermann (2011)	P_Domain: Market risk management P_User: Risk manager	R_Task(s): Prediction or intraday stock price volatilities R_Purpose(s): Trading decision	C_Data: Unstructured and structured (news stories and related stock prices) C_Method(s): Naïve Bayes, kNN, Neural Net, SVM	Simulation and evaluation metric: Best results with SVM 78.49% accuracy
17. Huang & Li (2011)	P_Domain: Market risk management P_User: Investors, accountants	R_Task(s): Extraction of risk factors R_Purpose(s): Risk management	C_Data: Unstructured (risk factors reported in SEC 10-K form) C_Method(s): k-Nearest Neighbour	Simulation and evaluation metric: Accuracy (74.94%) and four metrics for multi-label classifications, search for optimal kNN parameters
18. Kirkos, Spathis, & Manolopoulos (2007)	P_Domain: Fraud detection P_User: Compliance officers	R_Task(s): Analysis of financial statements R_Purpose(s): Detection of fraudulent statements	C_Data: Structured (financial ratios extracted from financial statements of 76 Greek manufacturing firms) C_Method(s): Decision Trees, Neural Net and Bayesian Belief Network	Evaluation metric: classification performance (decision tree 73.6%, neural network 80%, Bayesian belief network 90.3%)

### 5.3.1 Problem Domain and related Requirements

We find that past research has addressed three different problem domains, which we discuss separately in the following.

**Domain of financial analysis.** Research in this domain covers the following tasks: Prediction of stock price movement (eight studies), prediction of exchange rate movements (three studies), prediction of index movement (one study) and prediction of bond ratings (one study). We found that the observed articles focus mostly on unstructured data (eight studies of thirteen). With regards to applied methods, we found that there is no coherency between unstructured data and applied methodology in this problem domain. The reason for this may be situated in the computational complexity of natural language, causing researchers to evaluate different methods in order to find the most appropriate one for the particular task. We also find that coherence between structured data and applied method exist. The most popular method here is the SVM, reasonably because it achieves very good prediction performance when applied to structured data (Huang, Nakamori, & Wang, 2004; Huang, Hsinchun, Hsu, Chen, & Wu, 2004).

**Domain of risk management.** This domain comprises research of market and credit risk management. Four studies cover one of the following tasks: discriminatory risks detection, extraction of risk factors from disclosures, credit rating, and credit scoring. Two studies are based on structured and two studies on unstructured data. Further, we found no coherency between data and applied methods. This may be explained by the small number of research studies.

**Domain of fraud detection.** This domain has a focus on the detection of fraudulent financial statements. Since we found only one research study, it is safe to say that research in this category is still in its infants. This finding is noteworthy because of the following reason: Manipulated financial statements are attributed to market abuse and subsequently cause improper/inadequate behavior of investors (Financial Services Authority, 2012). Despite the financial crisis in the last years, there is lack of academic research about fraud detection and market surveillance. Apparently, there is a lack of understanding of how to detect fraudulent information circulated in the financial domain.

**User.** Target users of a DSS can either be a member or customer of an organization, including both individuals and groups (Power, 2004). Present studies primarily investi-

gate requirement aspects of financial DSS without explicit involvement of these users. It appears that target user is mostly only mentioned in the present studies and requirements and problem statements have been derived from the literature only.

In the following, we discuss cross-domain findings, according to the components and evaluation categories of our framework.

### 5.3.2 Components

**Data.** The evidence of types of data analyzed in the studies shows that recent studies on DSS in finance tend to use unstructured data in form of company announcements, news stories, or text data downloaded from internet sources, including user-generated contents. Including unstructured data into the analysis for improving decision support in finance domains becomes more popular.

Table 4: Classification of articles by data

	1998	2001	2004	2005	2006	2007	2008	2009	2010	2011
Structured		1	1	1	1	1	1		2	1
Unstructured	1	2		1				2		3

Since the beginning of the year 1998, we observe the regular publication of financial DSS related research studies. While between 1998 and 2007 nine relevant articles are published, for the years 2008 to 2011 we count also 9 relevant publications. This finding affirms increased relevance of financial DSS.

**Method(s).** Our analysis of the applied methods reveals that 11 different data mining techniques have been used in the reviewed research articles. In the next paragraph we briefly discuss the three most frequently applied techniques. Support Vector Machine (SVM): SVM is the most frequently used data mining technique in our sample. SVM is an algorithm where the classifier is a hyperplane, which separates the feature space into different categories (Witten & Frank, 2005; Feldman & Sanger, 2007). SVM is a supervised learning method, which has been developed by Vapnik and Chervonenkis (1974). Neural Networks (NN): NN emulate human pattern recognition. It consists of connected neurons, which are able to receive and send impulses to and from its neighbors. Deci-

sion Trees (DT): DT's classifier consists of nodes, where internal nodes are labelled by the features, each having its own weight (Witten & Frank, 2005). The documents are categorized starting by the root node and moving to the leaves, which are the classes of the document (Feldman & Sanger, 2007). The following Table 54 summarizes all methods applied in the studies with Probabilistic Rules (PR), Rule Based (RB), Mean Absolute Abnormal Return (MAAR), Naïve Bayes (NB), Bayesian Belief Networks (BBN), Genetic Algorithm (GA), k Nearest Neighbor (kNN), and Support Vector Regression (SVR), apart from the above discussed SVM, NN, and DT.

Table 5: Classification of articles by methods

	1998	2001	2004	2005	2006	2007	2008	2009	2010	2011
Method(s)	PR	NB, RB, Mul- tiple SVM, Single SVM	SVM, NN	SVM, SVM Com- bined	GA	DT, NN, BBN	NB, DT, NN, kNN, SVM	MAA R,SV R	RB, com- bi- nation of NN, DT, and GA	SVM, DT, kNN, NB, NN

Our analysis reveals that for decision support in financial analysis, the combined methods applied on structured data delivers promising results. In the research of Tay & Cao (2001) it was shown that a SVM combined with a self-organizing feature map (SOM) outperforms a single SVM by 0.25%. These research findings are consistent with the results of another study (Huang, Nakamori, & Wang, 2004), in which increased accuracy is reached by applying SVM with other methods including a neural networks.

**Technologies.** We find that a great majority of the selected papers do not provide information regarding the underlying technologies. This observation can be explained by the fact that most studies do not present an artefact instantiation (i.e. prototype), but mainly forecasting on classification models.

### 5.3.3 Evaluation Methods

All reviewed papers present an evaluation that has been conducted either on the basis of evaluation metrics (e.g. accuracy, precision and recall) or on the basis of a simulation (which may incorporate evaluation metrics allowing comparisons with alternative designs). This observation is noteworthy since the design science literature presents a rich set of design evaluation methods, including both qualitative (e.g. case studies and controlled experiments) and quantitative methods (e.g. optimization or simulation). Consequently, none of the papers observed the contribution's performance within its original organizational context.

## 5.4 Implications for Future Research and Conclusion

In this study we analyzed the state of the art of financial decision support systems. As a conceptual basis for this, we developed a framework, which consists of four major categories. The analysis results confirm the applicability of our framework and suggest directions for future research along the examined categories:

**Problem and related requirements.** Future research might focus on those domains that remain underexplored. Compared to the field of financial analysis we found only a very limited number of studies in the risk management and fraud detection domains. Considering the financial crises of recent years, these two domains appear highly relevant. Future research in these fields could also build upon domain expert knowledge or the increasing amount of unstructured user-generated contents.

**Components.** While the use of structured data in DSS in the financial analysis domain has been extensively utilized, the exploitation of unstructured data in order to provide decision support is still very limited. The reason may be situated in the complexity of natural language as a computational problem (Burger & Du Plessis, 2011). Accordingly, more research in computational intelligence is needed. Next, we found that the decision support of risk management might need more research in order to refit the organizations in the endeavors of managing the risk using both structured and unstructured data available to the company. These findings relate to those found by Geva and Zahavi (2010) confirming that using both kinds of data could enable better decision making in diverse financial domains. Next, it might be interesting for practice to strengthen the organizational compliance offices by providing information that is useful for decision making.

This information could be for example derived from unstructured data like financial tweets or blogs, providing the insights into current mood states in the market.

**Evaluation.** In general, it appears that the evaluation from the organizational and/or user perspective has been excluded so far. This might be an opportunity for IS researchers to explore and apply the rich set of different evaluation methods, e.g. in order to receive valuable feedback from domain experts. It is widely accepted in the literature that engaging those who are experiencing and know the addressed domain problem can be very beneficial (Van de Ven, 2007). Those focusing solely on generic evaluation metrics and simulation will definitely miss this research opportunity.

In conclusion, in this study we analyzed the current state of the art of financial DSS by developing and applying an analytical framework that may also serve future researchers in this domain to structure their investigations.





## 6 Paper 2: A Qualitative Multi-Attribute Model for Financial Market Surveillance

<b>Title of Article</b>	<b>Hot Stock or Not? A Qualitative Multi-Attribute Model to Detect Financial Market Manipulation</b>
<b>Author</b>	Irina Alić, Georg-August-University Göttingen Michael Siering, Goethe University Frankfurt Marko Bohanec, Institute “Jozef Stefan” Ljubljana
<b>Published</b>	26th Bled eConference eInnovations: Challenges and Impacts for Individuals, Organizations and Society June 9, 2013 – June 13, 2013; Bled, Slovenia
<b>Abstract</b>	The emergence of online financial information channels, such as web portals and financial blogs, eases the challenge process for scammers of publishing fraudulent contents in order to manipulate share prices. To maintain market integrity, financial market surveillance authorities monitor these different information channels to detect suspicious behavior. However, as the available amount of online information increases, analyses become more costly and time-consuming. In order to support related decisions, we have developed a model to identify fraudulent situations. Based on interviews with domain experts, we first identified the factors determining suspicious situations and then applied a qualitative multi-attribute modelling technique. Thereby, our resulting model builds upon valuable knowledge of domain experts and provides means to address the challenge of information based market manipulation.
<b>Keywords</b>	Market Manipulation, Market Surveillance, Qualitative Modelling, Decision Support

Table 6: Fact sheet of publication nb. 2

## 6.1 Introduction

With an increasing number of market manipulation cases observed in recent years, financial market surveillance has gained increased attention in both practice and academia. In one recent case<sup>1</sup>, a prominent US-celebrity published a recommendation for a penny stock investment on Twitter of which he held a significant position. Generally, such stock promotions, if distributed to a large audience, can lead to significant price effects for the respective penny stock, causing abnormally high returns. After these returns are realized, the promoter sells his stocks and since there was no significant change of the company's real value, the other investors run the risk of losing their money due to falling stock prices. Many private investors lack the necessary financial knowledge to judge this situation and are thus particularly vulnerable to such fraudulent stock promotions that make up these so-called "Pump and Dump" (P&D) market manipulation schemes<sup>2</sup>. In order to address such market integrity threats, surveillance authorities need to gain insights into the manipulative behavior of market participants. However, being aware of different information channels and diverse manipulation activities, this task remains cost-intensive and requires a lot of effort. As the available budget of market surveillance authorities is limited (Aggarwal & Wu, 2003), decision support systems may therefore contribute within this context. This research contributes to the on-going discussion of how to support financial market surveillance authorities by analyzing data published in several web-based social networks and portals. Based on expert interviews, during several iterations, we derive the essential indicators for the decision regarding whether a certain financial instrument is suspected of manipulation by a P&D scheme. We present our research contribution in the form of an IT artefact, developed within a multinational design science research project. In doing so, we present a qualitative model that can support users in the decision-making process. We apply a qualitative multi-attribute modelling method to develop a corresponding model, which belongs to the hierarchical decision-making models being suitable for unstructured decision problems (Aggarwal & Wu, 2003), (Ou, Cao, Yu, & Zhang, 2007). This paper is organized as follows: In the next section, we provide a review of the related work followed by a description our methodology. In the subsequent sections, we present our design princi-

---

<sup>1</sup> Chris Barth, Forbes staff: <http://www.forbes.com/sites/chrisbarth/2011/01/11/get-rich-or-die-tweetin/>

<sup>2</sup> The U. S. Securities and Exchange Commission (SEC): P&D Schemes, <http://www.sec.gov/answers/pumpdump.htm>

ples, the model and the experimental results. Finally, we conclude and provide an outlook on future work.

## 6.2 Related Work

In the financial domain, there exists a variety of different market manipulation schemes. An overview and classification of these schemes is provided by (Allen & Gale, 1992; Bagnoli & Lipman, 1996; Aggarwal & Wu, 2003; Mei, Wu, & Zhou, 2004), covering action-based, trade-based, and information-based manipulation schemes. Trade-based manipulation is defined as the action of buying and then selling, whereas information-based manipulation is defined as the publication of false information or false rumors. Thus, action-based manipulation is defined as the actions which are nontrade-based and non-information-based actions. Related to price and volume manipulation, a variety of schemes exists: For example, ramping/gouging, where the broker bluffs the enthusiastic in a specific stock. Another scheme is so called pre-arranged trading, where the participants enters identical price and volume orders at the same time. The next scheme is P&D manipulation scheme. P&D manipulations aims at manipulating the share price by disseminating untrue information in order to make profit from an increased price level (Cumming, Zhan, & Aitken, 2012; Aggarwal & Wu, 2003). If P&D manipulation is defined as a kind of information-based manipulation, we can thus argue that 50-Cent's behavior can be classified as information-based market manipulation due the following reasons: First, faulty and misleading information was spread in a persuasive manner, such as "You can double your money right now. Just get what you can afford". Second, the information was spread over his Twitter account, where it was received by his 3.8 million followers. Finally, he promoted a company whose shares he owns. Taken together, this behavior caused an artificial increase of the stock price, which, when shares were sold at the end of the day, resulted in a breath-taking profit of 8.7 Mio\$. Helpful insights on how to address such manipulation schemes from a market surveillance perspective are presented by (Aggarwal & Wu, 2003). Based on structured data such as the time series, the authors explore how market manipulation affects market efficiency. They show that prices rise during the manipulation phase, only to fall when the manipulation concludes. As noted by (Kirkos, Spathis, & Manolopoulos, 2007) there is little research that utilizes the rich universe of unstructured (i.e. textual) data to support market surveillance activities. Other research scrutinizes the real-time detection of fraudu-

lent activities (Mukherjee, Diwan, Bhattacharjee, Mukherjee, & Misra, 2010). The authors present an information system (IS) for compliance offices for monitoring investment staff by detecting outliers and by performing evaluations along predefined rules. The predefined rules for example assess significant trading volumes observed. With the objective to utilize both structured and unstructured data sources, qualitative multi-attribute model may serve as a basis to address this challenge.

Qualitative multi-attribute models utilize data and values proposed by decision makers, usually domain experts, in order to analyze and address a situation. The qualitative models thus remain highly suitable for unstructured decision problems where approximate judgment prevails over precise numerical calculations (Bohanec, 2003). In previous academic research, qualitative multi-attribute modelling was successfully applied in different domains, including e-learning and ecology (Arh & Blažič, 2007). While we observe a high grade of specialization in the financial domain, there is little research applying qualitative multi-attribute models to build upon the extensive knowledge of domain experts, especially in the field of market surveillance. Thus, in this research, we aim to contribute to the knowledge base and apply the qualitative multi-attribute modelling approach in the market surveillance domain. Therefore, we have developed a qualitative multi-attribute model, which aims at detecting information-based market manipulation in the form of P&D schemes utilizing both structured and unstructured data.

## 6.3 Methodology

### 6.3.1 Design Science Research Approach

Design science research is one of the prominent research paradigms that has driven a research stream in information systems discipline, with the goal of making contributions to the knowledge base on the basis of developed IT Artefacts (Hevner, March, Park, & Ram, 2004). According to the authors, four different types exist: constructs, models, methods, and instantiations. Constructs provide a language for the definition and the communication of the problem and its solution. Models represent the relationships between the constructs. Methods represent procedures to perform specific tasks. Finally, instantiations are based on constructs, models, and methods expressing the implementation in working systems. Because, our research effort aims on providing a set of steps

for a specific problem solution, therefore, our artefact belongs to the group of the model artefacts. As illustrated by (Vaishnavi & Kuechler, 2008), typical design science research projects follows a series of steps:

- I. Awareness of the problem: The first step aims at identifying the problem by conducting the literature review. Within the same step, a drill-down into the problem is required in order to explore the user needs.
- II. Suggestion: After the user needs are explored, potential decision alternatives can be suggested. In this step, the complexity of the whole problem is decomposed to problems of lower complexity.
- III. Development: The aim of the third step is to deliver an artefact, which in our case is a qualitative multi-attribute model.
- IV. Evaluation: After the artefact is being developed, the evaluation aims at exploring its functionalities and performance.
- V. Conclusion: The design cycle ends by providing judgements on the developed artefact.

In our research efforts, we adapt these general process steps to guide our development of the artefact.

### 6.3.2 Qualitative Multi-Attribute Modelling Methodology

Within the development phase III, we aim to develop a qualitative multi-attribute model to assess decision alternatives. The models can be developed in several different ways; the most common is via expert modelling. The model is developed on the basis of interviews with experts. Qualitative multi-attribute modelling is being conducted in a series of four steps (Bohanec, 2003):

1. Identifying attributes: Aims at identifying the important attributes of the decision problem.
2. Structuring attributes: Aims at composing the attributes into hierarchical groups, and enabling decomposing into smaller and possibly more manageable sub-problems. Thus, in this step, we are able to present our model. The model is refined within of the following two steps of defining scales and rules.
3. Defining attribute scales: aims at describing the scales of each attribute (e.g. very-low, low, medium, high, very-high).

4. Defining aggregation rules: based on the step before, the scales are evaluated individually and then aggregated by the model into an overall utility: The higher the utility, the appropriate alternative.

Once developed, qualitative models specify a working method of the evaluation of “objects”, which can be easily embedded into software systems such as a decision support system. For the development and the experimental evaluation of our qualitative multi-attribute model, we use the DEXi software (Bohanec & Rajkovič, 1990).

### 6.3.3 Proposed Research Design

In our research, we combine the general design process cycle of design science research (Vaishnavi & Kuechler, 2008) and the qualitative multi-attribute modelling methodology (Bohanec, 2003) to guide our model development. Our resulting research approach is represented in Figure 10.

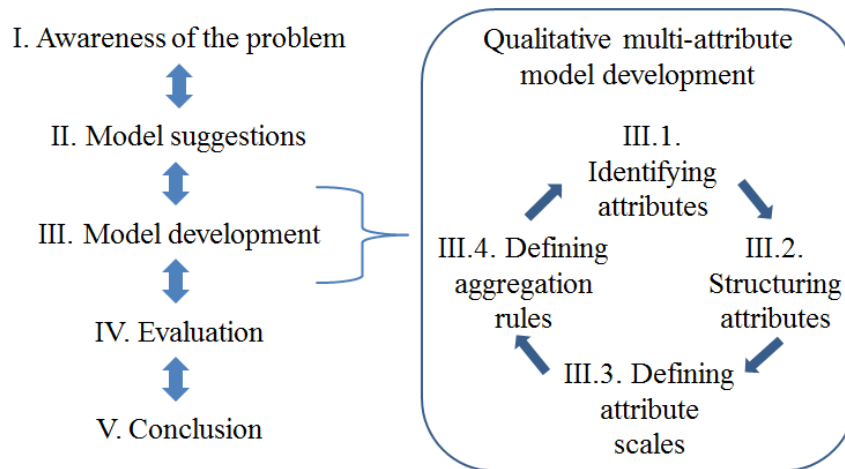


Figure 10: Research approach based on (Vaishnavi & Kuechler, 2008) and (Bohanec, 2003)

In the subsequent section, we explain the development of our artefact. Here, we follow design step I, II as suggested by (Vaishnavi & Kuechler, 2008); step III constitutes the development of our qualitative model; steps IV and V are presented thereafter.

## 6.4 Problem Definition

The goal of the first process step is to generate problem awareness through identification and definition of the specific decision problem. We conduct this step by means of both a survey of related academic literature exploring the problem and an investigation of the user-specific needs. We initially agree with the definition of financial surveillance as stated by (Heping, 2006), namely that such surveillance “[...] refers to a comprehensive non-stop process of a fully automated or interactive intelligent financial system(s) for continuous monitoring of the target markets.” (p. 2/15). Furthermore, we use the definition of P&D schemes as stated by (Kyle & Viswanathan, 2008), that “In a pump-and-dump manipulation scheme, the perpetrator first acquires a large long position, then publishes false information to induce market participants to push prices up by buying the asset, and finally liquidates his own long position at a profit”, (p. 276). Due to the fact that the amount of information published in different online media is increasing continuously, our study focuses on the detection of P&D market manipulation scenarios like the ‘50 Cent’ example. To fully understand the problem at hand, we identified four people working in a market surveillance context who have both substantial knowledge about various market manipulation schemes and experience with systems for detecting financial market manipulations. One expert has 20 years of experience as a decision maker, another two with approx. 10 years’ experience in developing market surveillance solutions. The fourth one is an expert of a European financial regulatory authority. In several interviews, the experts were asked to explain typical factors for P&D market manipulations. The questions posed to them are grouped into three categories (Table 7).

Table 7: Interview structure for problem definition

Category	Explanation
Problem description	Precisely define the problem. What do we access? What is the decision about? Is the multi-attribute modelling a suitable approach to the problem? Who are the actors? Who is affected and who is responsible for the decision?
Requirements description	What are the goals and functionalities of an appropriate problem solution?
Relevant inputs for addressing requirements	What are the characteristics of typical pump-and-dump scenarios? How can P&D be detected, what parameters/variables need to be observed? What are the parameters/variables and their meanings? Which factors/aspects must one take into account when scanning the market by hand in order to detect market manipulation? Can those factors/aspects be measured? Are ordinal qualitative scales of measurement appropriate (e.g., high, low, medium)? What could an entity-relationship-model look like? What are the relations between them? What kind of data do we require?

## 6.5 Decision Model Suggestion

In the interviews, one of the main requirements identified is the generation of a market surveillance indicator, i.e., an alert. Such a surveillance indicator would support the decision of the compliance officer (e.g., regulatory authority) which is decomposed into investigating the suspicious trading behavior, then communicating the suspicious behavior with the source before escalating the situation and taking further steps if necessary. The detection of suspicious market behavior appears to be a complex problem in the means of deciding which of the observed patterns are suspicious and which are not. Therefore, we aimed at developing a qualitative multi-attribute decision model to support decision alternatives by assessing suspicious or non-suspicious market situations.

## 6.6 Qualitative Multi-Attribute Model Development

### 6.6.1 Attribute Identification

In the interview context, the experts stated that a main determinant of P&D manipulations is “The publication of untrue information within different news channels”. Furthermore, the experts stated that this kind of news covers certain financial instruments, which are issued by certain companies. Consequently, the P&D problem is divided into



these main attributes: news, financial instruments and companies. Thereafter, in telephone interviews and further face-to-face meetings, the lists of attributes have been refined. Thus, the detection of suspicious situations is based on the following considerations:

- **Company:** Previous research shows that companies whose stocks are recommended in P&D market manipulation schemes mostly lack prospect business (Rockness & Williams, 1988). In order to determine whether a company is suspicious, the experts consider two possibilities. First, if a company has already been part of such a manipulation. As one of the experts stated: “Financial regulators frequently issue warnings or litigation releases, and the company is put on a blacklist of suspicious firms”. Second, the history of the company is taken into account. The experts state that: “Market manipulators usually target new companies or companies that have been bankrupt”. An examination into the company’s history can uncover aspects of company’s past practices which could justify doubts regarding the reliability of its market activity. Accordingly, the attribute ‘Company’ is refined by attributes ‘Blacklists’ and ‘History’.
- **Financial Instrument:** The financial instrument also needs to be assessed in order to detect potentially suspicious situations. In this case, the experts focus on the question of whether the financial instrument is listed in a suspicious market segment, i.e., “...in a segment with low regulatory requirements, it is easier to published manipulated information in the form of corporate disclosures, among other things”. Furthermore, the experts state: “low market capitalization” is seen as an “additional indicator of a suspicious financial instrument” since corresponding stock prices can adjust on the basis of lower trading volumes, (as opposed to large-capitalized stocks). Finally, a significant change in trading volume or trading behaviour can also be seen as suspicious. Accordingly, the attribute ‘Financial Instrument’ is refined by attributes ‘Market Segment’ and ‘Market Capitalization’.
- **The Attribute ‘News’** estimates the suspiciousness of published news based on following criteria:
  - **Content:** Covers, as stated by one of the experts: “Whether the web publication includes specific content, e.g. increase in revenue, new product development”. The sub-model assesses suspiciousness according to content considered over predefined periods of time.

- Sentiment: Incorporates the sentiment expressed within the news source. In this respect, the experts state that “A positive significant change in sentiment could indicate a suspicious situation”. Therefore, the submodel assesses the suspiciousness of estimated sentiments over a predefined period and compares it with the sentiments of a longer period.

### 6.6.2 Attribute Structure

As a result of the conducted interviews, the model was extended. In this context, the interrelations between the attributes have been defined. Hereafter, three main groups of attributes were identified as illustrated in Figure 11:

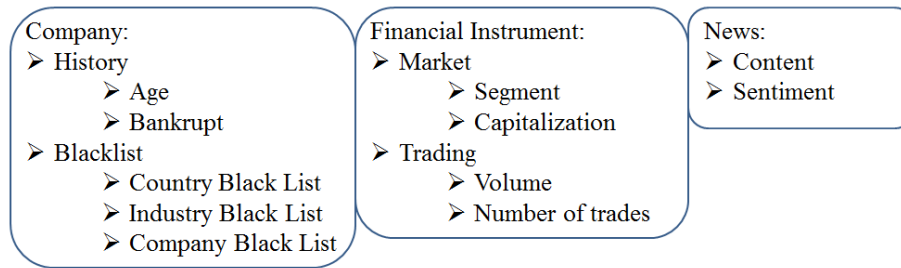


Figure 11: Structured attributes for assessment of P&D cases

On the basis of further interviews, we developed an attribute structure in the form a hierarchical tree. The model is refined into attributes which can be structured and measured so that finally the attributes can be represented as tree of attributes (Bohanec, 2003). In the model, the problem is decomposed into various components, namely:

- Root node, as the target attribute, representing an indicator which determines whether a suspicious market situation prevails.
- Internal set of aggregated attributes (e.g. ‘History’), which is used to structure the attributes relevant to the decision.
- Final set of basic attributes (e.g. ‘Country Blacklist’), representing attributes that can be measured, e.g. by means of data analysis.

Figure 12 represents the tree structure of the proposed model. We derive two sub trees: A sub tree for the news, and a sub tree for the company and the related financial instrument.

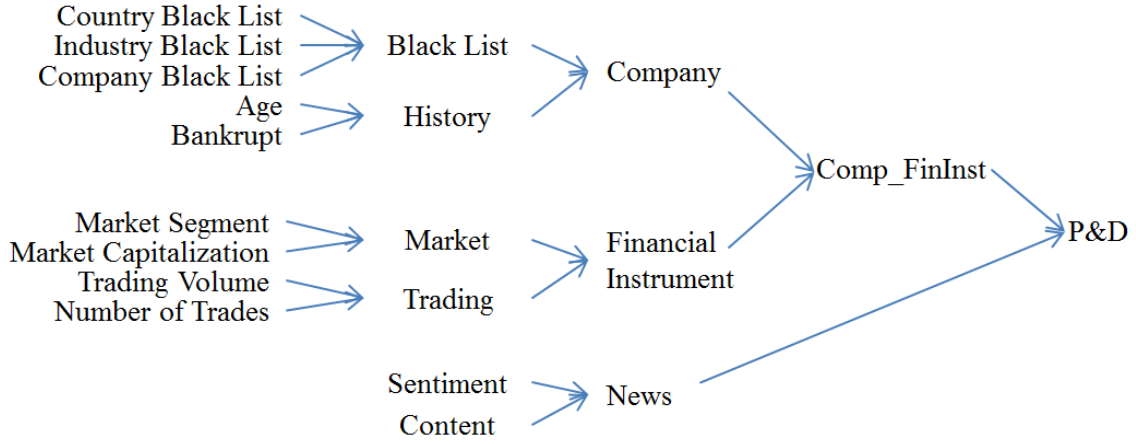


Figure 12: Model of attribute structure

### 6.6.3 Attribute Scales

The value scales for each attribute are set in cooperation with the domain experts. Each attribute can take values from the corresponding scale. Most scales are ordered from ‘good’ values (non-suspicious situation) to ‘bad’ (indicating a highly suspicious situation). For example, the attribute ‘CountryBlackList’ can either be ‘yes’ or ‘no’. If a company originates from a country which is black-listed, the corresponding value will be set to ‘yes’. Such country blacklists are provided by regulatory authorities. The complex aggregated attributes (Comp\_FinInst, Company, Black Lists, History, Financial Instrument, Market, Trading, and News), are dependent upon the lower level attributes (Country-, Industry-, Company Black List, Age, Bankrupt, Market Segment, Market Capitalization, Trading Volume, Number of Trades, Sentiment, and Content). The scales then consist of three to five values:

- P&D: The high level attribute is an aggregated attribute. It indicates the suspiciousness of P&D situations. The values v-high, high, med, low, v-low indicate the suspiciousness.
- Comp\_FinInst: The aggregated attribute. It detects the suspiciousness for the attributes Company and Financial Instrument, and is labelled as v-high, high, med, low or v-low.

- **Company:** The aggregated attribute of Black Lists and History. Assess the potential suspiciousness of the company as v-high, high, med, low or v-low.
- **Black Lists:** If the company appears in any of the blacklist, then the aggregated attribute is labelled as low, medium or high.
- **Country-, Company-, and Industry blacklists:** The values are either yes, indicating the appearance in the list, or no.
  - **Country Blacklist:** Countries which do not rely on the global international standard (FATF – The Financial Action Task Force) to combat money laundry and terrorism. The current lists can be accessed via the FATF<sup>3</sup>.
  - **Industry black list:** There is no black list predefined by a reputable organization such as a regulatory authority. However, it became apparent that suspicious financial instruments are oftentimes issued by companies for which it is not clear in which industry they are operating in. The experts state that, if this information is available from data vendors for financial instruments and the issuing companies, then it is considered to be reputable. In Contrast, missing industry type data can be seen as suspicious.
  - **Company black list:** Companies who are either not approved by a regulatory body or are involved in stock fraud, can be found on e.g. the SEC list<sup>4</sup>.
- **History:** The aggregated value indicates the suspiciousness as low, medium or high.
  - **Age:** The low level attribute indicates the suspiciousness as old, med or new. Where old stands for older companies. In our case, the experts define old >10 years.
  - **Bankrupt:** If the company was insolvent/bankrupt in its history, the low level attribute can be labelled as was, no or is.
  - **Financial Instrument:** The aggregate attribute, which assess the potential suspiciousness of the financial instrument, is labelled as v-high, high, med, low or v-low.
- **Market:** The aggregated attribute assess the potential suspiciousness of the market with the labels low, med or high.

<sup>3</sup> <http://www.fatf-gafi.org/topics/high-riskandnoncooperativejurisdictions/documents/fatfpublicstatement-16february2012.htm>

<sup>4</sup> <http://www.sec.gov/litigation/suspensions/2012/34-66980.pdf>

- Market Segment: If the market segment is potentially suspicious then, the label ‘yes’ appears. Otherwise, the label ‘no’.
- Market Capitalization: Small capitalization is labelled ‘low’. In our case, according to the experts, small capitalization appears to be less than 5Mio\$. High capitalization is defined as more than 30 Mio\$. The range between 5 and 30 Mio\$ is labelled as ‘med’.
- Trading: The aggregated attribute assesses the potential suspiciousness of the trading behavior and is labelled low, med, and high.
  - Trading Volume: Recent changes in the market volume are labelled low, med or high.
  - Number of Trades: Recent changes of number of trades are labelled low, med or high.
- News: Aggregates the attributes Sentiment and Content. Assesses the potential suspiciousness of a news, and is labelled as v-high, high, med, low or v-low.
  - Sentiment: Assessment of the long and short term sentiment. Indicates the change between long and short term sentiment, and is labelled as vhigh, high, med, low or v-low.
  - Content: Assessment of the long and short content. Indicates the change between long and short term content. Assess the recent changes in the suspiciousness of the news, and is labelled as v-high, high, med, low or v-low.

#### 6.6.4 Aggregation Rules

In qualitative models, decision rules serve as the aggregation of values from the basic to the root attribute. For each aggregate attribute in the model, a table of rules specifying the values of the said attributes for all combinations of values in the lower-level attributes is defined by the interviewed domain experts. The root attribute ‘P&D’, for example, depends on the lower-level attributes ‘Comp\_FinInst’ and ‘News’. The corresponding decision rules have been defined as shown in Figure 13. Rules 24, and 25 illustrate situations of very high suspiciousness, which occur whenever ‘News’ and ‘Comp\_FinInst’ are either high or very-high. Rule 1, however, demonstrates that the suspiciousness of a situation is very low only when the attributes ‘News’ and ‘Comp\_FinInst’ are both of very low suspiciousness.

	Comp_FinInst	News	P&D
1	v-low	v-low	v-low
2	v-low	low	low
3	v-low	med	low
4	v-low	high	med
5	v-low	v-high	high
6	low	v-low	low
7	low	low	low
8	low	med	med
9	low	high	med
10	low	v-high	high
11	med	v-low	low
12	med	low	med
13	med	med	med
14	med	high	high
15	med	v-high	high
16	high	v-low	med
17	high	low	med
18	high	med	high
19	high	high	high
20	high	v-high	v-high
21	v-high	v-low	high
22	v-high	low	high
23	v-high	med	high
24	v-high	high	v-high
25	v-high	v-high	v-high

Figure 13: Decision rules for P&amp;D attribute

## 6.7 Experimental Evaluation of the Qualitative Multi-Attribute Model

In order to demonstrate the validity of the developed model, we go along with (Hevner et al., 2004)'s suggested experimental evaluation. Following the approach by using the evaluation functionality of the DEXi software, we are able to evaluate the behavior of the model. In the following paragraphs, we explain the functional experiment. The proposed model is implemented with DEXi software and used for the evaluation and analysis of suspicious situations. The evaluation is designed to reveal the different levels of suspiciousness in potential P&D cases. Figure 14 shows the input data and results of the evaluations of five hypothetical situations. The first of these situations appears not suspicious given the specific attribute values. The consequent evaluation yields only low and very low values with all the aggregate attributes. In contrast, the last two situations are assessed as high and v-high suspicious, as an inspection of the internal model variables reveals. The fourth and fifth situations are suspicious, the former due to the high volume of highly positive news and the latter as a result of several elements, among

them, the company's appearance on the black list, its recent entry into the market, its unusually high trading volume of the financial instrument. In the last situation, there is also an indication that short-term sentiment in the news is perhaps too positive. Once complete, the results of this evaluation can be presented to the domain experts in order to appraise the decision rules of the model. Although these results are based on those decision rules developed in cooperation with the same domain experts, these exemplary results can lead to further reflections on the model and a further adjustment or refinement of its components.

Attribute	V-low susp.	Low susp.	Med. susp.	High susp.	V-high susp.
P&D	<i>v-low</i>	low	med	high	<i>v-high</i>
Comp_FinInst	<i>v-low</i>	low	med	med	high
Company	<i>v-low</i>	low	high	high	high
BlackLists	<i>low</i>	med	<i>high</i>	<i>high</i>	<i>high</i>
CountryBlackList	<i>no</i>	<i>no</i>	<i>no</i>	<i>no</i>	<i>yes</i>
IndustryBlackList	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
CompanyBlackList	<i>no</i>	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
History	<i>low</i>	<i>low</i>	<i>low</i>	med	med
Age	<i>old</i>	med	med	<i>new</i>	<i>new</i>
Bankrupt	<i>no</i>	<i>no</i>	<i>no</i>	<i>no</i>	<i>no</i>
FinancialInstrument	<i>v-low</i>	low	low	low	high
Market	<i>low</i>	<i>low</i>	<i>low</i>	<i>low</i>	med
MarketSegment	<i>no</i>	<i>no</i>	<i>no</i>	<i>no</i>	<i>no</i>
MarketCapitalization	<i>high</i>	med	med	med	<i>low</i>
Trading	<i>low</i>	med	med	med	<i>high</i>
TradingVolume	<i>low</i>	med	med	med	<i>high</i>
NumberOfTrades	<i>low</i>	med	med	med	<i>high</i>
News	<i>v-low</i>	low	med	high	<i>v-high</i>
Sentiment	<i>v-low</i>	low	med	high	<i>v-high</i>
Content	<i>v-low</i>	low	med	high	<i>v-high</i>

Figure 14: Evaluation Examples

Within our expert interviews as well as during the regular project meetings, the outcomes of these evaluations have been presented and discussed with both domain experts as well with a representative of a European national financial supervisory authority. Feedback has been used to refine the decision rules, which finally brought us a stable set of these rules.

## 6.8 Conclusion

The fraudulent behavior of market participants represents a topic that has gained increased importance within financial markets. Market surveillance combined with decision support systems enable appropriate analyses of large amounts of data to support decision making in this field. One severe market manipulation scenario is P&D

schemes, where manipulators aim to increase stock prices by dissemination positive but false information in order to sell these stocks at a profit. Within our research, we experienced that detecting related market manipulations requires a great amount of expert knowledge. Therefore we applied qualitative multi-attribute modelling in order to derive a decision model from expert interviews that includes the different factors influencing whether or not a financial instrument is affected by a P&D scheme. The proposed model is developed by means of the DEXi software. From a theoretical perspective, we contribute to the literature on financial market surveillance, especially with regard to the explanation of what factors characterize P&D market manipulations. Based on expert interviews, we find that company and financial instrument characteristics, as well as news-related characteristics like sentiment and content, play a crucial role in identifying P&D market manipulations. Furthermore, the qualitative model also provides insight as to how these different factors are interconnected. From a practical point of view, we provide a decision model that can be included in decision support systems to assist surveillance authorities in identifying suspicious market situations. With further research, we plan to extend the evaluation of our model with both real and artificial data. Our intention is to acquire fraudulent cases from market surveillance authorities that can be used to evaluate the decision model; however, we are aware that related information is not easy to obtain and that a consequent evaluation with artificial data would be more appropriate. For this purpose, we will generate different attribute values for the input parameters and consider the decision models' output. As a next step, we will evaluate whether the domain experts agree with the models' classifications. The model at hand, however, already provides valuable insights into which factors are essential to assess whether a certain financial instrument is suspected of manipulation by a P&D schema.



## 7 Paper 3: IT Artifact Instantiation and Evaluation

<b>Title of Article</b>	<b>Supporting Financial Market Surveillance: An IT Artifact Evaluation</b>
<b>Author</b>	Irina Alić Georg-August-University Göttingen
<b>Published</b>	28 <sup>th</sup> Bled eConference June 7 - 10, 2015; Bled, Slovenia
<b>Abstract</b>	In this paper, an IT artifact instantiation (i.e. software prototype) to support decision making in the field of financial market surveillance, is presented and evaluated. This artifact utilizes a qualitative multi-attribute model to identify situations in which prices of single stocks are affected by fraudsters who aggressively advertise those stocks. A quantitative evaluation of the instantiated IT artifact, based on voluminous and heterogeneous data including data from social media, is provided. The empirical results indicate that the developed IT artifact instantiation can provide support for identifying such malicious situations. Given this evidence, it can be shown that the developed solution is able to utilize massive and heterogeneous data, including user-generated content from financial blogs and news platforms, to provide practical decision support in the field of market surveillance.
<b>Keywords</b>	Big Data, User-Generated Content, Decision Support Systems

Table 8: Fact sheet of publication nb. 3

## 7.1 Introduction

Financial market manipulation has received much attention from regulatory authorities, resulting in trading suspensions of those companies that may have been hijacked by fraudsters. Recent suspensions included two large groups of penny stock companies suspended by the U.S. regulatory authority in a single day (SEC 2012) before they could harm investors. All companies have been traded in over-the-counter (OTC) markets with low regulatory standards. To assess which companies were potentially affected, the supervisory authority used information technology to recognize thinly traded, low-capitalization stocks that could be affected by fraudsters. The system is calibrated to detect companies with low-priced penny stocks that are traded in low volumes or not at all; such companies must also be delinquent in their public disclosures (SEC 2012). Unfortunately, the system reveals its weak point: namely, insufficient real-time surveillance of voluminous and heterogeneous data.

The pump-and-dump manipulation scheme is the most common form of information-based market manipulation in financial markets (Dunham, 2007; Zaki, 2013). It first appears as the spreading of false positive information to market participants by fraudsters; this information, typically user-generated content (UGC) on financial blogs and news platforms, is enthusiastically spread by fraudsters attempting to pump up a stock price to an artificial level. Usually, the affected stocks are penny stocks that trade below \$5. Before starting the manipulation, fraudsters buy a significant quantity of the stocks over a longer period of time. Motivated by the spread of positive information, uninformed market participants rely on the bogus story and buy the stocks, effectively forcing abnormal price increases (pump) (Bouraoui, 2009; Huang & Cheng, 2013). Finally, fraudsters make a profit by selling their stocks at the increased price level, which accordingly causes the stock price to drop (dump).

This paper addresses the question of whether assessing UGC can help regulatory authorities and financial institutions to detect such situations. Therefore, the focus of this study is to present a quantitative evaluation of the final IT artefact that was developed within a major EU project in order to demonstrate the feasibility of the approach. Accordingly, an evaluation based on a voluminous dataset is performed; for this purpose, OTC trading data and corresponding UGC related to 1700 companies were collected during 2012 and 2013. The collected data are used by the developed and instantiated IT artefact to identify situations of abusive behavior, and to assess whether those identified

situations do actually demand more attention than those that were not identified as suspicious.

At this point, it is important to note that the EU project involved several activities in order to provide services, including technologies development as well as the development of data mining algorithms suitable to handle big data complexity, and metadata for further usage for the different use cases. The corresponding development processes and associated findings resulted in more than 20 research articles and project deliverables ("Project FIRST" 2013). Thus, in this present paper, the voluminous data utilized for this evaluation are provided by the mentioned services.

In the next section, the related background research is introduced. Afterward, I provide an overview of the developed IT artefact instantiation. Section 4 describes the quantitative evaluation based on real data, and the final section concludes this paper.

## 7.2 Background

The study areas relevant to this research in terms of stock fraud detection include the literature on: (1) the impacts of spam email-based stock touting; (2) the impact of user-generated news on stock activity; and (3) financial market services architecture. Each area will be briefly discussed in the following paragraphs.

Several studies examine of the impact of stock touting using spam emails. To generate profit, scammers spread misleading information. One study (Frieder & Zittrain, 2008) investigated market activity prior to and following a stock touting email campaign, making use of a dataset that contained about 75,000 spam emails. The research revealed high market activity beginning one day before email spamming commenced, which continued until the day with the greatest number of touting mails. The authors found that volume and return respond positively to touting, whereas the returns dip significantly after the campaign concludes. Hence, the study suggests two main indicators for market manipulation: abnormal price changes and trading volumes; these indicators will be utilized by our IT artefact. Research conducted by (Bouraoui, 2009) demonstrates similar findings. The author provided an evolution model of volatility to assess the impact of stock spam emails based on a sample of 110 penny stock companies. The research showed abnormal returns three days after the commencement of spam emails. In both studies, the outcomes have been explained as the behavioral effect of market partici-

pants who responded positively to the touting. Thus, these studies considered statistical approaches to explicate the influence of email stock promotion. Other studies introduced data mining techniques to identify stock touting spam emails. Research by (Zaki et al., 2011) observed spam messages to detect highly fraudulent stock activities by utilizing data mining techniques to identify stock touting spam emails. The accuracy detection of these experiments ranged from 58% to 71%. A considerable amount of research on the predictive power of UGC (such as tweets, financial forums, and blogs) on stock prices has been documented in the literature. In one instance, a study (Delort et al., 2011) introduced evidence of manipulation and examined the effect of such misuse in online financial forums. The authors showed that manipulative UGC regarding companies with lower prices and market capitalization positively correlated to stock returns, volatility, and volume. One recent study (Smailović et al., 2013) presented a support vector, machine-based sentiment classifier; here, a set of about 150,000 tweets thematizing eight companies (e.g., Apple, Google, and Microsoft) served as the data. The authors found that positive sentiment predicted positive movement in the closing price. Another study (Zhang & Skiena, 2010) examined the ways in which blog and news data are reflected in trading volumes and returns. The authors demonstrated a significant positive correlation between media content and trading volume as well as stock returns. Therefore, in this study, the UGC data will be incorporated.

Several commercial stock market systems for monitoring and detecting abuse in structured and unstructured data exist; some, such as the Securities Observation, News Analysis and Regulation Systems (SONAR), are presented in scientific research (Goldberg et al. 2003). SONAR, which aims to monitor the stock market, applies data mining, text mining, statistical regression, and rule-based detection to recognize both abuse patterns in structured data and unusual trading following publication of the news. A study by (Mangkorntong and Rabhi 2008) compared two different surveillance systems as event-processing systems in such areas as memory usage, scalability, and flexibility. The authors revealed the strengths and weaknesses of the two systems and suggested a generic approach that uses numerous different event-processing systems to support the detection process (Mangkorntong & Rabhi, 2007).

The next section introduces the Financial Market Surveillance Decision Support System (FMS-DSS), which was developed within an EU project between 2010 and 2013 ("Project FIRST" 2013). As an IT artefact instantiation, it demonstrates how regulatory

authority can be supported in detecting malicious pump-and-dump market manipulations utilizing voluminous and heterogeneous data streams.

### 7.3 Instantiated IT Artifact

During the research project, the opportunity to develop an instantiated IT artefact (FMS-DSS) was provided. Within the research consortium consisting of research institutions, industry partners, and financial regulatory authority, software components have been developed in close researcher-practitioner collaborations. The problem owners (i.e., domain experts and regulatory authority) intervened according to the project's needs and aligned the design principles with their surveillance issues (Alić, Siering & Bohanec, 2013).

Generally, decision support system configurations are built on the basis of three basic technology components related to: (1) data, (2) models, and (3) the user interface (UI) (Turban, Sharda, & Delen, 2010). The following subsections will present these components. The data management component pre-processes and stores the needed data. The model component assesses whether or not given stocks are under suspicion of being currently affected by pump-and-dump manipulation schemes. The user interface component allows the meaningful representation of suspicious situations. During operation, FMS-DSS continuously searches for the appearance of UGC related to monitored companies. For this purpose, unstructured input data are continuously retrieved, pre-processed, and stored in a database.

Based on rules and models that were developed in close collaboration with the domain experts that were involved in the project, the UI shows alerts that indicate how likely a stock is affected by manipulation (ranging from very high, high, medium, low, to very low).

#### 7.3.1 Data

Potential input data categories based on pump-and-dump manipulation scheme evidence were selected as follows. In meetings with experts (i.e., compliance officers and the regulatory body), the specific decision problem in revealing typical factors for pump-and-dump market manipulation was explained. The three main types of information

incorporated into the decision model are: the manipulation of information concerning the company, the manipulation of the financial instrument, and news as UGC about the company and its financial instrument. Consequentially, the three main input data categories of market abuse suspiciousness are Company, Financial Instrument, and News:

- **Company:** According to the experts, there are two determinants for company suspiciousness. First, in those cases where the company is already involved in financial market manipulation, the financial authority issues litigation releases and puts the company on a blacklist, which is later refined into company, country, and industry blacklists.

The second determinant of a company's reliability is its history; the manipulator often targets newer companies and companies that have gone bankrupt and have recovered again. The History attribute is thus refined into the attributes Age and Bankruptcy.

- **Financial Instrument:** This category is refined into the attributes Market and Trading. If a company's financial instrument is listed in a market segment with low regulatory requirements and the company itself has low market capitalization, then this instrument is seen as an additional indicator of suspiciousness. A change in trading volume or trading behavior is also seen as suspicious
- **News:** The UGC spread in social communities is closely analyzed by the model. The attribute News is refined into the attributes Content and Sentiment; the former analyses whether the web publication includes suspicious phrases (e.g., increase in revenue, new product development), and the latter captures the sentiment expressed within the news.

These input data are provided by the developed services and described in (Grčar, 2012; Smailović et al., 2012).

### 7.3.2 Model Description

Based on further interviews, the attributes structure was transformed into a hierarchical tree (Alić et al., 2013), with the root node 'Pump and Dump' in the following P&D, based on a qualitative- multi-attribute-method as proposed in (Bohanec & Zupan, 2004), differentiating between the pure UGC data (News) and the heterogeneous data regarding the company and related financial instruments (Comp\_FinInst). Hence, the tree con-

sists of the two sub-trees: one for ‘Company’ and related financial instrument ‘Comp\_FinInst’ and the second one for ‘News’. The proposed model aggregates the attributes into an assessment of pump-and-dump market manipulation:

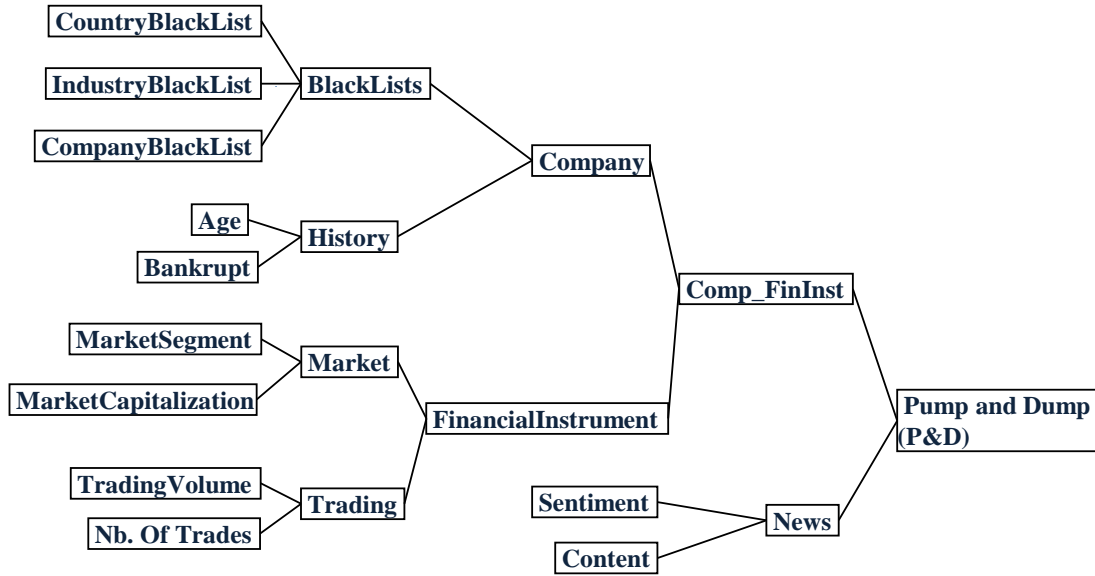


Figure 15: The hierarchical tree of attributes (Alić et al., 2013)

**Attribute scales:** For each attribute, the qualitative values are scaled in the range from highly suspicious (red colored) to not suspicious (green colored), where v-low is an abbreviation for very low and v-high an abbreviation for very high (Figure 16).

Attribute	Scale
P&D	v-low, low; med; high; v-high
Comp_FinInst	v-low, low; med; high; v-high
Company	v-low, low; med; high; v-high
BlackLists	low; med; high
CountryBlackList	no; yes
IndustryBlackList	no; yes
CompanyBlackList	no; yes
History	low; med; high
Age	old; med; new
Bankrupt	no; was; is
FinancialInstrument	v-low, low; med; high; v-high
Market	low; med; high
MarketSegment	no; yes
MarketCapitalization	high; med; low
Trading	low; med; high
TradingVolume	low; med; high
NumberOfTrades	low; med; high
News	v-low, low; med; high; v-high
Sentiment	v-low, low; med; high; v-high
Content	v-low, low; med; high; v-high

Figure 16: The attribute scales (Alić et al., 2013)

The scales for each attribute value are defined by the regulatory authority members and can be reconfigured when in use by the end user.

In our case, for example, the default setting for the attribute ‘Market Capitalization’ is ‘**low**’ (highly suspicious) if the company’s value is under 5 million (given currency), ‘medium’ if the value is between 5 and 30 million, and ‘**high**’ (not suspicious) if the value is greater than 30 million. In the next example the attribute ‘Market Segment’ has two values ‘**no**’ and ‘**yes**’. In this case the value ‘no’ means that the stock is not traded at the market segment with low regulations. Accordingly, the value ‘yes’ means, that the stock is traded at the market with low regulations (e.g. the OTC market).

**Manipulation scheme indicators:** Within the predefined timespan the proposed calculation intends to identify abnormal changes that can be seen as indicators (Eren & Ozsoylev, 2006; Goldstein & Guembel, 2008; Zaki et al., 2012) for pump-and-dump abuse. Hence, the suspiciousness is assessed as follows:

Firstly, to assess recent changes in trading, long-and short-term average trading volumes are computed by taking the monthly and three-day averages of the trading volume:

$$\text{Trading Volume Short Term} = \frac{(\sum_{i=1}^n TV_i)}{n}$$

$$\text{Trading Volume Long Term} = \frac{(\sum_{i=1}^m TV_i)}{m}$$

where  $TV_i$  is the trading volume of the  $i$ -th day;  $n = 3$  days;  $m = 30$  days.

Secondly, to assess recent changes in the number of trades, long-and short-term average trading volumes are computed by taking the monthly and three-day averages of the trading volume:

$$\text{Number of Trades Short Term} = \frac{(\sum_{i=1}^n NT_i)}{n}$$

$$\text{Number of Trades Long Term} = \frac{(\sum_{i=1}^m NT_i)}{m}$$

where  $NT_i$  is the number of trades of the  $i$ -th day;  $n = 3$  days;  $m = 30$  days.



Additional indicators for pump-and-dump market manipulation are calculated as presented in Table 9.

Table 9: Calculation of average values of the input variables

Name	Description/Definition
Sentiment Long-Term Period	<p>Sentiment of news based on assessment of long-term sentiment. Based on the overall picture of the mood of the news</p> $SentiLong = (\sum_{i=1}^m Si)/m$
Sentiment Short-Term Period	<p>Sentiment of news based on short-term sentiment (daily). Based on the overall picture of the mood of the news of one to three days</p> $SentiShort = (\sum_{i=1}^n Si)/n$
User-Generated Content (UGC) Long-Term Period	<p>Content of news based on assessment of specified terms. Based on the overall picture of the mood of the news</p> $ContentLong = (\sum_{i=1}^m Ci)/m$
UGC Short-Term Period	<p>Content of news based on short-term specified terms (daily). Based on the overall picture of the mood of the news of one to three days</p> $ContentShort = (\sum_{i=1}^m Ci)/n$

Thirdly, in order to calculate jumps in e.g. price (Frieder & Zittrain, 2008), the deviation of the short-term as related to the long-term average is calculated by dividing the short-term average by the long-term average and multiplying by 100, as presented in Figure 17. Three cases are assessed: when the short-term value is smaller than, greater than, or equal to the long-term value. Suspiciousness is assessed using aggregated numerical input values, which are then mapped according to qualitative scales, as defined by the problem owner as high, med or low or as v-high, high, med, low, or v-low.

Accordingly, structured data (e.g. trading volume) and unstructured data (e.g. UGC) are thereby taken into account in order to identify abnormal changes that may indicate pump-and-dump market manipulations. The recalibration of the indicator values or even the deployment of predefined default values can be adjusted by the end user.

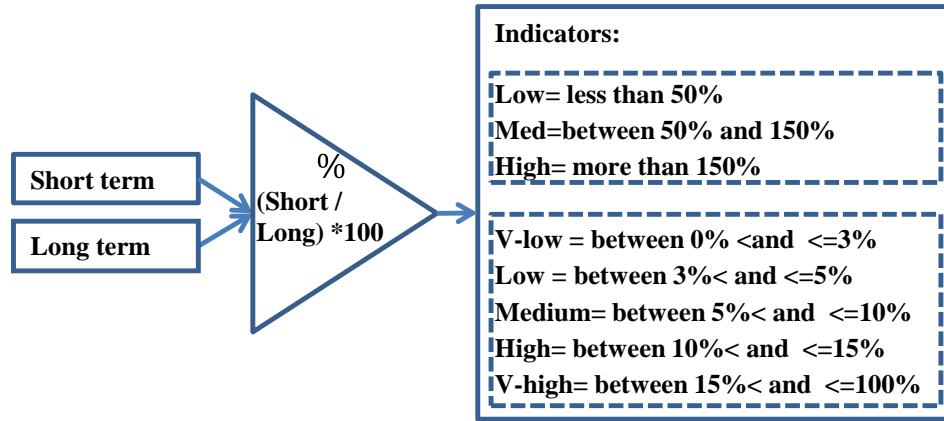


Figure 17: Calculation of artificial jumps as decision rules calibration

**Output Calculation:** The final pump-and-dump alert output value (P&D) is an aggregation of the lower-level attributes Comp\_FinInst and News, whereas the Comp\_FinInst aggregates Financial Instrument and the issuing Company. News aggregates Content and Sentiment of the News regarding the Financial Instrument and the issuing company. P&D aggregates News and Comp\_FinInst and presents the final indicator which indicates whether a suspicious market situation prevails. The scales consist of five values representing the decision rules as depicted in Figure 18.

	Comp_FinInst	News	P&D
1	v-low	v-low	<b>v-low</b>
2	v-low	low	<b>low</b>
3	v-low	med	<b>low</b>
4	v-low	high	<b>med</b>
5	v-low	v-high	<b>high</b>
6	low	v-low	<b>low</b>
7	low	low	<b>low</b>
8	low	med	<b>med</b>
9	low	high	<b>med</b>
10	low	v-high	<b>high</b>
11	med	v-low	<b>low</b>
12	med	low	<b>med</b>
13	med	med	<b>med</b>
14	med	high	<b>high</b>
15	med	v-high	<b>high</b>
16	high	v-low	<b>med</b>
17	high	low	<b>med</b>
18	high	med	<b>high</b>
19	high	high	<b>high</b>
20	high	v-high	<b>v-high</b>
21	v-high	v-low	<b>low</b>
22	v-high	low	<b>med</b>
23	v-high	med	<b>high</b>
24	v-high	high	<b>v-high</b>
25	v-high	v-high	<b>v-high</b>

Figure 18: P&amp;D alert output aggregation

For the v-high alert, three possible combinations exist; the alerts ‘high’ and ‘med’ are respectively defined by seven and eight possible combinations whereas the alert ‘low’ has six possible permutations and the alert ‘v-low’ has one combination.

### 7.3.3 User Interface

Initialized by a routine at the end of the day (midnight), the Surveillance System begins the search for the appearance of UGC regarding the monitored company every day for a predefined time period (e.g., within the last 30 days). For this purpose, the unstructured data are continuously retrieved, pre-processed by the developed services, and stored in the database where the data can be assessed by the FMS-DSS. The FMS-DSS subsequently collects other related input data.

Based on end-user defined rules, the output appears as a signal that can be either a v-high, high, medium, low, or v-low alert. The following figure depicts a detailed view of alerts as displayed on the end-user screen. This view allows the user to drill down to a more detailed level for each alert. The related P&D output values are then stored in the database, enabling the user to analyze the suspect alerts.



Figure 19: FMS-DSS prototype

The FMS-DSS user interface depicted in Figure 19 consists of a screen showing the listed ‘Alerts’ and ‘Details’ with specific information on each alert.

The warnings listed in the ‘Alerts’ screen provide general information to the end user. Here, the company name in the list is replaced with an International Securities Identification Number (ISIN) and the exchange where the security is traded. The screen also displays the alert number under which the specific alert is stored in the database.

The Date specifies the exact alert date. The user can select an alert from the list to view more detailed information.

The ‘Details’ screen depicts the alert selected above. Here, the display divides the information into two parts. The upper part displays unstructured data, and the lower part displays corresponding structured data.

In the unstructured data area, the graph indicates the appearance of positive (green) and/or negative (red) news. The yellow line represents the difference between positive and negative news. In the alert shown in this screen shot, the yellow line appears mostly in a positive sentiment area and the light blue bars represent the sentiment amount; in this example, we note the high appearance of sentiment on the 28<sup>th</sup> and 29<sup>th</sup> of March 2012.

In the structured data, we can see three colored horizontal lines showing the highest (green), lowest (red), and average (black) daily stock prices. The dark blue vertical bars represent trading volume. The alerts, appearing as yellow dots, are evident on the 29<sup>th</sup> and 30<sup>th</sup> of March and the 3<sup>rd</sup> of April 2012.

On the right side of the screen some of the phrases collected from social-net news that have been used in the calculation of sentiment values are shown; this feature allows the end user to observe the corresponding text and follow it to its source and author. Also on the right side of the screen, the Calculation tab contains further information on the calculation of the P&D output alert.

For the user’s convenience, a history of alerts and the corresponding database ID of the specific company are also listed. The end user is thus provided with a range of information to use for financial market monitoring.

## 7.4 IT Artifact Evaluation

This section presents an empirical evaluation of the presented FMS-DSS. In particular, the P&D output data are used for the performance assessment.

This section is organized as follows: first, the data required for the evaluation will be described; then, after presenting descriptive statistics of these data, the relationship between the P&D output values and actual observed stock price changes is analyzed.

There are two different sources of unstructured data are considered in this dataset: UGC gathered from a variety of financial blogs, and news platforms. The unstructured input data are pre-processed and stored in a database, which is then assessed by the other system components as described previously; the data published on regulatory authority pages are likewise stored as input data in the database.

Structured financial data are automatically downloaded by the system from a data vendor (which was involved as one industry partner in the research project), while regulatory data regarding the blacklist are stored as a list in a database and can be modified by the end user.

Overall, three different data sources are considered: UGC, regulatory authority data, and stock price data for the period from 1<sup>st</sup> of January 2012 to 3<sup>rd</sup> of September 2013. As a result, the P&D output sample generated by the FMS-DSS consists of 1700 OTC stock trades, and the corresponding structured and unstructured data contains 118096 entries.

### 7.4.1 Descriptive Analysis

The P&D output data are clustered within the five alert groups: very high, high, medium, low, and very low suspiciousness. The evaluations of suspiciousness in the given dataset of 118,096 entries are presented in Table 10: Examination of the highest and lowest daily price changes Table 10.

Table 10: Examination of the highest and lowest daily price changes

Alert	Count	Percentage
v-high	982	0.83%
High	22215	18.81%
Med	92062	77.96%
Low	2780	2.35%
v-low	57	0.05%

Statistical analysis reveals prevalent medium- and high-value alerts. This may be explained by the sensitivity of the model. Certainly, the model is developed as a qualitative model where pre-defined default values might be too receptive. Allowing the end user to configure their own aggregation rules and set their own thresholds would therefore allow for improved alert sensitivity, to the effect that false-positive alerts can be reduced. The output data presented above already incorporate abnormal volume changes and other market abuse indicators, as defined in previous sections.

#### 7.4.2 Manipulation Examination

According to the domain experts involved in the research project, v-high and high alerts are the most relevant situations that demand further investigation. On that account, the evaluation proceeds by grouping v-high and high alerts into a suspicious class (v-h, h), and the other alerts into a non-suspicious class (m, l, v-l).

In the literature, significant price changes have been described as an indicator (Hanke & Hauser, 2008; Zaki et al., 2011) of stock manipulation. In the following, this measure is applied to determine if the developed IT artifact can help detect potentially suspicious pump-and-dump abuse cases.

In order to evaluate the system's capability to detect such relevant cases, actual price changes in situations that were assessed by the system as highly suspicious (v-h, h) are compared with those that were assessed as less suspicious (m, l, v-l). In doing so, for each v-high, high, medium, low and v-low P&D output value, the changes between the highest ( $p_H$ ) and the lowest ( $p_L$ ) daily stock price at the alert date are calculated. To

assess the changed market price value for the lowest and highest daily prices, the following measure is calculated:

$$f = \frac{p_H - p_L}{p_H} * 100$$

A hypothesis is thus formulated to verify whether actually observed price changes related to the (v-h, h) class are significantly higher than the actual price changes related to the (m, l, v-l) class:

$$H_0: \Delta p_{(v-h, h)} \leq \Delta p_{(m, l, v-l)}$$

$$H_1: \Delta p_{(v-h, h)} > \Delta p_{(m, l, v-l)}$$

In doing so, and given suitably-sized samples ( $N_{v-h, h} = 23197$ ;  $N_{m, l, v-l} = 94899$ ) a t-test is applied to test the formulated hypothesis. The null hypothesis can be rejected at the 1% level. The results are presented in Table 11.

Table 11: Evaluation results based on market price value changes for the lowest and highest daily prices

		Assessed changed market price values		
	N=118096	Mean	Std. dev.	Median
(v-h, h)	N=23197	1.24	3.65	0.41
(m, l, v-l)	N=94899	0.4	1.46	0
p-value	<0.01			
t-value	34.31			

Undoubtedly, it appears that focusing on the group of v-high and high alerts is advantageous, since the t-value is highly significant. Hence, it can be stated that suspicious abuse cases, where fraudsters first buy shares and then attempt to manipulate their price

by spreading extremely euphoric content on social media, can be detected by the implemented FMS-DSS.

This statistical evaluation suggests that the default values, as defined by the practitioners and regulatory authority, might be seen as the best default values. A potential shortcoming from perspective of the end user (e.g. compliance officer in a bank) might be the huge number of potentially suspicious daily alerts (55 in averages) the system provides.

In order to reduce the overwhelming number of positive alerts and to capture only the most probable suspect patterns, further refinements might be undertaken by the end user, such as a threshold routine that calculates the changes between the highest and the lowest daily stock price. For every price change greater than or equal to a specified percentage, the routine will trigger an alert.

An examination of such price changes results in a reduced list of 268 alerts, in other words, about one alert every two days. Thus, the procedure might reduce the noise of less relevant alerts and pinpoint the most suspicious manipulation cases. It appears that filtering based on price changes significantly increases the appearance of v-high values from 0.83% to 23.88% and of high values from 18.81% to 34.33%. The appearance of a medium alert, however, falls significantly from 77.96% to 41.04%. Hence, the final procedure will filter the most promising suspect alerts and help reduce their daily appearance to the lowest reasonable number.

## 7.5 Conclusion

This work has documented the effectiveness of the FMS-DSS and the importance of a market surveillance solution in improving the detection of information-based market manipulation. It has shown how the regulatory authority can be supported by the use of a market surveillance service. The evaluation results show that UGC can be utilized in near-real time using the developed FMS-DSS. The P&D alert results show that the FMS-DSS can handle voluminous and heterogeneous data and provide timely daily alerts by distinguishing suspect and non-suspect market situations. Most notably, this is the first study that investigates the effectiveness of market surveillance decision support that considers three different data sources: UGC, data provided by regulatory authorities, and time series data from a data vendor. Our results provide convincing evidence



for a long-term analysis (of approximately two years) of real data and suggest that the developed artifact may be effective in detecting real abusive cases of pump-and-dump market manipulation. However, some limitations are worth noting; even if the research is supported statistically, the artifact has not yet been evaluated in the real-world conditions of a compliance office. Future research should therefore include a subsequent effort to evaluate the acceptance and use of the running artifact.



## 8 Paper 4: Explanatory Design Theory for Financial Market Surveillance Decision Support

<b>Title of Article</b>	<b>Financial Market Surveillance Decision Support: An Explanatory Design Theory</b>
<b>Author</b>	Irina Alić Georg-August-University Göttingen
<b>Published</b>	<b>28th Bled eConference</b> June 7 - 10, 2015; Bled, Slovenia
<b>Abstract</b>	Sufficient Financial Market Surveillance Systems require both structured and unstructured data. In this research an explanatory design theory for decision support systems is presented which addresses both assets enhanced with further regulatory demands. The developed general requirements and general components of the proposed design theory may serve as guidance for successful implementation of highly flexible and real-time surveillance systems for financial services.
<b>Keywords</b>	Financial Markets, Surveillance Systems, Explanatory Design Theory

Table 12: Fact sheet of publication nb. 4

## 8.1 Introduction

Information sources, such as financial blogs and tweets, seduce nonprofessional investors into investing in potentially suspicious financial instruments (SEC 2012). Many investors struggle with their involvement in faulty investments. For the identification of market abuse, Information Systems (IS) for market surveillance include the detection of notable market abuse patterns in structured data (Eren & Ozsoylev, 2006). However, as of yet, there has been no research that integrates both structured and unstructured user-generated content with the information provided by the regulatory authority in a system that supports financial institutions in their surveillance tasks. The aim of the proposed design theory is to address this research gap by formulating design recommendations for an IS that supports market surveillance decision making.

The research presented in this paper is based on a three-year research project that provided the opportunity to develop an IT artefact to detect market manipulation. From October 2010 through September 2013, the market surveillance system was developed, implemented and evaluated in close researcher-practitioner collaboration (“Project FIRST” 2013). The domain experts and regulatory authorities intervened as needed to align the design theory with their surveillance issues.

This paper intends to contribute to the explanatory picture of market surveillance by providing insights into an explanatory design theory for financial markets as well as to support regulatory authority decision making by proposing a design solution for market surveillance. Thus, the study is led by the following research question: What are the general requirements and general components of financial market surveillance systems that are capable to detect market manipulation initiated via social media?

This paper is organized as follows: the next section provides a brief research background regarding market manipulation and design theories for IS, followed by proposed study design. Finally, a design theory is described, followed by a conclusion.

## 8.2 Research Background

### 8.2.1 Market Manipulation

Allen and Gale (1992) investigated different manipulation schemes, distinguishing between three groups of manipulation strategies. The first group consists of trade-based manipulations used as strategies for buying and selling that do not result in changes to beneficial interests or market risks. The second group is made up of information-based manipulation strategies, where false and misleading information is published in order to manipulate prices. The third group is made up of action-based manipulation strategies, in which compromising actions are undertaken by the management in order to affect the value of the company. Market manipulation related to the illegal disclosure of untrue information by the sender via unstructured data has been explored (van Bommel, 2003). The “pump and dump” market manipulation is one of the most widespread fraud schemes (Securities and Exchange Commission [SEC], 2012), manipulating share prices by first buying a specific stock and then spreading untrue positive information about the company in order to push share prices to an artificial level. The Internet (e.g., financial news platforms and blog forums) is used to spread the misleading information. Profit is then made by selling the stock at this artificial price level (Aggarwal & Wu, 2006). Affecting the share price of penny stock companies is therefore much easier than of large cap companies whose shares are traded by professional institutional investors. In summary, to detect the various types of market manipulation, a corresponding surveillance system needs to handle traditional data (e.g., time series) as well as the non-traditional data (e.g., news, blogs, and twitter platforms).

### 8.2.2 Design Theories for Information Systems

Several studies on theory-building approaches in Design Science Research (DSR) have been published in recent years. In (Hevner, March, Park, & Ram, 2004), seven guidelines were proposed to assist design science researchers in both contributing to IS theory and creating and evaluating as-of-yet unknown and innovative information technology (IT) artefacts. Particularly relevant to this study was the recently-developed IS design theory proposed by Baskerville and Pries-Heje, (2010), which was an explanatory model of design IT artefacts. The theory distinguishes between general components and

general requirements where the components are justified by the requirements. The explanatory design theory explains why a set of requirements is satisfied by a set of components. Hence, only two essential parts are needed for a complete explanatory design theory: general requirements and general solution components. Nevertheless, an evaluation with the domain experts and regulatory authority will be provided.

### 8.3 Study Design

#### 8.3.1 Action Design Research

The general requirements and components of interest are identified and the action design research (ADR) methodology for design science research problems is utilized to merge science and practice (Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011).

The method bridges the gap between research and practice (Baskerville & Myers, 2004) and is appropriate for collaborative projects between scientists and practitioners who wish to develop or improve solutions for real practical problems (Marshall, Willson, Salas, & McKay, 2010). Thus, ADR is appropriate for this project because it is expected to provide a solution to a real-world problem while reflecting on lessons learned (i.e., by formalizing the design theory).

ADR is by its nature an intervention, in this research, not in a unique organizational setting, but on the European regulatory background where financial authorities face the problem of market abuse and the need to counteract such abuses. In order to satisfy the reliability of this research the findings were steadily counterchecked with practitioner of the project consortium including stakeholders from a European financial supervisory authority. More precisely, during sequentially held consortium meetings, developed IT component were presented and the practitioner provided feedback if they provide a solution to the problem.

#### 8.3.2 Research Stages

In our case, the ADR stages are maintained iteratively in cycles of theory and practice steps (Baskerville & Wood-Harper, 1996) and in close collaboration between the participants, leading to the generation of general requirements and general components that

constitute an explanatory design theory. The ADR stages are detailed as follows (Baskerville & Wood-Harper, 1996):

### **Stage 1: Problem formulation**

The project task is to develop an IT artefact to detect market manipulation. Thus, the main driver for development is the support of market surveillance tasks via systematic collection and analysis of any data that can be utilized for decision support. The project involved other researchers and practitioners from both the financial and IT domains; the market surveillance project team consisted of 15 partners (five scientists from Slovenia, Spain, and Germany; two practitioners from the German and Italian regulatory authorities; and eight practitioners from Germany and Italy). The role of the researchers was to consider the problem in order to assess the situation from a scientific perspective and contribute accordingly to the knowledge base. The practitioners worked in a market surveillance context as financial domain experts.

Communication was maintained over a project-web service platform that contained all project-relevant documents (e.g. models, prototypes, documents). The platform was extensively used by both practitioners and researchers.

In this first stage of the project, user needs were identified and problem awareness for a specific goal was generated<sup>5</sup>. From a theoretical perspective, the literature stream on decision support systems (DSS) was examined, the initial questions to be discussed with regulatory authorities and practitioners were settled on, and possible methods were debated. The first meeting was set for this discussion.

### **Stage 2: Building, intervention, and evaluation**

In these process steps, collaboration between practitioners and scientists was motivated by specifying the activities that should lead to the desired solution for the problem. In doing so, the researchers initiated the first semi-structured interview, which included the following questions: What is to be accessed? What is the decision about? Who is the decision maker? Who is affected by the decision? In several further meetings and telephone conferences, the tacit knowledge regarding how to assess the market abuse driver was explored.

---

<sup>5</sup> FIRST Consortium D1.2 Use case requirements specification, [http://www.projectfirst.eu/public\\_deliverables](http://www.projectfirst.eu/public_deliverables).

The data collected in collaborative meetings was analysed instantly within the team of practitioners, users, and scientists. In each meeting, the initial question served as both a starting point for discussions and a focus point for the resultant discussion on gaining a better understanding of market abuse. The attributes were used to enhance understanding of the phenomenon (Hadasch, Mueller, & Maedche, 2012).

Over the course of the project, the entire team met in person several times in annual meetings, each of which lasted three days. In these meetings, development stages were presented, possible improvements and ideas were suggested, and subsequent steps were discussed. The market surveillance team additionally met in person twice a year. In addition, several telephone conferences were conducted.

Initially, the system was designed as a prototype qualitative model (Alić et al., 2013) allowing the derivation of initial design foundations. The prototype was evaluated in two ways. First as a simulation where artificial data was utilized to simulate and prove the usability of the prototype and second as a verification of whether the model addressed the problem. In the subsequent phases, further developments were continuously made, discussed, and evaluated, resulting in a final IS.

### **Stage 3: Reflection and learning**

This was the continuous stage, conducted synchronously with the two first stages. Across all three stages, possible problem solutions were re-conceptualized, ensuring greater generalizability of learning. During the entire project, the permanent involvement of a regulatory authority member and the evaluation phases resulted in the development of general requirements and general components.

### **Stage 4: Formalization of learning**

The learning was incorporated into the outcome, representing a generalized solution to the problem (Sein et al., 2011). In this stage, nine general requirements and five general components were formulated as the design theory for a market surveillance system. Table 13 presents the summary of ADR cycles in the project.



Table 13: ADR Stages based on Sein et al., (2011)

ADR Stages and Principles		Outcome
Problem Formulation		
Principle 1: Practice-inspired re- search	The main driver for this research was the need to support market supervisory authorities in market surveillance tasks.	Recognition: Based on recognized shortcomings the IT Artefact should operate on:  -structured time series data  -unstructured user-generated content data  -and information provided by the regulatory authority.
Principle 2: Theory-ingrained Artefact	General theoretical background related to model-driven DSS (Turban, Sharda, & Delen, 2010)	
Building, Intervention and Evaluation		
Principle 3: Reciprocal shaping	Infrastructure for the retrieval, storage and knowledge extraction from social network was expected to be an ongoing problem.  The developed prototypes were steadily counter-checked with the regulatory authority.	The prototype was designed as a qualitative model.          The prototype was iteratively developed and evaluated within the team resulting in a final IS (Alić et al. 2013)
Principle 4: Mutually influential roles	The role of the researchers was to assess the situation from a scientific perspective. They also acted as the artefact developers.	
Principle 5: Authentic and concurrent evaluation	During the development, the artefact (i.e. the instantiated prototype) was continuously evaluated within the project team including the regulatory authority members. The final IS was evaluated by the potential end-users from financial institutions.	
Reflection and learning		
Principle 6: Guided Emergence	Constant intervention and evaluation lead to re-conceptualization of possible design components.	Refined version of the design.
Formalization of learning		
Principle 7: Generalized outcomes	Formulation of financial market surveillance constituting explanatory design theory: interconnection between theory components and goals to apply the knowledge to the problem class.	A set of general requirements and general components.

## 8.4 An Explanatory Design Theory for Market Surveillance Decision Support

This section provides the general descriptions, units of analysis, and requirements in the construction of the desired system as the results of ADR stages. Further, it explains artefact classifications in order to greater conceptualize generalized components. The meaning of the word “requirements” as it is used by (Baskerville & Pries-Heje, 2010), refers to a “condition or capability needed by a user to solve a problem or achieve an objective.”

### 8.4.1 General Requirements

Through meetings and telephone conferences with the experts, a set of general requirements was established. One of the practitioners pointed out that: “The target users are surveillance staff members who are employed by a regulatory authority.” Other experts on the team highlighted the importance of daily observations: “The surveillance staff members need to prove daily if some bad guys are out there.” As a consequence, the DSS focused on compliance staff members and their daily work activities in the context of market surveillance. DSR on DSS has shown that most systems are designed to support IS practitioners and managerial users as a single user (Arnott & Pervan, 2012). Focusing on classes of systems that support decision-making processes of regulatory authorities, compliance officers in financial institutions can be expected to benefit from market surveillance DSS. The importance of this research is grounded on the nature of financial DSS and the consequences that such newly-introduced methodologies and artefacts can have upon its users. Hence, the general requirements of the system were assumed to be as follows:

**The task:** The market surveillance officers attempt to ensure the proper functioning of capital markets in accordance with the regulation rules (R1).

**The decision support:** The market surveillance officers are supported in their daily efforts to maintain observations of market participants’ abusive behaviours (R2).

Compliance offices are not profitable cost centres (Cumming, 2008), so the user needs to ensure market surveillance is as time-efficient as possible in order to reduce costs. As a result, the following requirements are defined:

**The signalling:** If an anomaly occurs, an alert needs to be generated (R3).

**The surveillance:** Monitoring the market and the market's behaviour implies timely analysis of a large number of financial instruments (R4).

The data monitored is primarily structured (e.g., in a time series). Detecting trade-based manipulation by finding suspicious trading patterns in structured data has already been well examined and employed in market surveillance IS (Cao & Ou, 2008). Regarding the one behind the manipulation, one expert states: "A bad guy is engaged in the market, is interested in selling after he buys low, and starts to spread highly positive news on the social net." The detection of information-based manipulation in recognition of suspicious information published on social media, together with the detection of trade-based manipulation, was therefore mandatory for this research. The systems combine structured data with unstructured social media data to aid in decision making:

**The data:** The ability to deal with heterogeneous data (R5).

Regulatory authorities often recommend transparency of adaptive management while emphasizing specific processes (Linkov, Satterstrom, Kiker, Batchelor, Bridges, & Ferguson, 2006) such as the detection of suspicious patterns in a historical time series of data, the investigation of the transaction, and the escalation to the regulatory authority if necessary (Buta & Barletta, 1991; Lucas, 1993). According to one project expert, financial institutions have to provide the regulatory authority with "detailed information on every potentially abusive case". This implies the following general requirements:

**The rules:** Must be comprehensive (R6).

**Documentation of rules:** Alerts need to be processed and stored for investigative purposes (R7).

The subsequent general requirements are for the precise detection of abusive cases (suspicious behaviour) and the provision of signals if suspicious behaviour appears:

**The history:** The user must have the ability to prove the background of the case that caused an alert (R8).

So as not to overwhelm users with false alarms, as is the case when the rules are too sensitive, the user needs to be authorized to change the rules to a more balanced level. The ability to modify the elements in order to both receive all relevant abusive cases

and reduce the appearance of false alerts is further expected with the cost-reducing measures. This implies the final general requirement:

**Ability to modify the model configurations:** The values of the rules can be changed by the user (R9).

Hence, the unit of analysis in the proposed research was the financial market surveillance decision support system. This system provides all the relevant information necessary to support the regulatory decision making processes. The requirements were evaluated within the team with the purpose of ensuring design theory generalizability, which applies to the class of surveillance systems instead of an instance (Müller-Wienbergen & Müller, 2011). In addition, the developed solutions were presented to the European Commission by the project leaders, presenting achievements and discussing possible modifications of the solutions.

#### 8.4.2 General Components

General requirements derived from interviews with the practitioners in several cycles provided guidance in order to develop suitable IT solutions. Through abstraction and learning general IT components were identified on this basis. In the following, the abstract architecture of the proposed explanatory design theory is presented.

The data sources that will be considered in the market surveillance task are retrieved from the internal sources of the specific organization and from external data sources. The external structured data is usually provided by data vendors via proprietary IS and other delivery systems. The unstructured textual data is collected from the regulatory authority's web sites. Further unstructured data considered in this project was user-generated content collected from several social networks such as blogs. Surprisingly, previous research has ignored the regulatory data and user-generated content data in market surveillance tasks. Thus, a promising research approach may be achieved by assessing all three of these data sources (regulatory-, vendor- and user-generated content). The value-added components for modern surveillance solutions are:

**Internal and external data capturing systems (C1)<sup>6</sup>.**

---

<sup>6</sup> FIRST Consortium D3.1 Semantic resources and data acquisition; D3.3 Large-scale ontology reuse and evolution, [http://www.project-first.eu/public\\_deliverables](http://www.project-first.eu/public_deliverables)

The acquisition of a web data stream can be realized with web APIs, (e.g., Twitter™ API). Such stream-based workflows (up-to-date with the stream) can be built on data mining models, allowing client queries at any time (Saveski & Grcar, 2011). The unstructured data relevant for market surveillance retrieved from external sources, such as blogs, tweets, news web pages, and regulatory web pages, is stored here. This data is characterized as highly informative (Zhang & Skiena, 2010), and can be used to assess the investors' opinions (Klein, Altuntas, Riekert, & Dinev, 2013). The data retrieved from data vendors, such as structured financial time series data, can also be processed and analysed using data mining techniques (Gopal, Marsden, & Vanthienen, 2011). The general requirements for rules 2, 4, and 5 are thus satisfied. Consequently, the component that provides these services can be taken into account is:

### **Data storage and analysis (C2)<sup>7</sup>.**

For the huge amounts of unstructured data, techniques for extracting and adapting information from the text are necessary (Park & Song, 2011). Thus the component comprises the preparation of unstructured data for further use in the workflow process. For this purpose, the scientific literature offers two different approaches, namely ontology-based methods and data mining methods (Klein et al., 2011). Ontology is the formal specification of the vocabulary and its relationships in the domain (Gruber, 1993). The data mining method, particularly the text mining method, deals with the transformation of the natural text into numerical vector values (Feldman & Sanger, 2007). For the purpose of sentiment analysis, one of the sophisticated techniques is the 'active learning principle' where the output is represented by the model for sentiment classification, (e.g., positive or negative financial tweets) (Saveski & Grcar, 2011). In order to maintain the time-critical surveillance tasks of compliance officers, the methods for automatic sentiment classification are obligatory, satisfying general requirements 1 and 2. The component is therefore the further value-added component for modern surveillance solutions:

### **Processing of unstructured data (C3)<sup>8</sup>.**

---

<sup>7</sup> FIRST Consortium D6.5 Highly Scalable Interactive Visualization of Textual Streams v2, [http://www.project-first.eu/public\\_deliverables](http://www.project-first.eu/public_deliverables)

The data applied from the data processing unit serves as input to the knowledge repository, allowing the user to assess the data. Furthermore, as the repository meets general requirement 8 by comprising the information from internal databases and further external data sources, it stores all involved data in the alert signal. The data is further utilized by several models and rules and is stored in the repository. The models to which this research refers are quantitative data mining models, qualitative multi-attribute models, and further market surveillance rules that can detect market anomalies or abusive behaviour ("Project FIRST" 2013). Qualitative multi-attribute models were developed in the interviews with experts and were suitable for the evaluation and analysis of decision alternatives (Bohanec, Žnidaršič, Rajkovič, Bratko, & Zupan, 2013). The data mining models that handle forecasting from large unstructured and structured data sets for the detection of notable or suspicious patterns were also developed. Thus, the following component satisfies general requirements 3, 6, 7, 8, and 9:

#### **Glass-box model of the knowledge repository (C4)<sup>9</sup>.**

The component should ensure an enhanced understanding of the occurring phenomena and facilitate the decision making processes for the compliance officer. With visualization of the text mining results, along with the qualitative multi-attribute results, the user is deeply involved in the processes of alert generation. The most appropriate visualization can be represented as a decision tree (Liu & Salvendy, 2007) viewed graphically as a set of connected decision nodes and leafs. While the nodes carry the attribute values, the user can use their tacit knowledge regarding pattern recognition and change the attribute values if necessary. This ensures a better understanding of the data samples. This component fulfils general requirements 6, 8, and 9 by employing rule-based methodologies for comprehensibility of rules, vivid representation of the history of occurrence, and ease in rule modification:

#### **Graphical user interface (C5)<sup>10</sup>.**

---

<sup>8</sup> FIRST Consortium D4.3 Large-scale Semantic Information Extraction Components; [http://www.projectfirst.eu/public\\_deliverables](http://www.projectfirst.eu/public_deliverables).

<sup>9</sup> FIRST Consortium D6.2 Machine Learning and Qualitative Models; [http://www.projectfirst.eu/public\\_deliverables](http://www.projectfirst.eu/public_deliverables)

<sup>10</sup> FIRST Consortium D2.1 Technical requirements and state-of-the-art; [http://www.projectfirst.eu/public\\_deliverables](http://www.projectfirst.eu/public_deliverables).

For flexibility, market surveillance DSS needs to be modular, and the solution can be integrated into existing systems. Details regarding data integration and sentiment analysis are provided by (“Project FIRST” 2013). Table 14 summarizes the contributions this study makes to the scientific knowledge.

Table 14: Design theory for financial market surveillance DSS

General Requirements	General Components
(R1) Proper functioning of capital markets in accordance with the regulation rules.	(C1) Internal and external data capturing systems.
(R2) The user is supported in his daily efforts to maintain observations of market participants’ abusive behaviours.	(C2) Data storage and analysis.
(R3) If an anomaly occurs, an alert will be generated.	(C3) Processing of unstructured data.
(R4) Timely analysis of large number of financial instruments.	(C4) Glass-box model of the knowledge repository.
(R5) Use of heterogeneous data.	(C5) Graphical user interface.
(R6) Comprehensive rules.	
(R7) The rules can be configured by the user.	
(R8) Storage of alerts for investigative purposes.	
(R9) The user has the ability to prove the background of the case which caused an alert.	

## 8.5 Conclusion

The goal of this research on explanatory design theory development was to support decision making for market surveillance enforcement. The approach of theory development was based on the development of an instantiated IT artefact addressing identified user requirements. The emerging qualitative data exploration of semistructured interviews with team members was carried out with the goal of determining important decision attributes where the exploration was predicated by ADR. Further, an explanatory theory-building method was applied. From a practical perspective, the general requirements and components represent the design theory that provides guidance for the development of market surveillance IS. Furthermore, from the cost perspective, where market surveillance is emphasized as a time consuming cost centre, this study provided insights into the development of more efficient surveillance systems. From a theoretical perspective, this research contributes to the literature on financial market surveillance by enhancing future development strategies of explanatory design theories to solve a class of problems. The theory development approach was based on prescriptive research, and

---

accordingly, it built on the suggestions for development. This research is limited by the fact that it is based on interviews with European domain experts and regulatory authorities. It could be argued that non-European experts have a different point of view of market surveillance. Additionally, this research considers only English articles. Future research could be enhanced by adding non-EU regulatory authorities and by utilizing non-English data sources. To reduce bias during the project phase, the researchers tried to remain in close contact via email, Skype™, team views, and face-to-face meetings with the experts. Even so, there could be limitations in researcher bias due to the fact that the researchers' goals and those of the expert's sometimes differed, leading to restriction in generalization.



## 9 Executive Summary of Papers

### Paper 1: State of the Art in financial DSS and Problem Statement

This paper presents a structured analysis of existing research in the field of financial decision support systems (DSS). On the basis of a literature review, the paper presents an overview of the existing literature and derives areas for future research.

In the introduction, the subject is motivated by explaining the increasing volume of available financial data and more sophisticated methods in the field of data analysis and data mining. A comprehensive section follows that describes the adopted methodological approach, i.e., how the research articles were derived from the databases and how they were analyzed in depth using a tagging technique. In particular, the combination of two frameworks (decision support systems and information mining) appears convincing as well as the use of design science criteria for the denomination of the tags. The contribution results of the research are three-fold. First, three domains (investment management, risk management and market surveillance) are identified where DSS are reported to be applied in the relevant articles. Second, the data mining techniques that were observed in the articles revealed a trend towards the analysis of unstructured data and the combination of mining methods. Third, implications for future research are derived that include the exploitation of unstructured data, the increasing relevance for decision support in the domain of risk management and the use of the methods for market surveillance and the analysis of sentiments/opinions. The contribution of this paper is based on the developed analytical framework that may also help future researchers in this domain structure their investigations.

## Paper 2: A Qualitative-Multi-Attribute Model for Financial Market Surveillance

This paper addresses the increasingly important question of fraudulent behavior emerging on the Web designed to manipulate share prices. Here, a decision support system is presented to support the user in detecting malice behavior with respect to stocks price. Nevertheless, the amount of information available on the web is overwhelming, and individuals do not possess expert knowledge that would prevent them from being defrauded. Reports indicate that there are financial market surveillance authorities that monitor different information channels to detect suspicious behavior. Hence, this research provides a multi-attribute decision model based on expert knowledge that helps raise awareness and supports decision-making in the field. More specifically, the paper presents a prototype of implemented system and describes how the fraud detection alerts are generated.

An evaluation is presented using real and large amounts of data based on existing companies. Based on a user site investigation, the paper relates the key attributes of the investigated information-based market manipulation. The design science research study shows how the discovered attributes could be interconnected so as to build a decision tree.

The main conclusion of this paper is that financial market surveillance systems must deal with complex data, which in this research is the data gathered from social media combined with the data provided by the regulatory authority and other data vendors. The paper contributes to the regulatory authority by providing distinctive knowledge within the financial market surveillance systems.

### Paper 3: IT Artifact Instantiation and Evaluation

This paper examines a detection strategy of pump and dump manipulation to thwart fraudsters from unlawful profit techniques. Hence, an IT artifact instantiation in the form of a model-based decision support system that supports decision making in the field of financial market surveillance is presented. This artifact utilizes a qualitative decision model to identify situations in which prices of single stocks are affected by fraudsters who aggressively advertise the stock. An evaluation of the implemented system based on voluminous and heterogeneous data including user-generated content data is provided.

The empirical results indicate that the developed artifact can provide support for identifying such malicious situations. Given this evidence, it is apparent that the developed solution is able to utilize massive and heterogeneous data, especially user-generated content from financial blogs and news platforms, to provide practical decision support in the field of market surveillance.

In conclusion, the study documented financial management system DSS effectiveness and studied the role of market surveillance in improving market manipulation based on information. It illustrated how market surveillance service could support regulatory authority.

The results suggest that it is possible to utilize content generated by users in real time following DSS development. Moreover, this research demonstrated that the DSS were able to deal with heterogeneous and voluminous data, providing daily alerts by differentiating non-suspect and suspect market circumstances. The analysis provided adequate empirical evidence, suggesting that the developed artifact might be effective in the detection of real abusive incidences of pump-and-dump market manipulation.

## Paper 4: Explanatory Design Theory for Financial Market Surveillance Decision Support

The paper provides an explanatory design theory for financial market surveillance decision support systems (FMS-DSS) that incorporate structured and unstructured data and is enhanced to incorporate regulatory demands. The study employed action design research to develop general components and general requirements for an explanatory design theory. Moreover, the study provides a framework for designing and implementing real time and highly flexible surveillance systems appropriate for capital markets.

The objective of this paper was to provide design suggestions that enable effective development of FMS-DSS financial institutions. By shedding light on explanatory design theory applicable to financial markets and decision making by regulatory authority, the study aimed to contribute to market surveillance status. The study was guided by research questions aimed at determining general components and general requirements for financial market surveillance systems with the ability to detect a variety of market manipulations originating from social media usage.

The main conclusion of this paper was that FMS-DSS must have characteristics that are in line with the regulatory requirements. The paper presented its conclusion based on practical, cost and theoretical perspectives. From the practical viewpoint, the design theory represented by general components and requirements provides guidelines on market surveillance information system development. From the cost perspective, the study considered market surveillance as a major undertaking that consumes time. Theoretically, the investigation contributes significantly to financial market surveillance by highlighting explanatory design theory strategies for future studies. Thus, this paper contributes to market surveillance knowledge within the area of FMS–DSS.

The following table presents an overview of the proposed research stream, applied methodologies and achieved results.

Table 15: Paper contributions

Paper	Research Stream DSR/DSS (Banker &Kauffman, 2004)	Methodology	Results/Artifacts
Paper 1 State of the Art of financial Decision Support Systems	Design Science Research (DSR)	Literature Review as proposed by (Webster & Watson, 2002)	Analytical Framework for the literature review.  The analysis results confirm the applicability of developed framework and suggest directions for future research along examined categories.
Paper 2 Model to detect financial market manipulation	Design Science Research (DSR)	Process Model as proposed by (Vaishnavi & Kuechler, 2008)  Qualitative Multi-Attribute Modelling (Bohanec & Rajkovic, 1999)  Ex ante evaluation (Pries-Heje et al., 2008)	The qualitative multi-attribute model to assess decision alternatives.
Paper. 3 Evaluation of instantiated IT Artifact	Decision Support Systems (DSS)	Action Design Research as proposed by (Sein et al., 2011)  Empirical Evaluation  Ex post evaluation (Pries-Heje et al., 2008)	Evaluation of the instantiated IT Artifact (Financial Market Surveillance DSS);  Evaluation of the system's capability to detect relevant stock manipulation cases.
Paper 4 Explanatory Design Theory	Design Science Research (DSR)	Action Design Research (Sein et al., 2011)	Design recommendations as an Explanatory Design Theory for an Information System that supports market surveillance decision making.



# SECTION C: Integration of Findings

## 10 Contribution and Conclusion

### 10.1 Contribution to Theory

Paper 1 makes a contribution to theory by examining the evaluation process and proposing a framework that integrates elements of Decision Support Systems (DSS), design theory and information mining. The analysis reveals that more research is needed in the domain of fraud detection since very few relevant IS studies were found despite the importance of these topics due to increased turbulence in financial markets worldwide. Thus, paper 1, the first empirical article in this dissertation, analyzes the current state of the art of financial DSS by developing and applying an analytical framework that may also aid future researchers in this field structure their investigations.

The next empirical part of this dissertation makes a theoretical contribution by addressing the results of qualitative research. Based on interviews with problem owners, paper 2 extends the argument established in the first paper. Thus, it deals with the knowledge transfer and the development of key attributes for the detection of information-based manipulation. This paper demonstrated how the requirements are generated. Based on interviews with domain experts, we first identified the factors determining suspicious situations and then applied a qualitative multi-attribute modeling technique. Thereby, the resulting model builds upon valuable knowledge of domain experts and provides a means by which to address the challenge of information-based market manipulation.

The subsequent evaluation paper 3 builds upon the research results of paper 2. This evaluation paper draws on large-scale data to address the evaluation of an IT Artifact instantiation in the form of a model-based Decision Support System to support decision-making in the field of market surveillance. With a data architecture developed in the context of a collaborative EU-funded research project, an evaluation of the implemented system is presented based on voluminous and heterogeneously structured data, including social media data. Thus the developed artifact makes a theoretical contribution by utilizing a qualitative decision model developed previously and presented in paper 2 to identify situations in which the prices of single stocks are affected by fraudsters who aggressively advertise the stock. The empirical results indicate that the developed artifact can provide support for identifying such situations and provide practical decision support in the field of market surveillance.



Paper 4 brings the previous results together. It builds upon the research results from papers 2 and 3 and provides a formal representation of the developed design theory. Its main objective was to explain how to design efficient market surveillance systems by providing insights into needs of compliance officer. It analyzed the general components and general requirements by creating an innovative surveillance system. Thereby, it offers a theoretical explanation of how to develop corresponding systems. The research utilized the appropriate IS theory based on Baskerville and Pries-Heje (2010) to inform the research.

## 10.2 Contribution to Practice

The dissertation makes several noteworthy contributions to the field of financial market manipulation detection and prevention. The second paper offers the most important implication for the detection of financial manipulation. The paper addressed the increasingly important question of fraudulent behavior emerging on the Web 2.0 with respect to the manipulation of share prices. The purpose of the paper was to determine the requirements and components of information systems for information-based market manipulations, to pinpoint the importance of user-generated data and to develop an IT Artifact for the detection of financial market manipulation following the design science research paradigm. In this investigation, the aim was to assess decision support systems for information-based market manipulation in non-regulated markets. The evidence suggests that market manipulation is harmful. Hence, the dissertation reveals the relevant key indicators in a market surveillance context to tackle information-based market manipulation. With the negative effects of market manipulation cases in mind, the research can also aid in constructing the rationale for active efforts by market regulators to prohibit financial market manipulation. Lastly, the model developed in this paper can be used by practitioners and experts to develop and/or adjust their own manipulation detection systems.

## 10.3 Conclusion

This dissertation explains the central importance of detecting real time information during market abuse events using the latest financial applications to investigate ways to

strengthen financial market safety measures designed to detect and counteract market frauds.

Based on the research questions, the study aimed to determine the relevant key indicators in a market surveillance context required to tackle information-based market manipulation. Returning to the research questions posed at the beginning of this dissertation, it is now possible to state that the model for the identification of information-based manipulation can be developed by identifying and structuring attributes and defining attributes measurements and aggregate rubrics.

The literature review provided background information on the domain of fraud detection and revealed that very few IS studies have been performed. The provision of prior research on stock market manipulation lays the foundation for further research following the identification of gaps in the existing relevant literature. Examining prior research in risk management, financial analysis and fraud detection helped inform the type of DSS to develop for the appropriate detection of market manipulation in various financial settings.

The findings contribute to the current literature on market manipulation. This work contributes to the existing knowledge of financial market manipulation detection by providing a method for developing a model for the identification of information-based manipulation. The research has enhanced the understanding of design principles for IT artifact development.

The research provides additional evidence with respect to the use of software prototypes in providing support for decision making in financial market surveillance. Thus, this research extends our knowledge of developing IT software prototypes to support the identification of malicious activities in financial markets. Moreover, the study confirms previous findings and contributes additional evidence that suggests that the utilization of user-generated content on social media may help in supporting decisions made in the market surveillance domain. These findings enhance our understanding of the current state of financial decision support systems. The literature reviewed identified gaps in three major domains, including risk management, fraud detection and financial analysis. The information provided enhances our understanding of these areas and suggests areas that require improvements to ensure reduction in market manipulation. Furthermore, this research may serve as a base for future studies in the domains of financial market manipulation, fraudulent actions, risk management and financial analysis.

## 11 Discussion

### 11.1 Limitations

The following important study limitations need to be considered. First, the dissertation provides only general principles for artifact development; it does not specify information, technology and application principles. Secondly, the study did not evaluate the use of information as an asset or the easy accessibility of information in market manipulation.

### 11.2 Recommendations for Further Research Work

The findings of this dissertation have a number of important implications for future practice. One of the most important tasks of financial market regulatory authorities is to reduce market manipulation. Thus, manipulation detection ought to be an imperative theme in financial markets.

In paper 1 we suggested using fraud detection and risk management domains to build on expert knowledge and unstructured content increasingly generated by users. The study proposed future research in terms of problems and associated requirements, components and evaluation. While there were many studies in the field of financial analysis, there were few studies focused on fraud detection and risk management. Due to extensive literature on DSS for financial analysis, this study suggests computational intelligence research. Moreover, we recommend more studies in DSS in the risk management domain.

We found that most of the studies had excluded evaluations based on user views. Thus, we recommended the exploration and application of diverse evaluation techniques to obtain rich and useful responses from domain experts. This research has generated many questions in need of further investigation. In particular, the need for existing literature on decision support systems for detecting market manipulation became evident in paper 2. There were challenges in obtaining fraudulent cases from the market surveillance authorities, limiting information accessibility. More broadly, research is also

needed to determine if other experts are in agreement with the categorizations provided by the model.

### 11.3 Implications for Policy

While the market surveillance authorities, user involvement programs and market manipulation detection based on information are vital, policy makers should consider their objective of fraudulent reduction in financial markets. As such, there should be policies geared towards linking market surveillance authorities with DSS that foster the detection of market manipulations.

A manipulative trader is normally considered to be an individual independently acting to the damage of the clients and the institutions that have employed them. The damage inflicted by manipulative traders is not restricted to financial institutions; it affects the whole market since it undermines the credibility of financial markets and damages the trust built with investors. Thus, procedures and policies established by financial institutions must be robust as well as preventive to help detect manipulative trading. The policies should support the decision-making process in financial market surveillance via the use of artifacts. Surveillance market authorities can use DSS to reduce financial market manipulation through communication, guidance, regulation, the distribution of existing practices in the financial analysis domain and the provision of financial information. Furthermore, the introduction of extra policies to handle market abuse should explicitly highlight strategies for monitoring, preventing, detecting and responding appropriately to market abuses.



## References

- Abbasi, A., Sarker, S., & Chiang, R.H.L. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, 17(2), Article 3.
- Aggarwal, R. K., & Wu, G. (2003). Stock market manipulation - Theory and evidence. Working Paper, 1-47.
- Aggarwal, R., & Wu, G. (2006). Stock Market Manipulations. *The Journal of Business*, 79(4), 1915-1953.
- Allen, F., & Gale, D. (1992). Stock-price manipulation. *Review of Financial Studies*, 5(3), 503-529.
- Arasu, A., & Garcia-Molina, H. (2003). Extracting Structured Data from Web Pages. In proceedings of the Association for Computing Machinery Special Interest Group on Management of Data international Conference on Management of Data, San Diego, CA, 337-348.
- Arh, T., & Blažič, B. (2007). Application of Multi-Attribute Decision Making Approach to Learning Management Systems Evaluation. *Journal of Computers*, 2(10), 28-37.
- Arnott, D., & Pervan, G. (2012). Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines. *Journal of the Association for Information Systems*, 13(11), 923-949.
- Bagnoli, M., & Lipman, B. L. (1996). Stock price manipulation through takeover bids. *RAND Journal of Economics*, 27(1), 124-147.
- Baiyere, A., Hevner, A., Gregor, S., & Rossi, M. (2015). Artifact and or Theory? Publishing Design Science Research in IS. In proceedings of the International Conference on Information Systems, Fort Worth, 1-5.
- Banker, R. D., & Kauffman, R. J. (2004). The Evolution of Research on Information Systems: A Fiftieth-Year Survey of the Literature in Management Science. *Management Science*, 50(3), 281-298.
- Barnes, P. (2009). Chapter 6: The Regulation of the Markets. In *Stock Market Efficiency, Insider Dealing and Market Abuse*, Gower Publishing Ltd, 113-146.
- Barnhill, A. (2014). What is Manipulation? *Manipulation: Theory and Practice*, C. Coons and M. Weber (eds.), Oxford University Press, 51-72.
- Baskerville, R., & Myers, M. D. (2004). Special issue on action research in information systems: making is research relevant to practice - foreword. *Management In-*

- formation Systems Quarterly, 28(3), 329-335.
- Baskerville, R., & Pries-Heje, J. (2010). Explanatory Design Theory. *Business & Information Systems Engineering*, 2(5), 271-282.
- Baskerville, R., & Wood-Harper, A. T. (1998). Diversity in information systems action research methods. *European Journal of Information Systems*, 7(2), 90-107.
- Baskerville, R., & Wood-Harper, T. (1996). A critical perspective on action research as a method for information systems research. *Journal of Information Technology*, 11, 235-246.
- Bloom, K. (2011). Sentiment Analysis Based on Appraisal Theory and Functional Local Grammars. Dissertation thesis, Illinois Institute of Technology.
- Bohanec, M. (2003). Decision support. In D. Mladenic, N. Lavrac, M. Bohanec, & S. Moyle (eds.), *Data Mining and Decision Support: Integration and Collaboration*. Amsterdam: Kluwer Academic Publishers, 23-35.
- Bohanec, M., & Rajkovič, V. (1990). DEX : An Expert System Shell for Decision Support. *Sistemica*, 1(1), 145-157.
- Bohanec, M., & Zupan, B. (2004). A function-decomposition method for development of hierarchical multi-attribute decision models. *Decision Support Systems*, 36(3), 215-233.
- Bohanec, M., Žnidaršič, M., Rajkovič, V., Bratko, I., & Zupan, B. (2013). DEX Methodology: Three Decades of Qualitative Multi-Attribute Modeling. *Informatica*, 37, 49-54.
- Bouraoui, T. (2009). Stock spams: An empirical study on penny stock market. *Journal of Arts and Sciences*, 1(6), 196-208.
- Braeuer, S., Heddier, M., Knackstedt, R., & Platz, M. (2014). Integrating Regulatory Requirements into Information Systems Design and Implementation. In proceedings of the Thirtyfifth International Conference on Information Systems, Auckland, 1-12.
- Brandl, B., & Keber, C. (2006). An Automated Econometric Decision Support System: Forecasts for Foreign Exchange Trades. *Central European Journal of Operations Research*, 14(4), 401-415.
- Brocke, J., Simons, A., Niehaves, B., Riemer, K., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting. In proceedings of the Seventeenth European Conference on Information Systems, Verona, 1-13.
- Burger, C., & Du Plessis, M. C. (2011). Does Chomsky Complexity Affect Genetic Programming Computational Requirements? In proceedings of the South African Institute of Computer Scientists and Information Technologists Conference on Knowledge, Innovation and Leadership in a Diverse, Multidisciplinary Envi-

- ronment, Cape Town, 31-39.
- Buta, P., & Barletta, R. (1991). Case-based reasoning for market surveillance. In proceedings of the First International Conference on Artificial Intelligence Applications on Wall Street, New York, 116-121.
- Cao, L., & Ou, Y. (2008). Market Microstructure Patterns Powering Trading and Surveillance Agents. *Journal of Universal Computer Sciences*, 14(14), 2288-2308.
- Chan, S. W., & Franklin, J. (2011). A Text-based Decision Support System for Financial Sequence Prediction. *Decision Support Systems*, 2(1), 189-198.
- Chatterjee, S., (2015). Writing My next Design Science Research Master-piece: But How Do I Make a Theoretical Contribution to DSR? In proceedings of the Twentythird European Conference on Information Systems, Münster, Germany, Paper 28.
- Cumming, D. (2008). Global Market Surveillance. *American Law and Economics Review*, 10(2), 454-506.
- Cumming, D., Zhan, F., & Aitken, M. (2012). Exchange Trading Rules, Surveillance, and Insider Trading. Electronic copy available at: <http://ssrn.com/abstract=2101826>.
- de Moor, A. (2007). Using System Dynamics to Construct Design Theory for Community Information Systems. *The Journal Of Community Informatics*, 3(1), 1-24.
- Delort, J.-Y., Arunasalam, B., Leung, H., & Milosavljevic, M. (2011). The impact of manipulation in internet stock message boards. *International Journal of Banking and Finance*, 8(4), 1-18.
- Diaz, D., Zaki, M., Theodoulidis, B., & Sampaio, P. (2011). A Systematic Framework for the Analysis and Development of Financial Market Monitoring Systems. In proceedings of the Annual Service Research & Innovation Institute Global Conference, San Jose, CA, 145-153.
- Dibbern, J., Goles, T., Hirschheim, R., & Jayatilaka, B. (2004). Information Systems Outsourcing: A Survey and Analysis of the Literature. *Association for Computing Machinery Special Interest Group on Management and Information Systems Database*, 35(4), 6-102.
- Dunham, K. (2007). Pump and Dump Scams. *Information Systems Security*, 16(1), 65-71.
- Eren, N., & Ozsoylev, H. N. (2006). Hype and dump manipulation. *European Finance Association 2007 Ljubljana Meetings Paper*. Electronic copy available at SSRN: <https://ssrn.com/abstract=948814>
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press.



- Ferber, R. (2003). *Information Retrieval*. Heidelberg: dpunkt.Verlag.
- Frieder, L., & Zittrain, J. (2008). Spam works: Evidence from stock touts and corresponding market activity. *Hastings Communications and Entertainment Law Journal*, 479, 1-42.
- Geva, T., & Zahavi, J. (2010). Predicting Intraday Stock Returns by Integrating Market Data and Financial News Reports. In *proceedings of the Mediterranean Conference on Information Systems*, Tel Aviv, Paper 39.
- Gidófalvi, G., & Elkan, C. (2003). Using News Articles to Predict Stock Price Movements. Technical Report, Department of Computer Science and Engineering. University of California.
- Ginzberg, M. J., & Stohr, E. A. (1982). DSS: Issues and Perspectives. *NYU Working Paper*, 82(12), 1-42.
- Goldberg, H., Kirkland, J., Lee, D., Shyr, P., & Thakker, D. (2003). The NASD Securities Observation, New Analysis and Regulation System (SONAR). In *proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence*, Acapulco, 11-18
- Goldkuhl, G. (2004). Design Theories in Information Systems - A Need for Multi-Grounding. *Journal of Information Technology Theory and Application*, 6(2), 59-72.
- Goldstein, I., & Guembel, A. (2008). Manipulation and the allocational role of prices. *Review of Economic Studies*, 75(1), 133-164.
- Gopal, R., Marsden, J. R., & Vanthienen, J. (2011). Information Mining - Reflections on Recent Advancements and the Road Ahead in Data, Text, and Media Mining. *Decision Support Systems*, 51(4), 727-731.
- Gopal, R., Marsden, J. R., & Vanthienen, J. (2011). Information mining — Reflections on recent advancements and the road ahead in data, text, and media mining. *Decision Support Systems*, 51(4), 727-731.
- Grčar, M. (2012). FIRST Consortium D6.4 Ontology reuse and evolution; [http://www.project-first.eu/public\\_deliverables](http://www.project-first.eu/public_deliverables).
- Gregor, S., & Hevner, A. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *Management Information Systems Quarterly*, 37(2), 1-25.
- Groth, S. S., & Muntermann, J. (2011). An Intraday Market Risk Management Approach Based on Textual Analysis. *Decision Support Systems*, 50(4), 680-691.
- Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 4, 1-23.

- Hadasch, F., Mueller, B., & Maedche, A. (2012). Exploring Antecedent Environmental and Organizational Factors to User-caused Information Leaks: A Qualitative Study. In proceedings of the Twentieth European Conference on Information Systems, Barcelona, Paper 127.
- Hanke, M., & Hauser, F. (2008). On the effects of stock spam e-mails. *Journal of Financial Markets*, 11(1), 57-83.
- Heping, P. (2006). Multilevel Stochastic Dynamic Process Models and Possible Applications in Global Financial Market Analysis and Surveillance. In proceedings of the Joint Conference on Information Sciences Advances in Intelligent Systems Research, Kaohsiung, 1-15.
- Hevner, A. R., & Chatterjee, S. (2010). *Design research in information systems: Theory and Practice*. New York Dordrecht Heidelberg London: Springer Science+Business Media, LLC, 1-320.
- Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28(1), 75-105.
- Huang, K., & Li, Z. (2011). A Multilabel Text Classification Algorithm for Labeling Risk Factors in SEC Form 10-K. *Association for Computing Machinery Transactions on Management Information Systems*, 2(3), 1-19.
- Huang, M. L., Liang, J., & Nguyen, Q. V. (2009). A Visualization Approach for Frauds Detection in Financial Market. In proceedings of the Thirteenth International Conference Information Visualisation, Barcelona, 197-202.
- Huang, S.-H., Lai, S.-H., & Tai, S.-H. (2011). A Learning-Based Contrarian Trading Strategy via a Dual-Classifer Model. *Association for Computing Machinery Transactions on Intelligent Systems and Technology*, 2(3), 1-20.
- Huang, W., Nakamori, Y., & Wang, S.-Y. (2004). Forecasting Stock Market Movement Direction with Support Vector Machine. *Computers & Operations Research*, 32(10), 2513-2522.
- Huang, Y. C., & Cheng, Y. J. (2013). Stock manipulation and its effects: pump and dump versus stabilization. *Review of Quantitative Finance and Accounting*, 44(791), 1-25.
- Huang, Z., Hsinchun, C., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. *Decision Support Systems*, 37(4), 543-558.
- Iivari, J. (2015). Distinguishing and Contrasting Two Strategies for Design Science Research. *European Journal of Information Systems*, 24(1), 107-115.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining Techniques for the Detection of Fraudulent Financial Statements. *Expert Systems with Applica-*

- tions, 32(4), 995-1003.
- Klein, A., Altuntas, O., Häusser, T., & Kessler, W. (2011). Extracting Investor Sentiment from Weblog Texts: A Knowledge-based Approach. In proceedings of the Thirteenth Institute of Electrical and Electronics Engineers Conference on Commerce and Enterprise Computing, Luxembourg, 1-9.
- Klein, A., Altuntas, O., Riekert, M., & Dinev, V. (2013). A Combined Approach for Extracting Financial Instrument-Specific Investor Sentiment from Weblogs. In proceedings of the Eleventh International Conference on Wirtschaftsinformatik, Leipzig, 691-705.
- Kyle, A. S., & Viswanathan, S. (2008). How to define illegal price manipulation. *The American Economic Review*, 98(2), 274-279.
- Lee, S.-Y., Qiu, L., & Whinston, A. (2014). Manipulation: Online Platforms' Inescapable Fate. In proceedings of the Thirtyfifth International Conference on Information Systems, Auckland, 1-16.
- Linkov, I., Satterstrom, F. K., Kiker, G., Batchelor, C., Bridges, T., & Ferguson, E. (2006). From comparative risk assessment to multi-criteria decision analysis and adaptive management: recent developments and applications. *Environment International*, 32(8), 1072-93.
- Liu, Y., & Salvendy, G. (2007). Design and evaluation of visualization support to facilitate decision trees classification. *International Journal of Human-Computer Studies*, 65(2), 95-110.
- Lucas Jr, H. C. (1993). Market expert surveillance system. *Communications of the Association for Computing Machinery*, 36(12), 27-34.
- Mangkorntong, P., & Rabhi, F. (2007). A High-Level Approach for Defining & Composing Event Patterns and Its Application to E-Markets. In proceedings of the Second International Workshop on Event-driven Architecture, Processing and Systems at the Thirty Third International Conference on Very Large Data Bases, Vienna, 1-4.
- Mangkorntong, P., & Rabhi, F. a. (2008). A Domain-Driven Approach for Detecting Event Patterns in E-Markets. *World Wide Web*, 12(1), 69-86.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). Big data: The next frontier for innovation competition, and productivity. Report McKinsey Global Institute.
- Markus, L. M., Majchrzak, A., & Gasser, L. (2002). A Design Theory for Systems That Support Emergent Knowledge Processes. *Management Information Systems Quarterly*, 26(3), 179-212.
- Marshall, P., Willson, P., Salas, K. de, & McKay, J. (2010). Action research in prac-

- tice: issues and challenges in a financial services case study. *The Qualitative Report*, 15(1), 76-93.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data*. München: redline Verlag, Münchener Verlagsgruppe GmbH.
- Mei, J., Wu, G., & Zhou, C. (2004). Behavior Based Manipulation: Theory and Prosecution Evidence. *Working papers series*, (212), 0-49.
- Morton, M. S. S. (1971). *Management Decision Systems*. Boston: Graduate School of Business Administration, Harvard University.
- Mukherjee, A., Diwan, P., Bhattacharjee, P., Mukherjee, D., & Misra, P. (2010). Capital market surveillance using stream processing. In *proceedings of the Second International Conference on Computer Technology and Development*, Cairo, 577-582.
- Müller-Wienbergen, F., & Müller, O. (2011). Leaving the beaten tracks in creative work—A design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems*, 12(11), 714-740.
- Muntermann, J. (2009). Towards Ubiquitous Information Supply for Individual Investors: A Decision Support System Design. *Decision Support Systems*, 47(2), 82-92.
- Ou, Y., Cao, L., Yu, T., & Zhang, C. (2007). Detecting Turning Points of Trading Price and Return Volatility for Market Surveillance Agents. In *proceedings of the International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*, Silicon Valley, 491-494.
- Park, B.-K., & Song, I.-Y. (2011). Toward total business intelligence incorporating structured and unstructured data. In *proceedings of the Second International Workshop on Business intelligence*, Uppsala, 12-19.
- Peppers, K., Tuunanen, T., Rothenberger, M. a., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77.
- Peramunetilleke, D., & Wong, R. K. (2001). Currency Exchange Rate Forecasting from News Headlines. *Australian Computer Science Communications*, 24(2), 131-139.
- Power, D. J. (2004). Specifying an Expanded Framework for Classifying and Describing Decision Support Systems. *Communications of the Association for Information Systems*, 13(1), 158-166.
- Power, D.J. (2001). Supporting Decision-Makers: An Expanded Framework. In *Proceedings of the Informing Science Conference*, Krakow, 431-436.
- Project FIRST. (2013). Dissemination Material:

- [http://www.projectfirst.eu/public\\_deliverables](http://www.projectfirst.eu/public_deliverables).
- Pui, G., Fung, C., Yu, J. X., & Lu, H. (2005). The Predicting Power of Textual Information on Financial Markets. *Institute of Electrical and Electronics Engineers Intelligent Informatics Bulletin*, 5(1), 110.
- Putniņš, T. J. (2012). Market Manipulation: a Survey. *Journal of Economic Surveys*, 26(5), 952-967.
- Rockness, J., & Williams, P. F. (1988). A Descriptive Study of Social Responsibility Mutual Funds. *Accounting, Organizations and Society*, 13(4), 397-411.
- Ruggieri, S., Pedreschi, D., & Turini, F. (2010). Data Mining for Discrimination Discovery. *Association for Computing Machinery Transactions on Knowledge Discovery from Data*, 4(2), 1-40.
- Saveski, M., & Grcar, M. (2011). Web Services for Stream Mining: A Stream-Based Active Learning Use Case. In *proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge discovery in Databases*, Athens, 1-11.
- Schumaker, R. P., & Chen, H. (2009). A Quantitative Stock Prediction System Based on Financial News. *Information Processing & Management*, 45(5), 571-583.
- SEC, G. (2012). SEC Suspends Trading in Common Stock of Three Hundred Seventy-Nine Companies Quoted on OTC May 14 , 2012 (No . 66980). SEC (Vol. May).
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *Management Information Systems Quarterly*, 35(1), 37-56.
- Sinha, A. P., & Zhao, H. (2008). Incorporating Domain Knowledge into Data Mining Classifiers: An Application in Indirect Lending. *Decision Support Systems*, 46(1), 287-299.
- Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2013). Predictive Sentiment Analysis of Tweets: A Stock Market Application. *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. Lecture Notes in Computer Science*, Springer-Verlag Berlin Heidelberg, 77-88.
- Smailović, J., Žnidaršič, M., & Grčar, M. (2012). Web-based experimental platform for sentiment analysis. In *proceedings of the Third International Conference on Information Society and Information Technologies*, Coimbatore, 1-6.
- Sprague, R. H. (1980). A Framework for the Development of Decision Support System. *Management Information Systems Quarterly*, 4(4), 1-26.
- Sunstein, C. (2015). Fifty shades of manipulation. Electronic copy available at: SSRN 2565892, 1-32.
- Tay, F. E., & Cao, L. J. (2001). Improved Financial Time Series Forecasting by Com-

- binning Support Vector Machines with Self-organizing Feature Map. *Intelligent Data Analysis*, 5(4), 339-354.
- Tsai, C.-F., & Hsiao, Y.-C. (2010). Combining Multiple Feature Selection Methods for Stock Prediction: Union, Intersection, and Multi-intersection Approaches. *Decision Support Systems*, 50(1), 258-269.
- Turban, E., Sharda, R., & Delen, D. (2010). *Decision Support and Business Intelligence Systems*. New Prentice Hall (9th edition), 1-720.
- Vaishnavi, V. K. & Kuechler, W. (2008). *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*. Boston: Auerbach Publications, Taylor & Francis Group.
- Van Bommel, J. (2003). Rumors. *The Journal of Finance*, 58(4), 1499-1520.
- Van de Ven, A. H. (2007). *Engaged Scholarship: Creating Knowledge for Science and Practice*. New York: Oxford University Press.
- Vapnik, V., & Chervonenkis, A. (1974). *Teoriya raspoznavaniya obrazov: Statisticheskie problemy obucheniya*. (In Russian). [Theory of pattern recognition: Statistical problems of learning]. Moscow: Nauka, 1-416.
- Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, 3(1), 36-59.
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Management Information Systems Quarterly*, 26(2), 13-23.
- Wiener, M., Vogel, B., & Amberg, M. (2010). Information Systems Offshoring - A Literature Review and Analysis. *Communications of the Association for Information Systems*, 27(1), 455-492.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition. San Francisco: Morgan Kaufmann.
- Wüthrich, B., Leung, C. B., Permuntilleke, D., Sankaran, K., Zhang, J., & Lam, W. (1998). Daily Stock Market Forecast from Textual Web Data. In proceedings of the Institute of Electrical and Electronics Engineers International Conference on Systems, Man, and Cybernetics, San Diego, 2720-2725.
- Zaki, M. (2013). *An Ontological Approach for Monitoring and Surveillance Systems in Unregulated Markets*. Dissertation thesis, The University of Manchester, 1-229.
- Zaki, M., Diaz, D., & Theodoulidis, B. (2012). Financial Market Service Architectures: A "Pump and Dump" Case Study. In proceedings of Annual Service Research & Innovation Institute Global Conference, San Jose, 554-563.
- Zaki, M., Theodoulidis, B., & Solis, D. D. (2011). A Data Mining Approach for the

Analysis of “Stock-Touting” Spam Emails. *Journal of Manufacturing Technology Management*, 22(6), 70-79.

Zhang, W., & Skiena, S. (2010). Trading Strategies to Exploit Blog and News Sentiment. In proceedings of the Fourth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media, Washington DC, 1-8.

Zülch, M. J., Alic, I., & Muntermann, J. (2011). How ‘good’ is bad News? Exploring Sentiments of Corporate Disclosures. In proceedings of the Seventeenth Americas Conference on Information Systems, Detroit, Paper 221.

## Overview of author contribution in included publications

Table 16: Author contribution

Paper #	Title	Author	Approx, author contribution in percent
1	State of the Art of Financial Decision Support Systems based on Problem, Requirement, Component and Evaluation Categories	<b>Irina Alić</b>	<b>70</b>
		Jan Muntermann	20
		Robert W. Gregory	10
2	Hot Stock or Not? A Qualitative Multi-Attribute Model to Detect Financial Market Manipulation	<b>Irina Alić</b>	<b>75</b>
		Michael Siering	20
		Marko Bohanec	5
3	Supporting Financial Market Surveillance: An IT Artifact Evaluation	<b>Irina Alić</b>	<b>100</b>
4	Financial Market Surveillance Decision Support: An Explanatory Design Theory	<b>Irina Alić</b>	<b>100</b>



## **Promotionsstudiengang „Wirtschaftswissenschaften“**

### **Versicherung bei Zulassung zur Promotionsprüfung**

Ich versichere,

1. dass ich die eingereichte Dissertation

#### **Decision Support Systems for Financial Market Surveillance**

selbstständig angefertigt habe und nicht die Hilfe Dritter in einer dem Prüfungsrecht und wissenschaftlicher Redlichkeit widersprechenden Weise in Anspruch genommen habe,

2. dass ich das Prüfungsrecht einschließlich der wissenschaftlichen Redlichkeit – hierzu gehört die strikte Beachtung des Zitiergebots, so dass die Übernahme fremden Gedankenguts in der Dissertation deutlich gekennzeichnet ist – beachtet habe,
3. dass beim vorliegenden Promotionsverfahren kein Vermittler gegen Entgelt eingeschaltet worden ist sowie im Zusammenhang mit dem Promotionsverfahren und seiner Vorbereitung
  - kein Entgelt gezahlt oder entgeltgleiche Leistungen erbracht worden sind
  - keine Dienste unentgeltlich in Anspruch genommen wurden, die dem Sinn und Zweck eines Prüfungsverfahrens widersprechen
4. dass ich eine entsprechende Promotion nicht anderweitig beantragt und hierbei die eingereichte Dissertation oder Teile daraus vorgelegt habe.

Mir ist bekannt, dass Unwahrheiten hinsichtlich der vorstehenden Versicherung die Zulassung zur Promotionsprüfung ausschließen und im Falle eines späteren Bekanntwerdens die Promotionsprüfung für ungültig erklärt werden oder der Doktorgrad aberkannt werden kann.

15 August, 2016 Irina Alic