Dennis Kundisch
Daniel J. Veit
Tim Weitzel
Christof Weinhardt (Eds.)

Enterprise Applications and Services in the Finance Industry

LNBIP 23

4th International Workshop, FinanceCom 2008
Paris, France, December 2008
Revised Papers

Springer

Lecture Notes
in Business Information Processing 23

Series Editors

Dennis Kundisch Daniel J. Veit
Tim Weitzel Christof Weinhardt (Eds.)

Enterprise Applications and Services in the Finance Industry

4th International Workshop, FinanceCom 2008
Paris, France, December 13, 2008
Revised Papers

Springer

Volume Editors

Dennis Kundisch
University of Paderborn
33098 Paderborn, Germany
E-mail: dennis.kundisch@wiwi.uni-paderborn.de

Daniel J. Veit
University of Mannheim
68131 Mannheim, Germany
E-mail: veit@bwl.uni-mannheim.de

Tim Weitzel
University of Bamberg
96045 Bamberg, Germany
E-mail: tim.weitzel@uni-bamberg.de

Christof Weinhardt
University of Karlsruhe (TH)
76131 Karlsruhe, Germany
E-mail: weinhardt@iism.uni-karlsruhe.de

Library of Congress Control Number: 2009929372

ACM Computing Classification (1998): J.1, K.5.2, K.4.4, H.3.5

ISSN 1865-1348
ISBN-10 3-642-01196-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-01196-2 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12655533 06/3180 5 4 3 2 1 0

Preface

2008 was a disastrous year for the financial services industry. Like an earthquake that measured 10+ on the Richter scale, the industry and the markets were globally severely shaken—many financial institutions went bankrupt, were taken over or made use of massive state guarantees and equity injections. Governments all over, the globe announced bank rescue packages in staggering heights. So, the traditional investment banking type of market structure and its dominance is gone—what's next?

In these rough months it has become obvious how important a functioning financial system is in our globalized world with interrelated markets and industries. Unsurprisingly these eruptions have also hit the real economy and many countries are currently in a deep recession all over the globe. However, as a proverb has it: "If nothing is secure, everything is possible". Maybe the current status of the financial markets allows for the introduction of better rules and regulations and is the starting point for a different sustainable development of financial systems and economies. Providing for new perspectives and insights, constructive but critical analysis, while being on the quest for truth is a primary task of research. Especially in times when policy and decision makers are looking for guidance, it is important that research provides for rigorous and at the same time relevant results.

This is also one objective of the workshop series on *enterprise applications and services in the finance industry* (FinanceCom). After three very successful workshops in Sydney (Australia), Regensburg (Germany, co-located with ECIS 2005), and Montreal (Canada, co-located with ICIS 2007), FinanceCom 2008 was held on December 13, 2008 in Paris in co-location with ICIS 2008. The workshop spans multiple disciplines, including technical, economic, sociological and behavioral sciences. It reflects on technologically enabled opportunities, implications and changes due to the introduction of new business models or regulations related to the financial services industry and the financial markets. The guiding theme of this workshop was concerned with innovations in the financial services industry. For too long, many industry participants have focused on product innovations with apparent results. Process innovations have been neglected but seem promising when it comes to the facilitation of economic growth in the financial services industry.

Therefore, *innovation* was also the topic of the invited keynote contribution by Maurice Peat from the Securities Industry Research Centre of Asia-Pacific (SIRCA). He suggested an agenda for the development of financial computing, which will support a broad view of financial innovation. This agenda involves both the support and enhancement of standards setting exercises and the adoption of service-oriented architecture (SOA) as the platform for a wide range of process innovations. The keynote at the workshop set the stage for the program

of the rest of the workshop day, which focused on three distinct but interrelated areas: (1) "Financial Markets and Customers," (2) "SOA" and (3) "Regulation and Compliance." The contents of these proceedings follow the workshop's structure.

In the first part—after the keynote contribution—we have three contributions concerned with financial markets and the customers in these markets. First, Andreas Storkenmaier and Ryan Riordan look at the effects of the introduction of the NYSE Hybrid Market on market quality. Using an event study approach, the authors find evidence that trading costs were reduced and execution quality improved at the NYSE. Moreover, the importance of specialists and floor brokers was further downgraded. Second, Matthias Burghardt and Ryan Riordan study retail investor sentiment. They analyze the order flow based on a unique data set with 20.7 million transactions in bank-issued warrants from the European Warrant Exchange. They propose the construction of a retail investor sentiment index and find that retail investors are contrarian, that retail investor sentiment is an important part of the equity pricing process and that the proposed index is a good measure of the sentiment. Third, Dick Heinhuis and Erik J. de Vries propose a conceptual model that aims at explaining multi-channel customer behavior. They take the technology acceptance model as a starting point and extent their model with insights from expectation disconfirmation theory and customer choice theory.

In the second part of these proceedings, the focus is on SOA. First, Fethi Rabhi, Omer Rana, Adnene Guabtni, and Boualem Benatallah present a user-driven environment for financial market data analysis. Using a case study approach related to processing both news and financial market data, the authors introduce a software development environment which facilitates the analysis of large financial datasets by end-users. They also describe a prototype implementation that allows domain experts to compose components and services to build an application. Second, Daniel Beimborn and Nils Joachim try to get a hand on the determination of the business value of SOA and how it can be achieved. In their paper they present a conceptual model, which focuses on the impact of IT business alignment on the successful implementation of SOA, in terms of its business value. The findings suggest that the business strategy moderates the impact of SOA's general potentials on its actual business value and that this relationship is further moderated by IT business alignment, which must be thoroughly considered by practitioners deciding on introducing SOA in their firm. Third, Haresh Luthria and Fethi Rabhi focus on the same issue but take a different perspective. In their contribution, they empirically examine the decision to adopt service-oriented computing as an enterprise strategy across 15 firms, and investigate the business drivers that influence this decision.

In the third part, the focus is on regulation and compliance issues. First, Jan-Helge Deutscher and Carsten Felden present a metamodel that allows constructions of specific model instances to support the measurement of successfully tracking IT governance-related objectives. This aim is accomplished by a model-based support of the operationalization and the analysis of defined objectives

in the manner of determining compliance degrees for companies. Consequently, Peter Gomber, Gregor Pujol, and Adrian Wranik look at how German financial institutions and online brokers fulfill the best execution requirements according to the MiFID. In their empirical investigation they conclude that although the minimum legal requirements have recognizably been implemented in nearly all policies, there is substantial heterogeneity between the policies of various investment firms. In the existing studies from the time before the MiFID was applicable, the best execution principles are most frequently named as a key differentiator or competitive factor. However, it turns out that the use of the policies as a competitive instrument cannot at present be recognized in a large majority of German financial institutions. The proceedings are concluded with a contribution by Kathrin Braunwarth, Hans Ulrich Buhl, Marcus Kaiser, Alexander Krammer, Maximilian Röglinger, and Alexander Wehrmann that looks at the EU Insurance Mediation Directive from a data quality perspective. The authors argue that the directive offers an opportunity to enhance the overall data quality of customer data and exemplify their arguments with a set of scenarios from the field of customer relationship management.

Besides the paper contributions, there were many interesting discussions that revolved around the financial crisis. It is still an open issue in this context, whether IT has had a neutral, amplifying, or moderating effect on the crisis. This is an issue that could not be addressed in the contributions at the workshop, since they were prepared largely before the financial crisis really started. However, this will surely be of interest in follow-up FinanceCom Workshops in the coming years. The next one is scheduled for 2010. Please refer to www.financecom.org to keep up to date.

Many people were involved in the preparation of this workshop. First of all I would like to thank John Barr, Jochen Dzienziol, Torsten Eymann, Michael Grebe, Terry Hendershott, Carsten Holtmann, Steffen Krotsch, Sven Laumer, Marco Marabelli, Daniel Minoli, Jan Muntermann, Dirk Neumann, Fethi Rabhi, Omer Rana, Stefan Sackmann, Matthias Tomann, Loredana Ureche-Rangau, Daniel Veit, Henning Weltzien, and Tim Weitzel for preparing more than 40 reviews within a tight time schedule. Based on these reviews, nine full contributions were selected for publication (as revised versions) in this volume.

Special thanks go to Andrea Carugati, who helped us a great deal in finding the right location in Paris. Moreover, we are indebted to Sophie Guiroy and her team at ESCP-EAP Campus Paris. Food, beverages, signposting, infrastructure—everything was perfectly organized and it was very nice to be co-located with six other pre-ICIS workshops. Thanks go also to Lisa Rucker and Lise Fitzpatrick who were, amongst others, in charge of the registration system. For the substantial support in taking care of the website and conference review system, I would like to empathically thank Georg Buss. Moreover, I would like to thank Nils Joachim for the preparation of the Workshop CDs and Stefan Luckner for assisting in the preparation of the local organization.

Finally, I would like to thank Ralf Gerstner from Springer for his excellent support in producing this proceedings volume. Last but not least I would like

to thank the Organizing Committee Daniel Veit, Tim Weitzel, and Christof Weinhardt for the great cooperation. It was really a pleasure working with you!

January 2009 Dennis Kundisch

Organization

FinanceCom 2008 was organized by the Guest Professorship Information Systems, Value Management & E-Business, Brandenburg University of Technology, Germany. It was technically co-sponsored by the IEEE Computing Society. Financial support by Accenture was greatly appreciated.

The workshop was carried out at the International Conference of Information Systems 2008 (ICIS) on December 13, 2008, in Paris, France.

Organizing Committee

Conference Chair	Dennis Kundisch, Brandenburg University of Technology, Germany
Program Chair	Daniel Veit, University of Mannheim, Germany
	Tim Weitzel, University of Bamberg, Germany
Organizing Chair	Christof Weinhardt, University of Karlsruhe, Germany

Program Committee

John Barr	Research Director, Financial Markets and Head of EU Research, UK
Dieter Bartmann	University of Regensburg, Germany
Mike Briers	SIRCA, Australia
Andrea Carugati	University of Aarhus, Denmark
Jochen Dzienziol	finalix, Switzerland
Torsten Eymann	University of Bayreuth, Germany
Michael Grebe	BCG, Germany
Terry Hendershott	University of California at Berkeley, USA
Carsten Holtmann	FZI, Germany
Steffen Krotsch	Dresdner Bank, Germany
Daniel Minoli	SES AMERICOM, USA and EntNet Technical Committee
Marco Marabelli	Catholic University of Milan, Italy and Bentley College, USA
Jan Muntermann	University of Frankfurt, Germany
Dirk Neumann	University of Freiburg, Germany
Fethi Rabhi	University of New South Wales, Australia
Omer Rana	Cardiff University, UK
Robert Schwartz	The City University of New York, USA
Stefan Sackmann	University of Freiburg, Germany
Junjie Sun	Office of the Comptroller of the Currency, USA

Matthias Tomann	Senacor Technologies AG, Germany
Loredana Ureche-Rangau	Université Catholique de Lille, France
Bruce W. Weber	London Business School, UK
Henning Weltzien	x-markets GmbH, Germany

Referees

J. Barr	D. Kundisch	O. Rana
J. Dzienziol	S. Laumer	S. Sackmann
T. Eymann	M. Marabelli	M. Tomann
M. Grebe	D. Minoli	L. Ureche-Rangau
T. Hendershott	J. Muntermann	D. Veit
C. Holtmann	D. Neumann	H. Weltzien
S. Krotsch	F. Rabhi	T. Weitzel

Sponsoring Institutions

Accenture, Germany
IEEE Computing Society (technical co-sponsorship), USA

Steering Committee for the FinanceCom-Workshop Series

Peter Gomber	University of Frankfurt, Germany
Dennis Kundisch	Brandenburg University of Technology, Germany
Fethi Rabhi	University of New South Wales, Australia
Federico Rajola	Catholic University of Milan, Italy
Daniel Veit	University of Mannheim, Germany
Christof Weinhardt	University of Karlsruhe, Germany
Tim Weitzel	University of Bamberg, Germany

Table of Contents

Keynote

Markets and Customers

Service-Oriented Architectures

Compliance and Regulation

Data + Information Systems = Financial Innovation

Maurice Peat[1,2]

[1] Finance Discipline, University of Sydney and
[2] Securities Industry Research Centre of Asia-Pacific
m.peat@econ.usyd.edu.au

Abstract. Finance and Computing are linked in a symbiotic relationship, in this paper this relationship and its implications for both finance and computing will be explored. The outcome of the exploration is a suggested agenda for the development of financial computing, which will support a broad view of financial innovation. This agenda involves the support and enhancement of standards setting exercises and the adoption of services computing as the approach for a wide range of process innovation.

Keywords: Financial Innovation, Standards, Services Computing, Interoperability.

1 Introduction

Finance and Computing are linked in a symbiotic relationship, in this paper this relationship and its implications for both finance and computing will be explored. The outcome of the exploration is a proposed agenda for the development of financial computing, which is capable of supporting and enhancing a broad view of financial innovation.

There are numerous examples which show the close interrelationship of finance and computing. The introduction of new financial instruments requires an associated development of computing systems; to compute the prices of the new instrument, to record transactions involving the instrument and to support settlement of the instrument at its termination. The introduction of traded option contracts in the 1970's is a typical example of this interrelationship, market participants created new systems to support the new financial product.

There are also cases where developments in computing and communications technology have facilitated developments in financial services. The introduction of microcomputers and increases in network capacity was a necessary prerequisite to the development and widespread deployment of Automated Teller Machines (ATM) and Point of Sale (POS) terminals. These developments have fundamentally changed the payments system and the economics of banking.

Both these examples show that the interrelationship between finance and computing is an important driver of innovation in financial the financial system. To understand how financial computing can continue to drive financial innovation

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 1–10, 2009.

we will investigate the nature of the financial system and innovation in this system. We will then point to areas of research in computing that are important in supporting the future of financial innovation.

2 Financial Innovation

The financial system is not static, is constantly adapting to changing circumstances and the institutions within the system are also adapting to the environment while seeking to maximise their value. Innovation plays an important part in this process of adaptation to change; successful innovators are able to take advantage of the new environment to grow at a faster pace than their competitors. Given the need to innovate and the benefits that flow from innovation an important question is: what might future innovation look like and what will be required to underpin the process of innovation?

To address this question we will go back to first principles starting with the basic functions of the financial system. Merton [1] described a functional decomposition that identifies six functions delivered by financial systems:

1. moving funds across time and space;
2. the pooling of funds;
3. managing risk;
4. extracting information to support decision-making;
5. addressing moral hazard and asymmetric information problems; and
6. Facilitating the sale or purchase of goods and services through a payment system.

The institutional structure that supports these functions is dependent on history and culture. In some cases it will be largely market based, in others it will be a combination of intermediary institutions and markets.

Broadly speaking, financial innovation is the act of creating and then popularizing new financial instruments as well as new financial technologies, institutions and markets. Financial institutions engage in innovation to gain the following benefits: increased risk sharing opportunities, avoidance of regulations and taxes, reduced transaction costs and increased liquidity, reduced agency costs, capturing temporary profits and changing prices.

The "innovations" are generally divided into product or process innovation, product innovations are exemplified by new derivative contracts, new corporate securities or new forms of pooled investment products, process improvements are typified by new means of distributing securities, processing transactions, or pricing transactions. In practice, even this simple differentiation is not clear, as process and product innovation are often linked, see Tufano,[4].

Product innovation is the prevalent form of innovation; institutions are constantly developing new products and marketing them to clients; in most cases they are required to gain approval from regulators for the new product. Both these activities expose the detail of the new products to competitors and other market participants. This visibility which is inherent in product innovation is its

weakness, other market participants are able to offer an equivalent, or enhanced product within a short time frame. It is estimated [2] that globally institutions spend $11 BN per year on product innovation, which is shown to have a marginal effect on profitability and market share.

On the other hand process innovation, where financial services companies seek efficiencies in their internal operations, is seen by consultants such as Deloitte as the path to profitable innovation in the financial services sector. Pursuit of this type of innovation exposes all aspects of the operations of a financial services organisation to a process of analysis and change whose objective is the optimisation of business processes, leading to a minimisation of operating costs for a given level of service. Process innovation is the area where the methods and tools of financial computing can be successfully applied to improve profitability.

The example of the introduction of the ATM will be expanded to demonstrate the linkage between financial computing and innovation and areas that will be important in promoting innovation with some development. The ATM was an innovation in the payments system, which literally changed the way that funds are moved across time and space. Cash became available at a large number of locations around the clock, leading to a fundamental change in consumer behaviour. This is an example of an innovation that is a product and process innovation, bank customers were offered a new product, self service banking. To support this product the banks back office transaction processing systems were modified to support the security needs and increased transaction volumes generated by the ATM network.

The first wave of ATM's were owned by and exclusively operated by a single financial institution. To access funds you needed to use a terminal owned by your institution. These machines allow users to: withdraw cash, make deposits, view account balances and transfer funds between linked accounts. The economics of network ownership, and customer demand lead to the interconnection of ATM networks, leading to the current situation where any bank customer can transact at any ATM which is part of a network that the bank is a member of (charges for using ATM's of other banks depend on regulatory arrangements and the competitive environment).

This innovation typifies the interaction between financial innovation and financial computing. The development of ATM networks was based on a mix of information systems and communications technologies, the wide spread availability of microprocessors and packet networking. The systems are based on transaction processing, a common feature across the financial system. There is a need for the transactions to be secure. The interconnection for ATM networks was a driver in the development of standards, for example the ISO 8583 Standard for Financial Transaction Card Originated Messages.

The ATM network was the precursor for further innovation in the payments system such as Point of Sale terminals in retail outlets and the development of self service banking through internet banking.

3 Financial Computing in Support of Innovation

Let's concentrate on process innovation in financial services. The primary activity in financial services is the processing of various transactions. The sequences of steps that are required to process a transaction describe the business process that is associated with the transaction type. For example a transfer of funds between two accounts is a transaction. The objective in process innovation is to minimise costs per transaction, by optimising existing business process, creating a new process or outsourcing the processing.

Any attempt to create a new business process or to optimise an existing process will begin with a decomposition of the process into a sequence of discrete steps. Then decisions on how to group the processing steps in the implementation to achieve an efficient process will follow. When an institution is analysing a number of processes, the decomposition process will often lead to the discovery of processing steps that appear in the processing of a number of transaction types. These common steps provide the opportunity for the realisation of synergy benefits. If the common process steps can be implemented as components, in a way that facilitates reuse of the components then cost saving synergies will be realised.

The business process optimisation approach to financial innovation is driven by component construction and the reuse of these components in the processing of multiple transaction classes. A simple form of component reuse, the reuse of code libraries in multiple projects, is common practice in systems development. These libraries can be created within the organisation or purchased from third parties.

What areas of information systems research will assist a financial services company looking for process innovations? The argument presented suggests that innovative companies will be interested the modelling of business processes leading to the identification of process steps that are candidates for conversion into components. They will also be interested in minimising the costs of building or buying components and linking them together to process transactions.

Components need to have the property of interoperability; transaction processing requires that the systems of the companies involved in the processing of a transaction be able to work together in a secure and timely manner.

3.1 Interoperability

Interoperability is concerned with the ability of diverse systems and organisations to work together. This idea can be defined in a number of ways. In a technical sense the IEEE defines interoperability as the ability of two or more systems or components to exchange information and to use the information that has been exchanged. According to ISO/IEC 2382-01, Information Technology Vocabulary, Fundamental Terms, interoperability is defined as follows: "The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units". A broader definition sates that to

be interoperable, one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organisation are managed in such a way as to maximise opportunities for exchange and reuse of information, whether internally or externally.

In information systems, technical interoperability at its most basic is the ability to exchange data between different programs using an agreed set of exchange formats, file formats, and protocols. A more demanding form of interoperability, known as semantic interoperability, is defined as the ability of two computer systems to exchange information and have the meaning of the information received by each of the systems automatically interpreted and processed into a useful result as understood by the users of the communicating systems. For this type of interoperability to exist a common information exchange reference model must be agreed on by all parties involved.

Any program of process innovation in the area of financial services will involve the development of semantically interoperable systems, capable of dealing with a high degree of complexity in the specification of transaction classes and their processing requirements. Trade automation systems are an example of a class of developments requiring semantic interoperability; information about the order flow from various exchanges is collected and analysed to determine the nature of the order to be placed. This might be a decision on trade venue, asset to trade or the size of the order to be placed depending on the automation system being developed. The order that has been generated then has to be transmitted to a trading venue and the outcome of the trade has to be received from the exchange and recorded. For such a system to work all of the systems involved need to have an agreed message format. It is possible for each trading venue to have a unique message format, leading to a complex system being necessary to achieve interoperability. Where standardised message formats have been defined interoperability is easier to achieve and cost effective.

Going forward research and development in the creation of modelling tools and development methodologies based on these modelling tools will be necessary to support the degree of interoperability required for effective process innovation in financial services. The development and use of standards, to elicit consensus views on the structure of the transactions to be processed is another important supporting mechanism for the use of interoperable systems for financial innovation.

3.2 Standards and Mark-Up Languages

An activity which will facilitate the construction of reusable components is the definition of standards which describe the information required to complete the processing of a transaction, Sexton [3]. When the information describing a transaction is standardised is possible to construct components which know the form of the input they will receive and pass on the next component in the process. The adoption of standardised transaction descriptors allows third parties to develop generic components which can be supplied to multiple financial services organisations, who in turn can then use the component in a number of business processes.

There is a movement to standardisation in the transaction based financial services area. A number of standards are being developed under the auspice of the International Standards Organisation (ISO); these efforts are addressing messaging and descriptions of financial instruments. Another group of industry backed standards are being deployed as Extended Mark-up Languages (XML's); these efforts are largely related to business processes undertaken by financial services organisations. We will use the sequence of business processes, and standards that support these processes, required to support a trade automation system development to organise the presentation of the available standards.

A trade automation system would involve a sequence of four businesses processes. The first involves the acquisition and management of market information, which is then used by a decision support system to generate a trade recommendation. The recommended trade is then executed. After execution a settlement process transfers funds and ownership. The final process is the generation of business reports for internal and external consumption. A number of standards have emerged which can be used in this sequence of processes. They include RIXML, NewsML and MDDL - supporting the Information aggregation process and the decision making processes; FIX and FpML - supporting trade execution; ISO 15022 - supporting the settlement process; and XBRML supporting the corporate reporting process.

RIXML is an XML taxonomy for the representation of investment research information. It provides a structure for the publication of instrument and industry sector level data which focuses on metadata rather than on the way that research reports are structured. NewsML is an XML standard designed to provide a structural framework for the dissemination of multimedia news. Each news item may contain photos, graphics, video clips and the same text in different languages. While NewsML provides rich facilities for packaging news and adding metadata, it does not define the formats of the actual content. The Market Data Definition Language (MDDL) supports the publication of snapshots and historical time series of equity prices, financial indices, and mutual fund data. It can support the acquisition, dissemination and management requirements of reference data, including pricing, depth of market, trade reports and end-of-day close information. It supports the Unique Instrument Identification and it is designed to allow it to be extended with ease.

FIX supports the trade execution process. It is a set of message formats (in both XML and non-XML), and a session layer that allows buy and sell side organisations and transaction venues to transact effectively. FpML (Financial products Mark-up Language) is the business information exchange standard for electronic dealing and processing of over the counter financial derivatives instruments and structured products. FpML messages each have a known and predefined set of defaults which are be overridden explicitly by either party as required. This makes the process of not specifying something explicitly a well defined one, allowing confirmation to be done in a fast semiautomatic fashion, as only mismatched information is manually verified.

ISO 15022 is a non-XML standard that specifies a set of messages to support the settlement process. It provides a standard set of data fields for financial information and messages for financial transactions. These messages are transported over the SWIFT network infrastructure, which provides a secure messaging backbone for the interaction of industry participants. It incorporates and is compatible with the earlier securities message standards ISO 7775 and ISO 11521.

Extensible business reporting language (XBRL) is a financial specification which is focused on company filings and reports it supports the financial reporting needs of organisations, including the acquisition and dissemination of data to/from the General Ledger. It is a single language that provides the basis for the creation of taxonomies to meet business requirements. Each accounting standard requires a different XBRL "taxonomy". What differs in each case are: the list of defined accounting items and the rules on how lower level items are added/subtracted/multiplied to give higher level items. Taxonomies may also be geographically based, allowing it to meet the needs of the specific market in which the business entity exists. XBRL allows companies to add their own items by extending their local XBRL taxonomy. This allows those items which are important to a company's understanding of its own business to be directly related to the standard accounting terms required in its annual filing and report.

ISO 20022, UNIFI (UNIversal Financial Industry message scheme) aims to provide a toolkit for the development of XML financial messages to meet the needs of the financial services industry. The standard has published a set of documents that define the overall methodology, the mechanism for populating the message repository, the process to make existing message standards compliant; the modelling guidelines based on unified modelling language (UML) and XML schema design rules.

ISO 19312 (the securities data model) is under development. The aim of this standard is to cover all financial instruments, standardising the terms, definitions and relationships over the transaction cycle for the instrument. Its scope also includes instrument maintenance and change resulting from corporate events, as well as the terms needed to support the events that may result from a corporate announcement.

There is no shortage of standards available in the financial services arena. To employ these standards in the development of systems to support financial innovation we need an approach that will allow workflows to be developed using the appropriate standards for each stage of the workflow.

4 Services Computing and Innovation

Using a component based approach to the construction of systems to automate business processes is an important element of process based financial innovation. The emerging field of Services Computing, Zhang et. al. [5], is concerned with the use of information and computing technology to create, operate and manage business services effectively and efficiently. When the components identified in

the analysis of business processes are viewed as business services, there is alignment with the objective of improving of business processes to derive efficiency gains in financial services organisations.

The architectural model that supports Sercvices Computing is known as Service-oriented Architecture. A Service-oriented architecture is essentially a collection of services, which communicate with each other. The following describes a basic Service-oriented architecture. A service consumer sends a service query request message to a service provider. The service provider returns a response message to the service consumer. The request and subsequent response connections are defined in a way that is understandable to both the service consumer and service provider.

A service in SOA is an exposed piece of functionality with three properties:

1. The interface to the service is platform-independent.
2. The service can be dynamically located and invoked.
3. The service is self-contained. That is, the service maintains its own state.

A platform independent interface means that a client from anywhere, on any OS, and in any language, can use the service. For a service to have the property of dynamic discovery a discovery service must be available. The discovery service enables a look up mechanism where consumers can find a service based on some criteria. For example, if an authorization service for fund transfers was required, query request would be sent to the discovery service to find a list of service providers that could authorize a funds transfer for a fee. Based on the fee, one of the available services would be selected. The last property of a service is that the service be self contained.

Service providers and consumers communicate via messages. Services expose their interface. The exposed interface defines the behaviour of the service and the messages they will accept and return. Because the interface is platform and language independent, the technology used to define messages should also be platform and language independent. Messages are typically constructed using XML, which provides all of the functionality, granularity, and scalability required by messages. For effective communication to exit between consumers and providers they need a non restrictive system to clearly define messages; XML provides a mechanism for defining standardised message formats.

Conceptually, a realistic SOA is composed of three elements: service providers, service consumers, and the directory service. The directory service acts as an intermediary between providers and consumers. Providers register their services with the directory service and consumers query the directory service to find services. Most directory services typically organize services based on criteria and categorize them. Consumers can then search the directory services to find appropriate service providers. Embedding a directory service within SOA accomplishes the following:

- Scalability of services; services can be added incrementally.
- Decouples consumers from providers.
- Allows for real time updates of services.

- Provides a look-up service for consumers.
- Allows consumers to choose between providers at runtime based on search criteria.

Having defined the basic elements of SOA we will look at the benefits that are generally attributed to this approach.

Leveraging existing assets. It is possible to construct a business service as an aggregation of existing components, using SOA; the new service can then be made available to your enterprise. Using this new service requires knowing only its interface and name. Component anonymity lets organizations leverage current investments, building services from a conglomeration of components built on different machines, running different operating systems, developed in different programming languages. Legacy systems can be encapsulated and accessed using services interfaces, adding value as their functionality is transformed into services.

Infrastructure as a commodity. Existing components, newly developed components and components purchased from a range of vendors can be consolidated within a SOA framework. This collection of components will be deployed as services on the existing infrastructure. As services become more loosely coupled from the supporting hardware, the hardware environment can be optimised to the services environment.

Faster time-to-market. Services libraries will become the organization's core assets as part of a SOA framework. Building and deploying services with services libraries reduces time to market, new initiatives reuse existing services and components, reducing design, development, testing and deployment time in the process. As services reach critical mass it will be possible to assemble composite applications using services, rather than developing custom applications.

Reduced cost. As new requirements are identified, the cost to create new services by adapting the SOA framework and the services library is greatly reduced. Familiarity with existing components flattens the learning curve for developers, reducing development costs.

Risk mitigation. Reusing existing components reduces risk, enhancing or creating new business services. Risks associated with the maintenance and management of infrastructure supporting the services will also be mitigated.

Continuous business process improvement. SOA allows a mapping between process flows constructed from components and the particular business process being automated; providing the business with a framework for monitoring business operations. Process manipulation is achieved by reorganizing the pieces in a pattern (components that constitute a business service). This allows for changes to the process while monitoring the effects, enabling a continuous improvement cycle.

Financial services organisations embark on process based innovation to generate cost savings. Using Services Computing to implement reusable components that can be used in multiple business processes will help in delivering cost savings. This leads to an obvious conclusion, research advances in the area of Services Computing will be of benefit to organisations looking for process based financial innovation. In particular progress in the development of modelling frameworks and tools for the mapping of business processes onto services, and financial models to value services will be particularly important.

Supporting this research, the establishment of broadly accepted standards which describe transactions and their transmission will also be of benefit to financial services organisations. New services for the processing of standardised transactions will be developed and published by a broad community of developers, enabling a new wave of process innovation.

References

1. Merton, R.: A functional perspective of financial intermediation. Financial Management 24(2), 23–41 (1995)
2. Riberio, J.: Glittering Prize; How financial institutions can drive growth through process and service innovation. Deloitte Research (2005)
3. Sexton, M.: MDDL and the Quest for a Market Data Standard: Explanation, Rationale, and Implementation. Butterworth-Heinemann (2007)
4. Tufano, P.: Financial innovation. In: Constantinides, G., Harris, M., Stulz, R. (eds.) Handbook of the Economics of Finance, vol. 1A, pp. 307–336. North-Holland, Amsterdam (2003)
5. Zhang, L., Zhang, J., Hong, C.: Services Computing. Springer, Heidelberg (2007)

The Effect of Automated Trading on Market Quality: Evidence from the New York Stock Exchange

Andreas Storkenmaier and Ryan Riordan

Institute of Information Systems and Management (IISM), University of Karlsruhe, Englerstr. 14, 76131 Karlsruhe, Germany
storkenmaier@iism.uni-karlsruhe.de, riordan@iism.uni-karlsruhe.de

Abstract. From the end of 2006 until the beginning of 2007 the NYSE introduced the NYSE Hybrid Market on a rolling basis. The NYSE Hybrid Market significantly changed the NYSE's market model and supports automated execution for almost unlimited order sizes and different order types. The introduction of the Hybrid Market was driven by fundamental changes in the securities trading industry over the last years. This paper analyzes the effect of the NYSE Hybrid Market on market quality through analyzing different spread measures and price impact. Results show that the introduction of the Hybrid Market reduced trading costs and improved execution quality at the NYSE.

Keywords: NYSE, Hybrid Market, Market Quality, Liquidity, Automatic Execution, Trading Technology.

1 Introduction

The securities trading industry has undergone fundamental changes during the last years. Electronic systems have become predominant on securities exchanges worldwide and alternative trading systems have undercut traditional exchanges' market shares. New technology has significantly decreased entry barriers for new market entrants and put pressure on market incumbents [10],[19]. Information technology has also altered the way traders handle orders and implement trading strategies, one example is the rise of algorithmic trading which constitutes a third of trading volume in U.S. equity markets [16].

The change in the securities trading industry has hit the New York Stock Exchange (NYSE) hard. As a large company in the trading industry, owned by members until shortly, the NYSE has been reluctant to changes and as a result lost market share to its competitors. The NYSE's market share in trading of NYSE-listed equity securities dropped from 80.1% in 2004 to 60.9% in 2007 [30]. With the introduction of the Hybrid Market from 6 October 2006 to 12 January 2007, the NYSE tries to approach those problems through a substantial change of its market model. With Regulation NMS the United States' regulatory authority also put pressure on exchanges and trading venues to change their market to

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 11–30, 2009.

comply with new regulation. The NYSE also adopted the Hybrid Market to satisfy regulatory demands.

The new NYSE Hybrid Market significantly shifts the NYSE's underlying market model towards an electronic public limit order book. The new system and market model feature automated execution even for large trades and different order types circumventing interaction with the specialist. This is supposed to decrease execution latency significantly from an average of nine seconds to 0.3 seconds[1].

While our analysis reveals that the introduction of the Hybrid Market benefits the market as a whole it further downgrades the importance of specialists and floor brokers on the NYSE. The fraction of trades with floor broker participation has already fallen from 50% at the end of the nineties to about 10% after the introduction of the NYSE Hybrid Market [17].

Section 2 provides an overview of related work. Section 3 presents the NYSE Hybrid Market. Data and methodology are explained in section 4. Section 5 provides the results with an interpretation and section 6 finally concludes.

2 Related Work

Several papers e.g. Venkatamaran [29] or Gajewski and Gresse [13] research into the differences of electronic public limit order books and more dealer oriented forms of market organization. However, results are not consistent and do not clearly favor one market type. Boehmer [7] compares the NYSE and Nasdaq and finds that additional to execution costs, speed is also an important dimension of execution quality. He states that usually a trade-off between execution speed and trading costs exists. Madhavan and Sofianos [23] empirically analyze the specialist's role on the NYSE whereas Sofianos and Werner [28] investigate floor brokers on the NYSE. However, both analyses are based on data more than ten years old which do not capture significant changes in the securities trading industry. Boehmer, Saar, and Yo [9] highlight that a change in the NYSE's market microstructure through the introduction of the OpenBook service, which improved pre-trade transparency, has increased liquidity.

Hendershott and Moulton [17] analyze the introduction of the NYSE Hybrid Market with a focus on interaction and cooperation of floor brokers. In contrast to our findings they suggest that trading costs have increased. When we conducted our analysis, Hendershott and Moulton's version [17] did neither include an analysis of different trade sizes nor an analysis of the information content of trades. With a revised version of their paper [18] they include a Hasbrouck decomposition and an analysis on the information content of trades. Our results differ from theirs in that ours suggest improved liquidity whereas their [18] results suggest a deterioration of liquidity. To make our results comparable we also include the CBOE VIX as a daily volatility measure. However, our results also hold for regressions without the CBOE VIX. In contrast to our analysis Hendershott and Moulton [18] use a matched sample with Nasdaq for their analysis. The matched

[1] Hybrid-System verbessert Schnelligkeit: New Yorker Börse startet elektronischen Handel, Handelsblatt 5 October 2006.

sample is not the reason for the different results though. Hendershott and Moulton [18] report descriptive statistics without using their matched sample. Their descriptive statistics already show an increase in quoted and effective spreads in contrast to our descriptives showing exactly the opposite direction. Although our and Hendershott and Moulton's [18] market capitalization means are rather close, the most reasonable explanation for the different results are the different samples. Unfortunately they [18] do not report which stocks are in their sample such that we cannot apply our calculation to their sample stocks for comparison. Additionally, our findings are consistent with Abrokwah and Sofianos [1]. Although they only conduct a brief analysis they also find decreasing effective spreads. Easley, Hendershott, and Ramadorai [11] find that latency reduction increases liquidity which relates to the introduction of the Hybrid Market as a latency reducing technology. In a recent paper Hendershott, Jones, and Menkveld [16] find that algorithmic trading increases liquidity for large cap stocks.

3 The NYSE Hybrid Market

Due to fast developing information technology, many exchanges have introduced automated trading systems [19]. In 2001 the NYSE followed this trend and introduced NYSE Direct+, its automated trading systems, as a pilot. Although a large step towards automation, Direct+ had two major restrictions. First, it was limited to 1,099 shares and a minimum reload time of thirty seconds had to be observed until another order could be submitted. Second, Direct+ was restricted to limit orders. From a regulatory view markets without automated execution are slow markets which can be traded through. Trade-throughs occur when a trade is executed at one market center despite the fact that another national market center offers a better quote [26]. Through investors' demands and pressure by regulation, the NYSE decided to implement a fully automated trading system, the NYSE Hybrid Market.

The most important change in the NYSE's market microstructure through the introduction of its Hybrid Market in 2006 and 2007 was the removal of strict constraints on the automatic execution of orders. The NYSE Hybrid Market features an upper limit of 1,000,000 shares, no time restriction, and market orders are also eligible for submission [25]. With the NYSE Hybrid Market, automated execution is available for large orders offering immediacy and subsecond execution. The event of lifting the restrictions on automated execution is the one that we consider for our analysis.

To enable market participants to cope with the new possibility of executing large orders automatically, the NYSE introduced new order types including new tools for floor brokers and specialists. They also significantly changed the possibilities of market participants to sent orders to the market and to define orders eligible for automated execution. Orders for automatic execution are also able to sweep the book, meaning they not only execute at the best bid or ask if possible but walk their way into the book also taking hidden liquidity from floor brokers and specialists [24].

Orders at regulated markets are routed to other fast away markets as required by Reg NMS [26] in case an away market offers a better quote. However, away markets need to qualify as fast markets to be eligible to receive orders. Fast markets are defined as markets which offer automated execution. Since the NYSE Hybrid Market qualifies as a fast market the NYSE does not see routing to fast away markets as required through U.S. regulation as a problem. The NYSE sets the national best bid and offer (NBBO) 93% of the time [20]. Prices of orders sent to the market for automatic execution can only receive price improvement if the specialist uses algorithmic price improvement [27]. To receive manual price improvement a trader can still route her order through SuperDOT to a specialist or commission a floor broker.

4 Data and Methodology

4.1 Data Source

The data source for our trade and quote data (TAQ) is the Securities Industry Research Centre of Asia-Pacific (SIRCA)[2]. SIRCA provides data that originally are Reuters data. Instruments are identified through the Reuters Instrument Code (RIC). The data provided features trades and the best bids and asks. Monetary measures are in U.S. Dollars.

Before using the data for statistical analysis it has to be cleaned to ensure consistency of results. The first and the last five minutes of a trading day are removed to avoid biased results. Data from 9:36 am to 3:55 pm local time are included in the analysis.

Additionally we retrieve the Chicago Board Options Exchange Volatility Index (VIX) to use as a daily market volatility measure. The VIX is a volatility measure derived from the S&P 500 stock index option prices. It measures expected stock market volatility over the next thirty calendar days [3]. To retrieve a daily volatility measure we calculate the daily mean price of the VIX for every trading day within the observation period.

4.2 Sample Selection

Stocks in the final sample have to comply with several criteria. Some of those criteria relate to TAQ data whereas other criteria are based on master data collected independently of Reuters data. The sample is based on one hundred randomly selected common stocks from the NYSE Hybrid Market activation list. The Hybrid Market was introduced on a rolling basis. Each instrument has an individual event date. For each individual security in the sample 125 trading days before the change to the Hybrid Market and 125 trading days after the change to Hybrid including the event date are taken into account for the analysis.

[2] http://www.sirca.org.au/

[3] cf. CBOE VIX White Paper, http://www.cboe.com/micro/vix/vixwhite.pdf

Three criteria related to TAQ data, consistent with [17], have to be satisfied for an instrument to be taken into the sample. An instrument must never be traded below $1 and it must never be traded above $500. Additionally, an instrument must never be traded twice or less per day and it has to be continuously traded over the observation period.

Securities whose stock were split during the observation period, which have been delisted during or after the observation period, or which had symbol changes during the observation period are dropped from the sample. For the analysis the stock sample is enriched with market capitalization data. Market capitalization for a company is calculated as the product of shares outstanding and the average mean price over the observation period. Finally the sample is divided into quartiles by each stock's market capitalization. Required additional information is retrieved from the website of the NYSE[4]. Those data is checked against Yahoo! finance[5], Google finance[6], and OnVista[7].

The final clean sample consists of sixty-seven common stocks.

4.3 Market Measures

Several market measures need the direction of a trade to be calculated. Direction of trade reveals if a trade is buyer or seller initiated. Our data only features trade and quote data without a trade direction indicator. To infer the direction of a trade we use the algorithm of Lee and Ready [22] with contemporaneous quotes as proposed by Bessembinder [4]. Bessembinder [4] compares different heuristics to infer trade direction with proprietary data featuring the trade direction and finds that a comparison of the trade with the contemporaneous quote using Lee and Ready's heuristic provides the best results.

Empirical measures of spreads are calculated as in Bessembinder and Kaufman [6] only we calculate full spreads instead of half spreads. All spread measures in our analysis are calculated in basis points. Quoted spreads are the easiest measures of trading costs and can easily be calculated from trade and quote data. However, quoted spreads only measure liquidity for small orders since depth is not taken into account. Spreads are calculated as relative spreads. Let $\text{Ask}_{i,t}$ be the ask price for a stock i at time t and $\text{Bid}_{i,t}$ the respective bid price. $\text{Mid}_{i,t}$ denotes the mid quote then the quoted spread is calculated as follows:

$$\text{Quoted Spread}_{i,t} = (\text{Ask}_{i,t} - \text{Bid}_{i,t})/\text{Mid}_{i,t}$$

The effective spread is the spread that measures actual trading costs. The effective spread also captures institutional features of markets like hidden liquidity or price improvement. Price improved trades are executed within the quoted spread. For such trades quoted spreads are inaccurate measures of liquidity. For

[4] http://www.nyse.com/
[5] http://finance.yahoo.com/
[6] http://finance.google.com/
[7] http://www.onvista.de/

the calculation of effective spreads trades in relation to the quote midpoint are considered and thus price improvement is incorporated in the effective spread as a liquidity measure. Price improvement is an institutional feature which does not exist in pure order driven markets since a market maker is needed to price improve orders. Additionally, effective spreads capture the effects of trades which sweep the order book because they are larger than the quoted depth. Let $\text{Price}_{i,t}$ be the execution price then the effective spread is defined as:

$$\text{Effective Spread}_{i,t} = 2 * D_{i,t} * ((\text{Price}_{i,t} - \text{Mid}_{i,t})/\text{Mid}_{i,t})$$

$D_{i,t}$ denotes the trade direction with -1 for market sell and $+1$ for market buy orders. The realized spread measures market makers revenue reduced by losses to better informed traders [6]. The loss to better informed traders is the adverse selection component (cf. [14]). We calculate the realized spread with the mid quote five minutes after the trade:

$$\text{Realized Spread}_{i,t} = 2 * D_{i,t} * ((\text{Price}_{i,t} - \text{Mid}_{i,t+x})/\text{Mid}_{i,t})$$

Price impact is a heuristic to approximate the adverse selection component in the spread. The price impact is the effective spread minus the realized spread and measures the information content of a trade. It approximates the permanent impact of a trade under the assumption that information impacts are permanent whereas other effects are transitory. Following a trade, market makers adjust their believes about the fundamental value of an asset depending on the information content of a trade (cf. [14]):

$$\text{Price Impact}_{i,t} = 2 * D_{i,t} * ((\text{Mid}_{i,t+x} - \text{Mid}_{i,t})/\text{Mid}_{i,t})$$

All measures are calculated for one minute intervals. Then they are aggregated to daily measures for further analysis. For the aggregated daily measures the mean of all daily minute data is calculated excluding the first and last five minutes of a trading day as well as opening and closing procedures. In order to avoid distorted results, minute observations featuring a quoted spread larger than 10%, an effective spread larger than 10%, a realized spread larger than 10%, or a realized spread smaller than -10% are removed from the data set. Such extreme measures have a high probability of being erroneous. Out of more than six million minute observations only 52 minutes are deleted through those requirements.

Since price impact is an imprecise measure of the adverse selection component in the spread, we calculate the permanent impact of trade innovation using Hasbrouck's approach [15]. Our implementation of Hasbrouck's algorithm is based on [8] only we use trade direction in our model as in [15] instead of net order flow. This model infers private information from the trade innovation. "The information impact of a trade may be formally defined as the ultimate impact on the stock price (or quote) resulting from the unexpected component of trade, i.e., the persistent price impact of the trade innovation" [15]. To infer the persistent price impact of the trade innovation a vector autoregressive (VAR) model is used. Let

x_{t-i} be the trade direction and r_{t-i} denotes the quote midpoint continuously compounded returns then the following model with five lags emerges:

$$x_t = \alpha_0 + \sum_{i=1}^{5} \alpha^x_{t-i} x_{t-i} + \sum_{t=1}^{5} \alpha^r_{t-i} r_{t-i} + u^x$$

$$r_t = \gamma_0 + \gamma^x_t x_t + \sum_{i=1}^{5} \gamma^x_{t-i} x_{t-i} + \sum_{t=1}^{5} \gamma^r_{t-i} r_{t-i} + u^r$$

In order to calculate the persistent price impact we look at the orthogonalized impulse responses of continuously compounded returns to trade direction. For our analysis we estimate the VAR model for each instrument for each trading day such that we have daily estimates. Summing up the orthogonalized impulse responses gives an approximation of the information impact of trades for each day.

4.4 Research Design

To analyze changes in market measures triggered through the introduction of the NYSE Hybrid Market and automated execution, we use a fixed effects model which captures cross sectional differences. Prior poolability tests reveal that cross sectional data cannot be pooled. The fixed effects model takes out stock specific effects. Additionally it can be assumed that the change to the Hybrid Market is strictly exogenous to the model. Average measures over a trading day are used which eliminates intra-day effects. Through the rolling introduction of the NYSE Hybrid Market the event dates in our analysis are scattered over almost four months. This significantly reduces the risk of measuring through panel regressions some effect not related to the introduction of the Hybrid Market.

The fixed effects model for panel regression uses trading days t as the time series component and individual stocks i as the cross section. 125 trading days before the introduction of the Hybrid Market for an individual stock and 125 trading days after the introduction including the day of introduction are analyzed for 67 stocks. $QM_{i,t}$ denotes the measure that is analyzed at a certain day and $\text{Hybrid}_{i,t}$ represents the two-valued indicator taking zero before the introduction of the NYSE Hybrid Market and one after. VIX_t represents the daily average VIX price as a volatility measure. Then the following fixed effects model emerges:

$$QM_{i,t} = \alpha_i + \beta \times \text{Hybrid}_{i,t} + \gamma \times \text{VIX}_t + \epsilon_{i,t}$$

The null hypothesis in our analysis is that the change in the market model has had no influence on market measures whereas the alternative hypothesis is that market measures were influenced through the introduction of the NYSE Hybrid Market. Through panel regression we test if we can reject the null hypothesis of insignificant regression coefficients. Our results are robust to the econometric specification as they still hold without including the CBOE VIX.

For our panel regressions we apply robust standard errors for within-groups estimators [2] which are essentially White's robust standard errors [32] adjusted for panel data. Those standard errors are consistent with heteroskedastic panel data.

Table 1. Descriptive statistics: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. For each figure mean and standard deviation are reported. Mean and standard deviation both refer to minute data. Market capitalization is calculated through multiplying the average price over the observation period with shares outstanding. Realized spread and price impact take the mid quote five minutes into the future into account for their calculations. Quoted spread, effective spread, realized spread, and price impact are reported in basis points.

		Market Cap ($Bil.)	Quoted Spread	Effective Spread	Realized Spread	Price Impact
Entire Sample	Mean	11.34	12.35	7.01	3.58	3.42
	Std.Dev.	25.38	16.35	8.95	25.02	24.73
First Quartile	Mean	37.82	5.44	3.56	1.46	2.11
	Std.Dev.	41.10	3.97	3.35	15.87	15.93
Second Quartile	Mean	4.93	9.66	6.52	3.71	2.81
	Std.Dev.	1.96	9.59	7.54	19.94	19.25
Third Quartile	Mean	1.98	14.40	8.58	4.67	3.91
	Std.Dev.	0.34	13.18	8.86	27.33	26.86
Fourth Quartile	Mean	0.61	21.01	11.15	5.73	6.03
	Std.Dev.	0.32	26.62	13.77	38.21	38.14

5 Results

In this section we present the empirical analysis. First, we provide descriptive statistics then we perform regression analyses and provide an explanation and interpretation of our results.

5.1 Descriptive Statistics

Table 1 shows descriptive spread and price impact statistics for the entire sample and market capitalization quartiles individually based on minute observations. Additionally, market cap data are shown. Descriptive statistics at first are indicative and are shown to get a feeling for the data and magnitude of values. In the next section a regression will further analyze the impact of the introduction of the Hybrid Market.

Table 2 shows the differences in mean before and after the change to the NYSE Hybrid Market for aggregated daily measures. Simple t-Tests reveal that differences between before and after Hybrid should be significantly different with respect to the given variance. For the entire sample every single spread measure decreases. Quoted spread decreases from 15.45 to 11.41, effective spreads from 10.89 to 7.85, realized spreads from 5.51 to 4.65, and price impact from 5.38 to 3.20. Smaller effective spreads suggest a decrease in the costs of trading.

Table 2. Descriptive statistics: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. For each figure mean, the difference in means, and its t-value are reported. Means refer to aggregated daily data. Realized spread and price impact take the mid quote five minutes into the future into account for their calculations. Effective spread, realized spread, and price impact are reported in basis points. Statistical significance is denoted as '***' for the 1% level, '**' for the 5% level, and '*' for the 10% level.

		Quoted Spread	Effective Spread	Realized Spread	Price Impact
Entire Sample	Pre	15.45	10.89	5.51	5.38
	Post	11.41	7.85	4.65	3.20
	Diff	4.04***	3.04***	0.86***	2.18***
	t-Val.	15.27	14.27	5.43	16.39
1st Quartile	Pre	5.92	4.01	1.40	2.61
	Post	4.94	3.28	1.66	1.62
	Diff	0.98***	0.73***	-0.26***	0.99***
	t-Val.	11.84	13.09	-5,59	19.97
2nd Quartile	Pre	10.67	7.76	4.25	3.52
	Post	9.17	6.95	4.65	2.30
	Diff	1.50***	0.81	-0.40**	1.22***
	t-Val.	5.69	3.71	-2.09	15.85
3rd Quartile	Pre	17.13	11.89	6.23	5.67
	Post	13.08	9.14	5.53	3.61
	Diff	4.05***	2.75***	0.70***	2.06***
	t-Val.	11.96	10.73	3.36	13.02
4th Quartile	Pre	28.87	20.45	10.46	10.00
	Post	18.90	12.30	6.91	5.39
	Diff	9.97***	8.15***	3.55***	4.61***
	t-Val.	11.12	10.85	6.31	9.55

Comparing quoted and effective spreads, the data show that public quotes do not reflect actual trading prices. An analysis of all spread measures and price impact by market capitalization quartile shows that all measures increase uniformly with decreasing market capitalization.

For all but two values market measures by market capitalization quartiles also decrease when comparing values before and after Hybrid. Only realized spreads for the first and second quartile of market capitalization do not decrease. It strikes that spreads and price impact are much higher for stocks with low market capitalization compared to first market cap quartile stocks. However, this is consistent with literature since spreads are wider in less liquid stocks, stocks with a lower market capitalization [5].

Table 3 shows descriptive statistics of turnover (in US$) and trading volume (in number of shares). It is obvious that turnover and volume per trade and per day decrease with decreasing market capitalization. Consistent with securities

Table 3. Descriptive statistics: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. Means refer to aggregated daily or per trade data. Turnover and volume are calculated on raw trade and quote data. Measures are shown for the entire sample as well as for individual market capitalization quartiles.

		Turnover Per Trade	Turnover Per Day	Volume Per Trade	Volume Per Day
Entire Sample	Pre	10,428	23,452,202	374	618,795
	Post	8,880	27,956,548	290	683,752
First Quartile	Pre	17,939	62,244,874	412	1,284,997
	Post	14,368	76,231,638	294	14,323,446
Second Quartile	Pre	10,585	18,211,026	452	643,278
	Post	9,084	21,181,512	364	761,291
Third Quartile	Pre	7,903	7,953,813	314	315,392
	Post	7,459	8,789,638	254	305,543
Fourth Quartile	Pre	4,961	4,270,773	314	207,307
	Post	4,342	4,227,581	246	207,833

industry development and studies (cf. [30]) turnover per day and volume per day have increased from before to after the introduction of Hybrid. More interesting is the influence of a change in the market model on per trade turnover and volume. Descriptives indicate that while overall volume increased trade turnover and volume have significantly decreased with the introduction of the Hybrid Market. For the entire sample turnover per trade dropped from \$10,428 to \$8,880, volume per trade dropped from 374 shares to 290 shares. This could be a result to changed order submission strategies like breaking up orders [21]. The effects seem to be stable over all market capitalization quartiles with the most dramatic drop in volume per trade for stocks in the first market capitalization quartile. For the regression analysis turnover per trade is aggregated to a daily measure by taking the each days mean of turnover per trade 125 trading days pre Hybrid and 125 post Hybrid for an individual stock.

It remains for the regression analysis in the following section to show which changes are statistically signficant with a subsequent assessment of its economic impact. But descriptive statistics already suggest that changes in measures for per trade turnover, spreads, and price impact will probably be statistically as well as economically significant.

5.2 Regression Analysis and Interpretation

To test if the introduction of the NYSE Hybrid Market and thus automated trading has a significant impact on market quality ($QM_{i,t}$) we use the previously

Table 4. Results Panel Regressions: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. Table 4 reports panel regression results for quoted spread, effective spread, realized spread, price impact, and per trade turnover. Results are reported for the entire sample and individually by a stock's market capitalization. Realized spread and price impact take the mid quote five minutes into the future into account for their calculations. Quoted spread, effective spread, realized spread, and price impact are reported in basis points. Monetary figures are reported in US-Dollar. Statistical significance is denoted as '***' for the 1% level, '**' for the 5% level, and '*' for the 10% level. t-statistics are reported in parantheses below panel regressions' coefficients.

	Quoted Spread	Effective Spread	Realized Spread	Price Impact	Per Trade Turnover
Entire Sample	-4.01***	-3.02**	-0.87	-2.15***	-2039***
	(-3.06)	*(-2.57)*	*(-1.28)*	*(-4.16)*	*(-6.16)*
First Quartile	-0.94***	-0.72***	0.25***	-0.97***	-4492***
	(-4.19)	*(-4.78)*	*(3.41)*	*(-6.84)*	*(-4.68)*
Second Quartile	-1.49***	-0.81***	0.40**	-1.20***	-2185***
	(-3.22)	*(-2.97)*	*(2.15)*	*(-5.54)*	*(-6.36)*
Third Quartile	-4.02***	-2.73***	-0.71	-2.02***	-664**
	(-4.22)	*(-3.16)*	*(-1.06)*	*(-5.21)*	*(-2.52)*
Fourth Quartile	-9.96**	-8.14*	-3.56	-4.57**	-736***
	(-1.97)	*(-1.78)*	*(-1.37)*	*(-2.29)*	*(-3.97)*

specified panel regression (See section 4.4). All panel regression are calculated for 125 trading days before and after the introduction of the NYSE Hybrid Market.

Almost all coefficients (Table 4) of spread measures, price impact, and per trade turnover for the entire sample are highly statistically significant. Only realized spread for the entire sample and market cap three and four are not statistically significant. The overall statistical significance allows us to generally reject our null hypothesis of insignificant regression coefficients. For the whole analysis "***" indicates statistical significance at the 1% level, "**" indicates statistical significance at the 5% level, and "*" indicates statistical significance at the 10% level. In the following we will analyze quoted spreads, effective spreads, realized spreads, price impact, per trade turnover, measures by trade size, and additionally the permanent price impact of trade innovation.

Quoted Spreads: The introduction of the NYSE Hybrid Market significantly reduces quoted spreads. The coefficient is statistically as well as economically significant with a reduction of 4.01 basis points on average. The impact strongly depends on market capitalization. The coefficient for the Hybrid Market dummy variable ranges from -0.94 for stocks in the first market capitalization quartile to -9.96 for stocks in the fourth market capitalization quartile which partly closes the gap between stocks with high and low market capitalization.

Since the change to the Hybrid Market also reduces exchange system latency [17] this result ist consistent with Easley et al. [11] who find that reducing latency increases liquidity due to increasing off-floor competition. With the introduction of the Hybrid Market specialists and floor brokers are exposed to more public limit order book competition in supplying liquidity. Since public limit orders do not feature price improvement and incur less costs an increase in public limit order competition could reduce the quoted spread. However, with respect to spreads our results are not consistent with Hendershott and Moulton [17]. Under the assumption that the introduction of automated execution facilitates algorithmic trading the result is consistent with Hendershott et al. [16].

A reduction of quoted spreads could also indicate a reduction in price improvement due to the new exchange systems that circumvent specialists. But we have to compare with effective spreads to assess that issue. The quoted spread only measures liquidity for small orders and thus is not precise in measuring liquidity. Effective spreads, realized spreads, and price impact have to be analyzed for more insights.

Effective Spreads: Like quoted spreads, effective spreads shown in Table 4 are also significantly reduced due to the introduction of the NYSE Hybrid Market. For the entire sample, the effective spread coefficient for the Hybrid Market dummy variable is -3.02. A reduction in effective spreads is consistent over all market capitalization quartiles with a considerably stronger reduction for stocks with small market capitalization ranging from -0.72 for large cap stocks to -8.14 for small cap stocks. Coefficients for market cap one to three are statistically significant at the 1% level. Market capitalization quartile four is significant at the 10% level which leads to an overall significance for the entire sample of 5%. Effective spreads are much more precise in measuring the actual trading costs than quoted spreads.

With the introduction of the NYSE Hybrid Market trading costs decreased significantly. Since effective spreads also capture immediate market impacts of large trades this suggests that liquidity increased especially for stocks with small market capitalization. These results are consistent with Boehmer et al. [9] who find that with decreasing floor broker and specialist participation due to increased pre-trade transparency and thus higher public limit order competition, effective spreads decrease and liquidity increases. According to Madhavan and Sofianos [23] specialists are most usefull in less liquid stocks. Thus a decrease in effective spreads especially for less liquid stocks due to automated execution indicates a reduction of the importance of specialists.

Interestingly, all measures for effective spreads feature a lower reduction than measures for quoted spreads. This implies that less trades are price improved by specialists. With less price improvement through specialists the NYSE loses more arguments to retain their trading floor. A decrease in price improvement is reasonable because with automated execution prices can only be improved through algorithms, not manually. Our results are consistent with Abrokwah and Sofianos [1]. After looking at quoted and effective spreads it is evident that the introduction of the NYSE Hybrid Market with its extension of automated trading decreased trading costs and increased liquidity.

Boehmer [7] suggests that speed of execution and costs are two dimensions of execution quality between which a trade-off exists. However, the change of the NYSE to the Hybrid Market does not only focus on latency reduction and pure speed but shifts the NYSE's whole market model towards an electronic public limit order book system. Different dimensions influencing liquidity are changed. Therefore, it is not imperative that an increase in speed increases trading costs and decreases liquidity. The positive impact of new trading and market organization technology could benefit execution quality in all dimensions. From our empirical data it seems to be evident that the introduction of the NYSE Hybrid Market increased speed and reduced trading costs at the same time. Since effective spreads precisely measure execution costs through capturing market specific characteristics but do not capture the difference between adverse selection and liquidity suppliers' - often specialists' - revenues we turn our analysis to realized spreads.

Realized Spreads: Realized spreads decrease for the entire sample by 0.87 basis points. This decrease is only statistically significant for the first and second market capitalization quartiles at the 1% and 5% level respectively. Thus, for realized spreads the picture is not consistent over all market capitalization quartiles. Realized spreads slightly increase for stocks from the first and second market capitalization quartile but decrease for stocks in market capitalization quartiles three and four. However only increases in the realized spread are statistically significant. This implies for the first two market capitalization quartiles that effective spreads which measure actual trading costs decrease less in magnitude than the value of the reduction of losses to informed traders. A decrease in revenues due to decreasing effective spreads is compensated through less losses to informed traders.

Liquidity suppliers' revenues however decrease for stocks in the third and fourth market capitalization quartile but not statistically significant. Especially in the fourth market cap quartile market makers revenues fall. It could be that off-floor traders rely less on specialists and floor brokers. Those orders generate competition especially for specialists. Overall however, the decrease of realized spreads in only market capitalization quartiles three and four is not as negative as effective spread measures at first suggest. Also, it is not sure what really happened to market cap three and four since the measures are not statistically significant.

Price Impact: Price impact is a proxy to analyze the information content of a trade. The adverse selection component of the spread can be approximated through the price impact measure. For the entire sample the change in price impact is significant at the 1% level with a coefficient of -2.15. As for other spread measures, coefficients are smaller for stocks with a small market capitalization. For all stocks but those in the fourth market captialization quartile, the price impact measure is significant at the 1% level. Price impact monotonically decreases from -0.97 for stocks in the first market capitalization quartile to -4.57 in the fourth market capitalization quartile.

Table 5. Results Panel Regressions: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. Table 5 reports panel regression results for the permanent price impact of the trade innovations. Results are reported for the entire sample and individual market capitalization quartiles. Figures are in basis points. Statistical significance is denoted as '***' for the 1% level, '**' for the 5% level, and '*' for the 10% level. t-statistics are reported in parantheses below panel regressions' coefficients.

	Entire Sample	M.Cap 1	M.Cap 2	M.Cap 3	M.Cap 4
Persistent Impact	-0.68***	-0.49***	-0.63***	-0.61***	-1.04***
of Trade Innovation	*(-10.85)*	*(-9.60)*	*(-6.61)*	*(-5.07)*	*(-6.11)*

These results are consistent with Hendershott et al. [16] who find that adverse selection decreases for all market capitalization quartiles due to algorithmic trading. Decreasing price impact measures imply that trades carry less information content. Another driver of reduced price impacts could be smaller trade sizes (see also section 5.2) through less information impact.

Another possible explanation besides algorithmic trading is that the NYSE manages to regain retail order flow that it has lost to other trading venues. Bessembinder and Kaufman [6] show that other trading venues cream skim retail order flow from the NYSE. However this is highly speculative, since we do not analyze the interdependence with other trading venues and exchanges comparable to Bessembinder and Kaufman [6]. It remains for future research to further analyze this question.

Looking altogether at spread measures and price impact, the NYSE's new Hybrid Market model with automated execution seems to be more competitive than the NYSE's old concentration on and support of floor-based and specialist supported trading. A caveat remains, with the analysis that we conduct it is difficult to infer a single reason for the reduction in price impact.

Persistent Price Impact of the Trade Innovation: To check the robustness of price impact measures we calculate the permanent price impact. The permanent price impact regressions in Table 5 support the explanations of the price impact. The results are smaller in magnitude but the general direction is the same and the panel regression coefficients are highly statistically significant. It is also observable that the decrease in price impact increases for smaller market capitalization quartiles.

Per Trade Turnover: Average turnover per trade (see also table 4) is an interesting measure to analyze in the light of an increasing market wide turnover. The coefficients for average per trade turnover are highly statistically and economically significant. For the entire sample, average per trade turnover falls by $2039 due to the introduction of the Hybrid Market. The coefficient is significant at the 1% level. Measures by market capitalization are significant at the 1% level

Table 6. Descriptive statistics: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. For each figure mean and standard deviation are reported. Mean and standard deviation both refer to minute data. Realized spread and price impact take the mid quote five minutes into the future into account for their calculations. Effective spread, realized spread, and price impact are reported in basis points.

		Effective Spread	Realized Spread	Price Impact
Trades above $100k	Mean	5.04	1.98	3.08
	Std.Dev.	12.44	36.45	37.09
Trades between	Mean	4.41	1.23	3.18
$25k and $100k	Std.Dev.	7.07	30.42	30.42
Trades below $25k	Mean	7.00	3.60	3.38
	Std.Dev.	8.87	37.62	37.41

for market cap one, two, and four. Market capitalization quartile three is significant at the 5% level. In absolute value the decrease is strongest for stocks in the first market capitalization quartile and least for stocks in the fourth market capitalization quartile ranging from -4492 to -736.

This development could be driven by automated execution since automated access to a market enables off-floor traders to split their orders and work large orders themselves instead of turning to a floor trader. Large orders bear higher costs of trading because they are potentially better informed [12]. Informed traders try to hide their trading interest through splitting up orders. This is a phenomenon which was already described by Kyle [21]. Faster access to markets enables order splitters to also work orders bearing lower risks. Decreasing per trade turnover also supports the results of decreasing price impacts. Trades are still informed but are not recognized as informed trades as fast as large trades without order splitting. With a reduction of about twenty percent in average per trade turnover, the reduction in trade size is economically significant and traders have reacted to changes in the NYSE's market model through adjusting their trading strategies.

Analysis by Trade Size: In this section we analyze market measures by trade size. Easley and O'Hara [12] developed an information based model that takes into account different trade sizes. Larger trades theoretically carry more information and thus have a stronger price impact.

Table 6 reports descriptive statistics for effective spread, realized spread, and price impact by trade size in basis points. Table 7 shows the differences in measures before and after the introduction of the Hybrid Market for aggregated daily measures. It also features a simple t-test. As for measures for all trade sizes it is already evident from descriptives that the change of the NYSE's market model has a significant impact on market measures.

Interestingly, lowest trading costs are shown for mid-sized trades and not large trades when considering the calculated one minute measures without daily

Table 7. Descriptive statistics: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. For each figure the pre and post mean, the difference in means, and its t-value are reported. Means refer to aggregated daily data. Realized spread and price impact take the mid quote five minutes into the future into account for their calculations. Effective spread, realized spread, and price impact are reported in basis points. Statistical significance is denoted as '***' for the 1% level, '**' for the 5% level, and '*' for the 10% level.

		Effective Spread	Realized Spread	Price Impact
Trades above $100k	Pre	12.46	7.50	4.86
	Post.	9.86	3.64	6.19
	Diff	2.60***	3.86***	-1.33**
	t-Val.	6.01	5.94	-1.98
Trades between	Pre	11.00	4.86	6.11
$25k and $100k	Post	8.69	3.83	4.85
	Diff	2.31***	1.03***	1.26***
	t-Val.	9.09	2.59	3.21
Trades below $25k	Pre	10.85	5.54	5.32
	Post	7.84	4.73	3.11
	Diff	3.01***	0.81***	2.21***
	t-Val.	14.14	5.08	17.99

aggregation. In our further analysis we consider daily aggregated data which weights each cross section and day equally despite different numbers of minute observations for individual stocks and days. Before the change to the NYSE Hybrid Market trades between $25,000 and $100,000 have an effective spread of 11.00 basis points, a realized spread of 4.86 basis points, and a price impact of 6.11 basis points. After the change to Hybrid mid-sized trades have an effective spread of 8.69 basis points, a realized spread of 3.83 basis points, and a price impact of 4.85 basis points. Before the change to the Hybrid Market the price impact is the highest for mid sized trades. This has already been observed by Barclay et al. [3] and Werner [31]. Informed traders try to hide their trading interest in order to avoid high trading costs. They use trading strategies like order splitting although they often originally have larger orders to fill. Thus, a change in order submission strategies due to automated execution also influences results by trade size.

Table 8 shows the results of panel regressions on market measures separated by trade size. Most coefficients are significant at the 1% level. Trades below $25,000 have a change in effective spread which is significant at the 5% level and a change in realized spread which is not statistically significant. The increase in price impact for trades above $100,000 is only significant at the 10% level. As expected from the spread and price impact analysis for the entire sample, values overall (but one) decrease for effective spreads, realized spreads, and price impact. The magnitude of change however is different depending on trade size.

Table 8. Results Panel Regressions: The Sample consists of 67 common stocks listed at the NYSE. The observation period comprises of 125 trading days before and after the introduction of the NYSE Hybrid Market for each individual stock. Table 8 reports panel regression results for effective spread, realized spread, and price impact. Results are reported for the entire sample. Realized spread and price impact take the mid quote five minutes into the future into account for their calculations. Effective spread, realized spread, and price impact are reported in basis points. Statistical significance is denoted as '***' for the 1% level, '**' for the 5% level, and '*' for the 10% level. t-statistics are reported in parantheses below panel regressions' coefficients.

	Effective Spread	Realized Spread	Price Impact
Trades above \$100k	-2.05*** *(-2.78)*	-3.62*** *(-5.18)*	1.63* *(1.71)*
Trades between \$25k and \$100k	-2.64*** *(-5.70)*	-1.18*** *(-2.77)*	-1.40*** *(-3.39)*
Trades below \$25k	-2.99** *(-2.54)*	-0.82 *(-1.16)*	-2.18*** *(-4.41)*

For effective spreads values range from -2.05 for trades above \$100,000 to -2.99 for trades below \$25,000.

Trading costs have decreased stronger for small trades than for large trades. This result probably interacts with the results that the impact on stocks with small market capitalization was stronger than on highly liquid stocks with large market capitalization. For less liquid stocks per trade turnover tends to be lower than for blue chips.

Liquidity providers' revenues measured through realized spreads decrease the highest for trades above \$100,000. The impact on large trades is about three to four times the impact on mid-sized and small trades. This is connected to an increase in price impact for large trades, the only coefficient that is not negative for market quality measures by trade size. Due to an interaction with changing trade sizes after the introduction of the NYSE Hybrid Market it is unclear which effects drive different changes for different trades sizes. The reduction of the price impact is strongest with -2.18 basis points for small trades driven by the strongest reduction in effective spreads. This result is consistent with the strong negative impact of Hybrid on price impact for stocks in the fourth market capitalization quartile for all trade sizes combined. Price impact for trades above \$100,000 is the only measure with a increasing coefficient. Altough it is statistically significant at the 10% level it is not as significant as most other measures.

6 Conclusion

In this paper we study the effect of the introduction of the NYSE Hybrid Market on market quality. The NYSE Hybrid Market is a new market model featuring automated execution which reduces specialist interaction. The NYSE's new

market model also alters market rules significantly, fundamentally changing the way trading is handeled on the NYSE.

Results suggest that market quality as a whole has improved. Speed has increased with the introduction of new trading technology. Our results show that at the same time trading costs significantly decreased. Quoted spreads, effective spreads, and realized spreads uniformly decreased over stocks with different market capitalization. Aside from large trades price impact also decreased over all market capitalization quartiles. Boehmer [7] states that usually there is a trade-off between execution costs and execution speed. With the change to the NYSE Hybrid Market however both speed and costs improved. The new market model not only changed execution speed but also changed trading rules and trading procedures fundamentally. The increase in market quality in terms of speed and costs suggests that the introduction of the NYSE Hybrid Market changed different fundamental quality determinants. Due to complex interdependencies between different effects of the new Hybrid Market it is difficult to extract single influences on market quality.

There are two groups however who lose due to the NYSE Hybrid Market: specialists and floor brokers. Specialists are less and less required to arrange trades and with automated execution and enhancements in algorithmic trading investors are not required to only rely on floor brokers to work large orders. The next years will be decisive for the future of specialists and floor traders at the NYSE: Will they be replaced by computer systems or will it be possible for them to generate additional value that customers are willing to pay for?

It remains for future work to more precisely isolate single effects that have changed market quality. Which determinants of the NYSE Hybrid Market have influenced which quality parameters? Proprietary data sets with more information on the type of traders and classification into algorithmic traders and human traders could help to conduct further analysis.

References

1. Abrokwah, K., Sofianos, G.: Execution quality at the new, fast nyse. Journal of Trading (2008) (Winter, forthcoming)
2. Arellano, M.: Computing robust standard errors for within-groups estimators. Oxford Bulletin of Economics and Statistics 49(4), 431–434 (1987)
3. Barclay, M.J., Litzenberger, R.H., Warner, J.B.: Private information, trading volume, and stock-return variances. The Review of Financial Studies 3(2), 233–253 (1990)
4. Bessembinder, H.: Issues in assessing trade execution costs. Journal of Financial Markets (6), 233–257 (2003)
5. Bessembinder, H., Kaufman, H.M.: A comparison of trade execution costs for nyse and nasdaq-listed stocks. The Journal of Financial and Quantitative Analysis 32(3), 287–310 (1997)
6. Bessembinder, H., Kaufman, H.M.: A cross-exchange comparison of execution costs and information flow for nyse-listed stocks. Journal of Financial Economics 46, 293–319 (1997)

7. Boehmer, E.: Dimensions of execution quality: Recent evidence for US equity markets. Journal of Financial Economics 78, 553–582 (2005)
8. Boehmer, E., Broussard, J.P., Kallunki, J.P.: Using SAS in Financial Research. SAS Institute Inc., Cary (2002)
9. Boehmer, E., Saar, G., Yu, L.: Lifting the veil: An analysis of pre-trade transparency at the NYSE. The Journal of Finance 62(2), 783–815 (2005)
10. Domowitz, I., Steil, B.: Automation, trading costs, and the structure of the securities trading industry. Brookings-Wharton Papers on Financial Services, pp. 33–82 (1999)
11. Easley, D., Hendershott, T., Ramadorai, T.: The price of latency. Working Paper, Cornell University, University of California at Berkley, and University of Oxford and CEPR (Version 21 May 2008), http://ssrn.com/abstract=961041 (accessed July 17, 2008)
12. Easley, D., O'Hara, M.: Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90 (1987)
13. Gajewski, J.F., Gresse, C.: Centralised order books versus hybrid order books:a paired comparison of trading costs on nsc (euronext paris) and sets (london stock exchange). Journal of Finance and Banking 31, 2906–2924 (2007)
14. Glosten, L.R., Milgrom, P.R.: Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100 (1985)
15. Hasbrouck, J.: Measuring the information content of stock trades. The Journal of Finance 46(1), 179–207 (1991)
16. Hendershott, T., Jones, C.M., Menkveld, A.J.: Does algorithmic trading improve liquidity. Working Paper, Haas School of Business University of California at Berkeley, Graduate School of Business Columbia University, and VU University Amsterdam Tinbergen Institute (Version 26 April 2008), http://ssrn.com/abstract=1100635 (accessed May 10, 2008)
17. Hendershott, T., Moulton, P.C.: The shrinking new york stock exchange floor and the hybrid market, Working Paper, Haas School of Business and Fordham Graduate School of Business (2007), http://www.bnet.fordham.edu/pmoulton/Hybrid_20070904.pdf (accessed May 20, 2008)
18. Hendershott, T., Moulton, P.C.: Speed and stock market quality: The nyse's hybrid. Working Paper, Haas School of Business and Fordham Graduate School of Business (2008), http://ssrn.com/abstract=1159773 (accessed October 5, 2008)
19. Jain, P.K.: Financial market design and the equity premium: Electronic versus floor trading. The Journal of Finance 9(6), 2955–2985 (2005)
20. Kinney, C.: Electronic and floor-based trading: The NYSE hybrid market. In: Schwartz, R.A., Byrne, J.A., Colaninno, A. (eds.) Electronic and Floor-Based Trading. Zicklin School of Business Financial Markets Series, ch. 7, pp. 111–120. Springer, Boston (2006)
21. Kyle, A.S.: Continuous auctions and insider trading. Econometrica 53(6), 1315–1336 (1985)
22. Lee, C.M.C., Ready, M.J.: Inferring trade direction from intraday data. The Journal of Finance 46(2), 733–746 (1991)
23. Madhavan, A., Sofianos, G.: An empirical analysis of nyse specialist trading. Journal of Financial Economics 48, 189–210 (1998)
24. NYSE: Hybrid market training program (2006), http://www.nyse.com/pdfs/hm_booklet.pdf (accessed May 21, 2008)

25. NYSE: NYSE hybrid market FAQ (2006), http://www.nyse.com/pdfs/hybridfaqs.pdf (accessed May 21, 2008)
26. SEC: Release No. 34-51808 (June 9, 2005): Regulation NMS. U.S. Securities and Exchange Commission (2005)
27. SEC: Release No. 34-53539 (March 22, 2006): Order Approving Proposed Rule Change and Amendment Nos. 1, 2, 3, and 5 Thereto and Notice of Filing and Order Granting Accelerated Approval to Amendment Nos. 6, 7, and 8 to the Proposed Rule Change to Establish the Hybrid Market. U.S. Securities and Exchange Commission (2006)
28. Sofianos, G., Werner, I.M.: The trades of nyse floor brokers. Journal of Financial Markets 3, 139–176 (2000)
29. Venkataraman, K.: Automated versus floor trading: An analysis of execution costs on the paris and new york exchanges. The Journal of Finance 56(4), 1445–1485 (2001)
30. Weber, B.: Transformation of trading at the new york stock exchange, 1980-2007. Working Paper, London Business School, Version 9 April 2008. Journal of Management Information Systems (submitted, 2008)
31. Werner, I.M.: Nyse order flow, spreads, and information. Journal of Financial Markets 6, 309–335 (2003)
32. White, H.: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48(4), 817–838 (1980)

Know the Flow: Sentiment Extraction from Retail Order Flow Data

Matthias Burghardt and Ryan Riordan

Institute of Information Systems and Management (IISM)
University of Karlsruhe
Englerstr. 14, 76137 Karlsruhe, Germany
{burghardt,riordan}@iism.uni-karlsruhe.de

Abstract. Retail investor sentiment has been a subject of interest in the finance literature for a number of years. Using the order flow based on a unique data set with 20.7 million transactions in bank-issued warrants from the European Warrant Exchange, we present the construction of a retail investor sentiment index. We show that retail investors are contrarian, that retail investor sentiment is an important part of the equity pricing process and that we have a good measure of the sentiment. We believe that this information can be used by exchanges to increase transparency in financial markets. As a whole our findings further support a role for retail investor sentiment in the equity pricing process.

Keywords: retail investor exchange, order flow, sentiment information extraction, market transparency, option trading.

1 Introduction

Investor sentiment has been a subject of interest in the finance literature for a number of years. The study of investor sentiment has its basis in the theories of noise trader models [1, 2]. Both suggest that, if some traders trade on 'noisy' signals, unrelated to fundamental data, then market prices can deviate from intrinsic value, thereby violating the Efficient Markets Hypothesis [3]. The 'behavioral' perspective continues to surmise that interplay between noise traders and arbitrageurs establishes prices [4]. This is contrary to the Efficient Market Hypothesis that theorizes the market price of an asset deviates minimally from the present values of expected future cash flows, and that the actions of arbitrageurs readily absorb demand shocks and shifts in investor sentiment.

Shleifer [5] mentions two major foundations of behavioral finance: limited arbitrage and investor sentiment. Investor sentiment is mainly driven by two phenomena: the representativeness heuristic, i.e. the tendency of people to view events as representative of some specific class and ignore the laws of probability in the process, and conservatism, which leads people to a slower updating of models in the face of new evidence than is necessary. These two drivers result in both overreaction and underreaction of investors on stock markets.

The debate as to the effects and relevance of investor sentiment continue unabated. At least since the De Long *et al.* paper on behavioral theory [6], financial researchers

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 31–46, 2009.

have been searching for measures of sentiment and the effect thereof on market returns. There is a growing body of empirical literature that investigates the dynamic relationship between investor sentiment and stock market returns. Two findings are important to support the influence of investor sentiment on stock returns: First, individual investor trades are systematically correlated, i.e. investors sell and buy stocks in concert. Second, rational arbitrageurs cannot use this fact to their advantage, either because the stocks in question are difficult to arbitrage, or the individual investor sentiment is not public knowledge. Several recent studies [7, 8, 9, 10] show – without definitely testing the noise trader theory – that information about retail investor sentiment can be used to reliably predict stock returns.

We argue that investor sentiment plays an important role in the price formation process, and adds to the transparency of financial markets, especially those for retail investors. In this paper, we model and construct a retail investor sentiment index from market data using a unique data set provided by the European Warrant Exchange (EUWAX). With this data set, we are able to avoid many limitations of the previous studies without a loss of generalization and without introducing new problems or biases. Our data, further described in section 3, covers all EUWAX transactions in warrants (bank-issued options) from January 2004 to May 2008. EUWAX is a market maker driven warrant exchange, and is the largest in Europe in terms of transaction volume. The distinct market model and the structure of the traded products allow us to create a unique retail investor sentiment index that correlates well with the market. Our paper also contributes to the e-market literature in that we show how sentiment information can be used to make electronic financial markets more transparent and more efficient. Especially since sentiment-information based on retail investor order flow can be computed and distributed real-time, it can become an integral part of electronic trading systems.

The paper is structured as follows: The next section gives an overview over related research with a clear focus on market-data-based sentiment measures and their implications on the equity pricing process. Section 3 describes the properties of the unique data set we use and the construction of our sentiment index. We perform several statistical analyses in section 4 in order to validate our findings. Section 5 concludes and gives an outlook over future research.

2 Related Research

2.1 Measures of Investor Sentiment

In the literature, two basic types of sentiment measure are identified [11, 12]: direct and indirect sentiment measures.

Indirect sentiment measures refer to financial variables and require a theory relating them to sentiment. The weakness of the indirect measures lies in the necessity of building up this theory and their respective interpretation. However, the use of indirect measures of investor sentiment is wide-spread in the academic literature because they are easily constructed and based on simple market data.

Shiller [13] also differentiates between two types of measures and establishes *direct sentiment measures:* the expectations of the market participants are measured directly

based on polling of investors. A famous example is the survey of the American Association of Individual Investors (AAII) or the Consumer Confidence Index. Harmful are many possible sources of errors, e.g. the interviewer, the questionnaire or the respondent [14]. There are often problems related to inaccurate responses, non-response and self-selecting biases that can influence the results. Polling also bears the risk of sample selection biases while market data allows for the extraction of investor sentiment from hard facts.

Another type of sentiment measure is considered [15]: The *combined direct and indirect measure* of sentiment, which includes the combination of different indicators and techniques, e.g. [11] and [16]. Novel methods developed in [17] or [18] aren't included. For this reason we refer to a third type of measure, the so called *meta-measure*, which includes analyst opinions [19, 20], columns, online forums [18, 21] and other non-standard measures of opinion. Typically, these measures are based on an amalgam of opinions.

2.2 The Relationship between Investor Sentiment and Stock Returns

There is a growing body of literature that examines the dynamic relationship between investor sentiment and contemporaneous as well as future stock returns. These studies typically use indirect sentiment measures such as flow data, put/call ratios, and buy/sell imbalances, and relate them to stock returns.

Jackson [7] analyzes a unique dataset of individual investor trades in Australia and examines whether the investment decisions of individual investors aggregate in a systematic way. He finds that the aggregation assumption holds across 56 unrelated brokerage firms and across sub-periods and across subsets of stocks. Regarding the relationship between individual investor trading and future stock returns, he finds that net trades of full-service brokerage clients positively and significantly forecast future short-term market and cross-sectional returns. Similarly, [22] show that the trading of individual investors at a large discount brokerage and a large retail brokerage is systematically correlated. Using, another brokerage data set from the U.S., Kumar and Lee [23] also show that retail investor trades are systematically correlated and retail investors buy and sell stocks in concert. Moreover, they find that monthly changes in retail sentiment induce co-movement in stock returns.

Baker and Stein [24] use trading volume as a measure of liquidity which can also be interpreted as a measure of investor sentiment. They identify short-selling as a problem and assume that investors generally only invest when they are optimistic and thereby assume that investors reduce liquidity as they become more pessimistic. In their data, they find a negative relation between annual turnover and subsequent annual stock returns suggesting that irrational investors overreact and cause stock prices to revert.

Pan and Poteshman [8] analyze option trading data and construct open-buy put-call ratios from option volume initiated by buyers to open new positions. When comparing the predictability of the other option volume types (open sell, close buy, close sell) they find that they do predict the correct direction but to a much lower extent than the open buy volume. Using their non-publicly available put-call ratio and publicly observable option volume, Pan and Poteshman examine the predictability of option trading for future stock price movements. They find that the open-buy put-call ratio positively predicts future stock prices on a weekly horizon.

Kaniel, Saar, and Titman [9] examine individual investors' trading behavior using a unique dataset of the New York Stock Exchange that contains the aggregated volume of executed buy and sell orders of individuals over four years. Very similar to

our methodology, they define individual investors' net trading as the difference between buy and sell volume. They find that the trades of individuals can indeed be used to systematically forecast future returns and show that statistically significant positive payoffs can be realized when this information is leveraged.

In a very similar study [10], Barber, Odean, and Zhu use TAQ/ISSM data to investigate the effect of order imbalance of individual investors. Like several studies before, Barber, Odean, and Zhu document strong herding by individual investors. Over shorter periods, they find that stocks heavily bought by individual investors one week earn strong returns in the subsequent week, while stocks heavily sold one week earn poor results in the subsequent week. Over longer periods, however, Barber, Odean, and Zhu document that stocks with small-trade selling pressures outperform those with buying pressures, indicating a negative relationship between investor sentiment and stock returns at annual horizons.

Using the same data set, Hvidkjaer [25] shows that stocks with intense sell-initiated small-trade volume, measured over several months, outperform stocks with intense buy-initiated small-trade volume. His findings support the role of retail investor sentiment as a contrary indicator for long-term stock price movements and positively relate to those of Barber, Odean, and Zhu. In a paper similar in spirit [26], Frazzini and Lamont study the effect of mutual fund flows on stock returns and find that stocks favored by retail investors tend to underperform in subsequent years.

Using the dataset of a large German discount broker [27], Dorn, Huberman, and Sengmueller show that retail investors tend to be on the same side of the market in a given stock. In particular, they calculate the herding measure by Lakonishok, Shleifer, and Vishny [28] and relate it to contemporaneous and future daily returns. They argue that aggregated market orders lead returns whereas aggregated limit orders appear to be negative feedback trades. They point out that the distinction between market and limit orders is crucial for understanding the relation between individual investor trading and returns.

Warrant trades are identified as an effective measure of sentiment in [29]. Their data is collected from a German discount broker and includes the portfolio data of a large number of investors. They calculate a simple sentiment measure based on whether or not investors hold, call, put or both. Investors that strictly hold calls (puts) are classified as optimistic (pessimistic), and investors that hold both are classified as neutral.

2.3 Discussion of the Literature

The evidence on the relationship between individual investor trading seems to differ depending on three dimensions that distinguish the different studies: (i) the horizon of the relationship, (ii) the underlying data set (trades, transactions, instruments, etc.), and (iii) the method of aggregating the data.

Concerning the horizon, all of the studies above document a positive relation of individual investor sentiment and future stock returns on a daily or weekly level, whereas many of the studies find a negative relation on much longer horizons, i.e. months and years. Of course, the availability of high-frequency data has improved considerably in the last couple of years so that short-horizon relations were difficult to detect in earlier studies.

The underlying data sets also differ in a couple of important dimensions: First, there are two major sources for transaction data: exchange data with executed transactions for a whole market, and brokerage data with trades and portfolio information on an

individual account level. The brokerage data has the advantage that trading decisions can be traced to individual accounts, and individual portfolios can be reconstructed. The exchange data usually does not contain information about individuals but contains trades from multiple brokerages and therefore allows for the distinction of different order flows. A second dimension is the differentiation of trade direction: Some of the papers rely on algorithms [30] to infer buyer- or seller-initiated trades, others use unique data sets where the trade direction is part of the data. The third dimension is the problem of disentanglement. Disentanglement is the act of separating institutional transactions from retail investor transactions. Some studies use retail brokerage data sets where all trades are supposed to come from retail investors, others rely on the small-trade volume or low-market-capitalization stocks. Fourth, the country of origin may be of importance since investors in Taiwan may act differently from those in the U.S. or in Germany. A fifth aspect concerns whether it is possible for retail investors to express negative sentiment. A retail investor has two practical alternatives in the equity market when trying to express a negative investor sentiment. They have the option to sell stock currently held, or if they already hold nothing they can decide not to buy. The option to sell stock short is practically impossible due to margin requirements, short-sell restrictions, and levels of investor knowledge. This severely limits a retail investor's ability to express a negative sentiment, and so an important part of the whole picture may be missing.

The aggregation methodology across the literature has common components but differs in certain aspects: Studies based on market data usually use order flow data of executed transactions and aggregate them creating measures of imbalance for buy and sell decisions, put and call volumes, small trades and large trades, etc. There are many possible ways to construct a sentiment measure out of order flow data, and many of them are highly correlated.

In our paper, we construct an *indirect sentiment measure* based on market data. We try to address most of the weaknesses mentioned above and avoid them by using a unique data set that (i) consists of retail investor orders only, (ii) captures orders from different brokerages, (iii) includes the explicit trade direction due to the market model, (iv) allows for the expression of negative sentiment through put options, and (v) captures differences in risk. We firmly believe that accurate measures of investor sentiment can be found in market data, and that investor sentiment information adds to the transparency of financial markets. The primary reason for our focus is the belief that market data measures what investors do and not what they say. By measuring what investors actually do, we do not have to interpret their responses nor control for possible biases arising there from.

3 Data

3.1 European Warrant Exchange

EUWAX is Europe's largest exchange for securitized derivatives in terms of turnover and number of trades. Securitized derivatives comprise a wide range of securities, e.g. covered warrants (bank-issued options), knock-outs (bank-issued barrier options), and investment certificates. They all have in common that their value depends on another financial instrument such as stocks, indices, commodities or fixed income products. Securitized derivatives are usually issued by banks which are obliged to continuously calculate and publish bid and ask prices for their products.

Turnover in derivative products accounts for about 3.5 billion Euros per month totaling to 184.3 billion Euros for the period of January 2004 to May 2008 (see Table 1). At the end of the sample period, there were more than 150,000 leverage products and more than 170,000 investment products listed. EUWAX is a retail investor exchange since the products offered are tailored to retail investors. The design of the issued securities is such that they are largely unattractive to institutional investors, but all the more appealing to retail investors. Products are not standardized and strike prices, duration, underlyings, payoff profile, and multiplier all vary from product to product. They are constructed in order to attract retail investors that value variability, dynamic combinations of underlyings, and optically low prices. These characteristics actively dissuade institutional investors that typically value standard and therefore easily hedged products, low counterparty risk, liquidity, and volume. These market characteristics and general acceptance that large investors avoid these types of markets obviate the need to control for institutional investors' influence in our dataset.

3.2 Key Facts

3.2.1 Quote-Driven Market

Trading at EUWAX takes place in a quote-driven market where market-makers publish quotes continuously. There are two kinds of trading participants in this context: Firstly, there are the market-makers (usually the product issuers) who publish bid and ask quotes for their products. Secondly, there are order flow providers (usually banks or brokerages) who route client orders to the exchange. Thirdly, there are quality liquidity providers acting on behalf of the exchange who assist in the matching procedure by providing quality and liquidity. They also continuously check the order book for executable orders and start the matching procedure.

This type of market model allows us to distinguish retail orders and market maker orders. Our main assumption is that retail investors express their expectations about the market by buying and selling derivatives whereas issuers have an obligation to trade but usually do not initiate trades themselves. A retail investor's order can therefore be matched either against an issuer's order or against another retail investor's order from the order book.

3.2.2 Risk

The derivative products listed at EUWAX can be categorized as either leverage or investment products. Leverage products comprise covered warrants (bank-issued options) and knock-out products (bank-issued barrier options), whereas investment products comprise all types of investment certificates.

Derivative products can be further categorized according to the risk inherent in their structure. Leverage products are usually regarded as riskier than investment products because investors face the risk of losing their whole investment in a relatively short time period. Knock-outs, for example, are worthless when the price of the underlying falls below (down and out) or goes above (up and out) a specific price. Investment certificates, on the other hand, usually exhibit some security features so that the investor is partially protected against a total loss.

Table 1. Executed order volume and number of executed orders in the sample period (05/01/2004 – 30/05/2008)

	Executed order volume (in million Euros)				
	Call		Put		Total
	Buy	Sell	Buy	Sell	
Warrants	14,709	16,138	4,096	3,839	38,782
Knock-outs	22,370	23,413	6,189	6,443	58,415
Investment certificates	46,070	39,073	1,038	916	87,097
Total	**83,149**	**78,624**	**11,323**	**11,198**	**184,294**

	Number of executed orders				
	Call		Put		Total
	Buy	Sell	Buy	Sell	
Warrants	3,198,637	2,514,233	862,641	693,377	7,268,888
Knock-outs	3,156,826	2,979,687	1,402,703	1,356,999	8,896,215
Investment certificates	2,500,146	1,899,514	56,302	38,990	4,494,952
Total	**8,855,609**	**7,393,434**	**2,321,646**	**2,089,366**	**20,660,055**

3.2.3 Positive and Negative Sentiment

For each product category there are two basic option types: calls and puts. The purchase of a call option is interpreted as a positive sentiment about the future return of the underlying whereas the purchase of a put option is interpreted as a negative sentiment about the future value of the underlying. Although the terms *call* and *put* are typically used in the options market, we can also make this distinction in the investment products market. The pay-off profiles of investment products also exhibit degrees of optionality. For example, the value of reverse convertible certificates rise as the price of the underlying(s) fall, exhibiting put option characteristics. Effectively, since investment certificates are structured using combinations of the underlying, plain-vanilla, puts and calls, and exotic options, lookback, asian, barrier, and others, we can easily interpret differentiated sentiment based on the transaction data.

3.3 Data Description

We have collected transaction data from EUWAX for the period of January 2004 to May 2008. This data set contains 20,660,055 executed orders from retail investors. The matching orders from market makers are not taken into account. The whole set of orders account for a volume of about 184 billion Euros. Table 1 reports the executed order volume and the number of executed orders by product category (rows), option type, and trade direction (columns).

As can be expected from a retail investor exchange, call options account for more than 88% of the traded volume whereas puts for only 12%. This is mostly due to the fact that most retail investors neither have the knowledge nor the willingness to trade put options. Interestingly, there is much more put volume traded in leverage products (warrants and knock-outs) than in investment products.

Table 2. Executed order volume by product type and underlying category

	Executed order volume (in million Euros)				
	Stocks	Indices	Fixed Income	Currencies	Commodities
Warrants	18,869	16,154	149	1,843	1,757
Knock-outs	9,292	41,851	705	1,722	4,620
Investment certificates	35,948	44,563	1,082	79	4,804
Total	**64,109**	**102,568**	**1,936**	**3,644**	**11,180**

Table 2 presents the executed order volume depending on the underlying types of the traded products. Indices and individual stocks are clearly the most used underlyings for derivative products and account for about 88% of the overall trading volume. Most traded underyings for single stocks comprise Allianz (9.3%), Daimler (6.2%), Deutsche Bank (5.3%), and Deutsche Telekom (5.0%). The most traded index is the DAX (63.7%) followed by EuroStoxx 50 (13.1%) and Nikkei 225 (3.7%).

3.4 Index Calculation

Intuitively, positive investor sentiment should be expressed by a high index level, and negative investor sentiment by a low index level. The intuition of the index level is the following: Bullish investors buy calls or sell puts whereas bearish investors sell calls and buy puts. Therefore, we categorize the order volume as driven by positive and negative sentiment and calculate the ratio of the net order volume to the overall order volume.

As a requirement for the index design, the index should be invariant to a rise or drop in overall order volume, i.e. the index values should not depend on the absolute order volume but rather on the proportion of buy and sell order volumes. Therefore, investor sentiment values are not influenced by an overall rise in the market volume.

The sentiment index is calculated for a given time period t as follows:

$$sentiment_t = \frac{\sum_i (v_i \cdot o_i \cdot t_i)}{\sum_i v_i} \quad (1)$$

where

v_i … transaction volume of order i (product of trade price and trade size)
t_i … trade direction of order i (t_i = 1 buy order, t_i = -1 sell order)
o_i … option type of the instrument traded through order i (o_i = 1 call option, o_i = -1 put option)

Sentiment can be calculated using either the aggregate transaction volume of the orders or the number of executed orders. Transaction volume measures take into account different order sizes whereas order based measures treat each executed order the same.

The time period used for calculation could be daily, weekly, or monthly. The larger the time period is, the noisier becomes the sentiment index. On the other hand, we have less granularity and descriptiveness in our resulting index.

Instead of choosing different calculation periods we can also use a moving average of the daily index calculation in order to reduce noise. A backward-looking moving average of the daily index calculation seems to be the most appropriate because noise is reduced, trends can be recognized, and a new index value is calculated every day. A disadvantage, however, is that sudden changes in the investor sentiment are not represented immediately as a delay is introduced as an artifact of the moving average.

It should be noted, however, that correlations are always calculated using the raw time series although we plot moving averages of the sentiment index for illustration purposes.

4 Analysis

In this section, we compare our sentiment measure with market returns because if sentiment is one of two drivers of equity prices we should find correlations between our measure thereof and market returns. Firstly, we present a covariance analysis of different order flow measures and show how the different measures correlate with market returns. Secondly, we present a cross-sectional analysis in which we form two portfolios according to sentiment values. We show that our sentiment measure can be used to generate excess returns.

4.1 Covariance Analysis

We first look at the different components of the order flow: buy call volume, sell call volume, buy put volume, and sell put volume (see Table 1 for absolute volume numbers). In order to be invariant from an overall rise or fall of absolute order volume, we calculate the ratio of all order flow subgroups to the sum of the order flow:

$$buy\_call\_ratio = \frac{buy\_call\_volume}{total\_volume} \tag{2}$$

$$sell\_call\_ratio = \frac{sell\_call\_volume}{total\_volume} \tag{3}$$

$$buy\_put\_ratio = \frac{buy\_put\_volume}{total\_volume} \tag{4}$$

$$sell\_put\_ratio = \frac{sell\_put\_volume}{total\_volume} \tag{5}$$

where *total_volume* is the sum of buy call, sell call, buy put, and sell put volume in a certain time period. We calculate the order flow ratios for all product categories, and for each single product category such as warrants, knock-outs, and investment certificates. We then compute the contemporaneous correlation between the different flow ratios and the DAX logreturns.

Table 3 shows the results of the comparison. The first part of the table (i) shows the correlation of the ratios based on order volume, the second part shows the results based on the number of executed orders. Correlations are qualitatively similar but higher when using the number of executed orders.

Table 3. Correlations of raw buy/sell ratios and contemporaneous DAX logreturns by product category. p-values in brackets. Panel 1 contains results for correlations based on the order volume (in Euro). Panel 2 presents correlation results based on the number of executed orders.

Panel 1: Ratios based on order volume

	buy call ratio	sell call ratio	buy put ratio	sell put ratio	sentiment
all products	-0.4129	0.5227	0.2859	-0.4039	-0.6498
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
warrants and knockouts	-0.5030	0.5716	0.2757	-0.4111	-0.7271
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
warrants	-0.3745	0.5143	0.1970	-0.4361	-0.6315
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
knockouts	-0.4725	0.4706	0.2268	-0.2424	-0.6839
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
certificates	-0.0891	0.1242	0.0930	-0.1803	-0.1613
	(0.0028)	(0.0000)	(0.0018)	(0.0000)	(0.0000)

Panel 2: Ratios based on the number of executed orders

	buy call ratio	sell call ratio	buy put ratio	sell put ratio	sentiment
all products	-0.6326	0.3246	0.6152	-0.1602	-0.7581
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
warrants and knockouts	-0.6658	0.3224	0.6154	-0.1485	-0.7612
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
warrants	-0.6461	0.6539	0.5759	-0.5836	-0.7376
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
knockouts	-0.5354	0.1038	0.4410	0.0655	-0.7164
	(0.0000)	(0.0005)	(0.0000)	(0.0282)	(0.0000)
certificates	-0.0677	0.0614	0.1621	-0.2048	-0.1921
	(0.0233)	(0.0398)	(0.0000)	(0.0000)	(0.0000)

A very interesting result is that the sentiment indices calculated from warrant and knock-out trades show a significantly higher correlation in absolute value than the sentiment index calculated using the investment certificates only. This is not surprising as warrants and knock-outs investors are more short-term oriented, leveraged, and more invested in each underlying than investors trading investment certificates. It also explains why there is a higher correlation when using warrants and knock-out products as opposed to all instruments. This confirms our intuitive understanding of these products and leads us to the conclusion that we should use only the leverage products in the index calculation.

When we look at the different ratios, we get the expected result that the buy call ratios and the sell put ratios are negatively correlated with the DAX logreturns, and that the sell call ratios and the buy put ratios are positively correlated with the DAX logreturns. We can even see that the absolute correlation of the order flow ratio using all three product categories can be higher than each of the single order flow ratios, i.e. all

Table 4. Correlation by order volume group. This table presents the correlation results of contemporaneous DAX logreturns and raw sentiment values by order volume group.

	Order Volume Groups			
	0 - 1,000€	1,000€-10,000€	10,000€-100,000€	> 100,000€
sentiment based on order volume	-0.7181 (0.0000)	-0.7693 (0.0000)	-0.6099 (0.0000)	-0.4020 (0.0000)
sentiment based on the number of executed orders	-0.6695 (0.0000)	-0.7748 (0.0000)	-0.6567 (0.0000)	-0.4056 (0.0000)

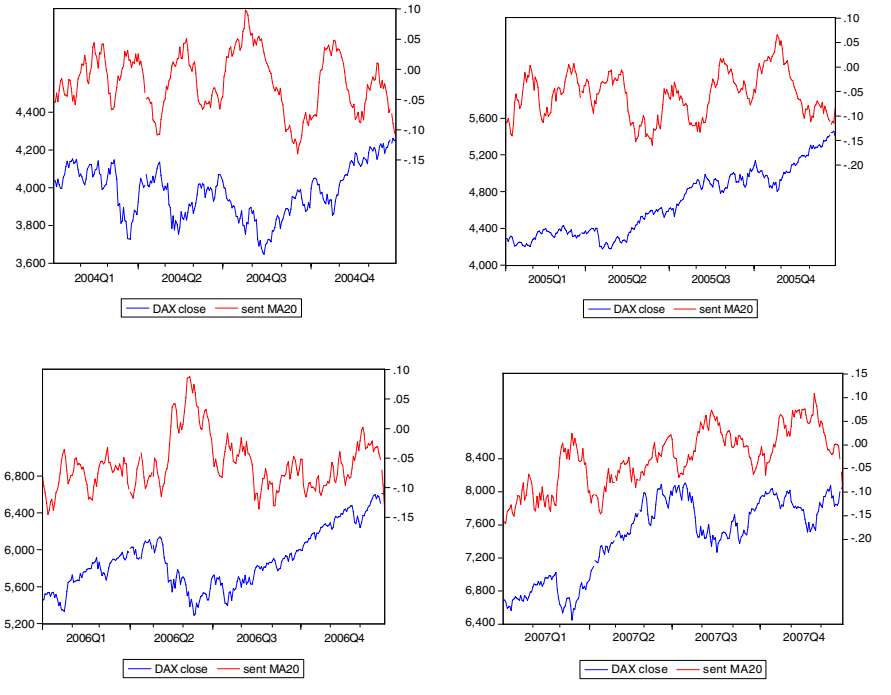


Fig. 1. Plots of the sentiment index and the DAX close prices (2004-2007)

categories together provide a better explanation of market returns than a single ratio. However, the correlations are even higher when we add the order volumes of warrants and knock-out products. Using a combined warrant and knock-out ratio gets us the best correlations with the market returns.

The difference in the correlation between sentiment measures based on order volume and those based on the number of executed orders could arise from a few orders with a high volume that somehow distort the overall sentiment. Orders above 100,000€ could originate from few wealthy individuals or semi-professional traders that are still regarded as retail investors but distort the overall retail investor sentiment measure.

To investigate this issue, we categorize orders into four categories according to their Euro volume: orders under 1,000€, orders between 1,000€ and 10,000€, orders between 10,000€ and 100,000€, and orders above 100,000€. Table 4 presents the results.

As expected, sentiment using all orders below a volume of 100,000€ show high correlation coefficients of -0.72, -0.77, and -0.61 whereas sentiment using orders with an order volume of above 100,000€ 'only' have a correlation coefficient of -0.40. This indeed explains the difference between our order volume based measure and the executed order based measure.

In the following analysis, we therefore report results for a sentiment index that is based on the number of executed orders in warrants and knock-outs, using orders with a volume of up to 100,000€. Figure 1 shows the 20-day backward-looking moving average of the sentiment index and the corresponding time series of the DAX close prices. For purposes of readability, we show the graph in four different pictures for each full year in the sample. Note that the DAX levels change from year to year since during the sample period we had an overall rising market.

We can clearly see that the sentiment index is strongly negatively correlated with the market index, as expressed in the correlations between sentiment values and logreturns.

The next statistical test we perform is a linear regression (see equation 6) with the DAX logreturns as our dependent variable and the leverage product order flow ratios as our independent variables. We use Newey-West standard errors that are robust to heteroskedasticity. We present the model summary for an indication of the amount of variance explained by our model in Table 5.

$$dax\_logreturns = \beta_1\, buy\_call\_ratio + \beta_2\, sell\_call\_ratio + \beta_3\, buy\_put\_ratio + \beta_4\, sell\_put\_ratio \quad (6)$$

The adjusted R-squared value for our model is 0.5813 and indicates that our model explains about 58% of the variance in the DAX close price time series. This value is quite good in our opinion. The sentiment index is intended to be a measure of retail investor sentiment, which at best, is only a component of the pricing process for equities.

Table 5. Linear Regression (OLS) using DAX logreturns as dependent variable and raw sentiment index values as independent variables. We use Newey-West standard errors that are robust to heteroskedasticity. Period is from 6/01/2004 to 30/05/2008. We have included 1,123 observations.

	Coefficient	Std. Error	t-Statistic	Prob.
buy_call_ratio	-0.0473	0.0045	-10.5198	0.0000
sell_call_ratio	0.0428	0.0049	8.6694	0.0000
buy_put_ratio	0.0399	0.0039	10.1779	0.0000
sell_put_ratio	-0.0339	0.0040	-8.3931	0.0000
R-squared	0.5824	Mean dependent var		0.0005
Adjusted R-squared	0.5813	S.D. dependent var		0.0102

4.2 Cross-Sectional Portfolio Formation

Another analysis we perform is to calculate sentiment measures for different underlying stocks in our sample and investigate their relationship with the underlying returns using portfolios formed according to the sentiment values.

In order to be able to calculate sentiment indices for each instrument, we include 100 underlying stocks with the most liquid instruments traded. The underlying stocks comprise companies from different European countries, from the U.S and Japan. Most of the companies, however, are listed in the DAX which is not surprising since EUWAX and most of its customers are based in Germany.

We construct daily cross-sections of the 100 stocks and calculate the sentiment measure as shown in formula (1) for each of the stocks using only orders in products with the respective stocks as underlyings. In addition, we calculate the daily logreturn series for each stock.

We sort all 100 stocks into terciles according to their sentiment value on each trading day t to form three equal-weight portfolios each containing about 33 stocks (the exact number of stocks included in each portfolio may vary since the portfolios are formed based on sentiment values which can be identical for different stocks). For each portfolio we calculate its average next-day return $\sum_i^n ln\,(close_{i,t+1}/close_{i,t})/n$ to test whether the returns of the low sentiment portfolio are different than the high-sentiment portfolio returns. This would imply that it is favorable to invest in those stocks that are in one of the extreme sentiment portfolios rather than investing in the whole market as defined as the portfolio of all 100 stocks.

Table 6. Continuously compounded returns and average sentiment for low and high sentiment portfolios

	Continuously compounded returns		
	All stocks	Low Sentiment Portfolio	High Sentiment Portfolio
2004-2008	0.3520	0.2984	0.5085
2004	0.0495	0.0250	0.0510
2005	0.2460	0.2369	0.2808
2006	0.1197	0.0729	0.1661
2007	0.0521	0.0299	0.0607
2008	-0.1153	-0.0663	-0.0501

	Average daily sentiment value		
	All stocks	Low Sentiment Portfolio	High Sentiment Portfolio
2004-2008	-0.0078	-0.6777	0.6800
2004	0.0266	-0.5975	0.6935
2005	-0.0050	-0.6559	0.6543
2006	-0.0221	-0.7275	0.6941
2007	-0.0383	-0.7472	0.6781
2008	0.0088	-0.6391	0.6803

Table 7. Excess returns for the high and low sentiment portfolios

	Excess returns	
	Low Sentiment Portfolio	High Sentiment Portfolio
2004-2008	-0.0536	0.1565
2004	-0.0245	0.0016
2005	-0.0091	0.0348
2006	-0.0468	0.0463
2007	-0.0222	0.0086
2008	0.0490	0.0652

Table 6 reports the continuously compounded returns for a portfolio containing all stocks, i.e. the 100 most liquid stocks in our sample, and for the low sentiment and the high sentiment portfolios. We also report the average daily sentiment values in each of the portfolios. We show the results separately for each year in the sample to check for consistency.

As can be seen, the low sentiment portfolio's average next-day returns are lower than the average next-day returns of the high-sentiment portfolios which clearly indicates that stocks that are bought by retail investors continue to do better on the next day than stocks that are primarily sold be retail investors.

For the whole period from 2004 to 2008, the portfolio comprising all stocks yielded a return of 0.3520. The low sentiment portfolio yielded a return of 0.2984, clearly underperforming the benchmark. On the other hand, the high sentiment portfolio yielded a return of 0.5085 which documents that this portfolio clearly outperformed the market. The excess returns of the high-sentiment portfolio over the market add up to 0.1065 as opposed to the lower returns of the low-sentiment portfolio which account for -0.0536. Table 7 reports the excess returns for the whole period including the five subperiods.

Results are consistent in the years 2004 to 2007. In the year 2008, however, we observe an outperformance of the low-sentiment portfolio instead of the underperformance shown in the previous years. At this point we do not have an explanation for this although we suspect that retail investor sentiment has to be treated differently in different market periods (2008 was a bearish period whereas most of the other years were bullish periods).

A trading strategy that involves buying the high sentiment portfolio would have yielded excess returns of 15.7% in the whole sample period. One caveat remains, however: In our portfolio construction we have not taken any transaction costs into account. Therefore, possible excess returns could possibly vanish when taking the costs for continuous portfolio changes into account.

5 Conclusion and Future Work

In this paper we explain why investor sentiment is important in equity markets. We develop the theoretical background for a study of investor sentiment and review existing work on the connection between sentiment and market performance. Further we

collect unique transaction data from retail warrant traders and use these data to compute our own sentiment index.

We show that our sentiment index correlates negatively with the underlying market returns and that it can explain a good portion of variance in the return time series. We also present a simple trading strategy which involves a daily construction of a portfolio of high-sentiment stocks. In the sample period we find significant excess returns of this portfolio over the market average as opposed to the underperforming low-sentiment portfolio. This finding is consistent with short-horizon returns documented by related work. We believe that our index, in concert with technical and fundamental analysis, provides insight into subsequent index moves and may in the future provide good prediction as to future equity prices. We do not, however, presume that we can now definitely predict equity price movement nor do we feel we can consistently demonstrate causality.

Next steps involve further analysis of a couple of points which we have mentioned in this paper: First of all, we still have to investigate the different effects of market and limit orders on sentiment development. We try to do this with a deeper understanding of why and when retail investors submit limit, market, or stop orders. Using the order submission time and the limit value, we try to get a grip on the investors' intention and therefore sentiment. Second, we will investigate whether it is possible to create meaningful sentiment indices for short as well as long-term sentiment. Here we are planning to distinguish the different product categories traded as well as the product properties indicating different investor horizons. Third, we will analyze the relationship between risk and sentiment more closely because we believe that the riskiness of the transaction provides insight into the confidence of investors.

We continue to present more statistical analyses that support the usefulness of a sentiment index for retail investors. Moreover, we plan to extend our analysis to future volatility of stock prices. Especially the derivatives market is suited to analyze the relation of investor sentiment and volatility.

References

1. Kyle, A.S.: Continuous Auctions and Insider Trading. Econometrica 53(6), 1315–1336 (1985)
2. Black, F.: Noise. The Journal of Finance 41(3), 529–543 (1986)
3. Fama, E.F.: Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25(2), 383–417 (1970)
4. Shleifer, A., Summers, L.H.: The Noise Trader Approach to Finance. The Journal of Economic Perspectives 4(2), 19–33 (1990)
5. Shleifer, A.: Inefficient Markets – An Introduction to Behavioral Finance. Clarendon Lectures in Economics. Oxford University Press, Oxford (2000)
6. De Long, J.B., Shleifer, A., Summers, L., Waldmann, R.: Noise Trader Risk in Financial Markets. Journal of Political Economy 98, 703–738 (1990)
7. Jackson, A.:: The aggregate behavior of individual investors. Working paper, London Business School (2003)
8. Pan, J., Poteshman, A.M.: The Information in Option Volume for Future Stock Prices. The Review of Financial Studies 19(3), 871–908 (2006)

9. Kaniel, R., Saar, G., Titman, S.: Individual Investor Trading and Stock Returns. The Journal of Finance 63(1), 273–310 (2008)
10. Barber, B.M., Odean, T., Zhu, N.: Do Retail Trades Move Markets? Review of Financial Studies 22(1), 151–186 (2009)
11. Brown, G.W., Cliff, M.T.: Investor sentiment and the near-term stock market. Journal of Empirical Finance 11(1), 1–27 (2004)
12. Qiu, L.X., Welch, I.: Investor Sentiment Measures. Working Paper (2006), http://ssrn.com/abstract=589641
13. Shiller, R.J.: Measuring Bubble Expectations and Investor Confidence. The Journal of Psychology and Financial Markets 1(1), 49–60 (2000)
14. Groves, R.M.: Survey Errors and Survey Costs. Wiley, New York (1989)
15. Beaumont, R., van Daele, M., Frijns, B., Lehnert, T.: On Individual and Institutional Noise Trading, Working Paper (2005)
16. Baker, M., Wurgler, J.: Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance 61(4), 1645–1680 (2006)
17. Tetlock, P.C.: Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance 62(3), 1139–1168 (2007)
18. Das, S.R., Chen, M.Y.: Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science 53(9), 1375–1388 (2007)
19. Das, S., Martinez-Jerez, A., Tufano, P.: E-Information: A Clinical Study of Investor Discussion and Sentiment. Working Paper, Harvard Business School (2003)
20. Ciccone, S.: Does Analyst Optimism About Future Earnings Distort Stock Prices? Journal of Behavioral Finance 4(2), 59–64 (2003)
21. Antweiler, W., Frank, M.Z.: Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. The Journal of Finance 59(3), 1259–1294 (2004)
22. Barber, B.M., Odean, T., Zhu, N.: Systematic noise. Journal of Financial Markets (forthcoming)
23. Kumar, A., Lee, C.M.C.: Retail Investor Sentiment and Return Comovements. The Journal of Finance 61(5), 2451–2486 (2006)
24. Baker, M., Stein, J.C.: Market liquidity as a sentiment indicator. Journal of Financial Markets 7, 271–299 (2004)
25. Hvidkjaer, S.: Small Trades and the Cross-Section of Stock Returns. Review of Financial Studies 21(3), 1123–1151 (2008)
26. Frazzini, A., Lamont, O.A.: Dumb Money: Mutual fund flows and the cross-section of stock returns. Journal of Financial Economics 88, 299–322 (2008)
27. Dorn, D., Huberman, G., Sengmueller, P.: Correlated Trading and Returns. The Journal of Finance 63(2), 885–920 (2008)
28. Lakonishok, J., Shleifer, A., Vishny, R.W.: The impact of institutional trading on stock prices. Journal of Financial Economics 32, 23–43 (1992)
29. Schmitz, P., Glaser, M., Weber, M.: Individual Investor Sentiment and Stock Returns - What Do We Learn from Warrant Traders? Working Paper, Universität Mannheim (2007)
30. Lee, C.M.C., Ready, M.J.: Inferring Trade Direction from Intraday Data. The Journal of Finance 46(2), 733–746 (1991)

Modelling Customer Behaviour in Multi-channel Service Distribution

Dick Heinhuis[1] and Erik J. de Vries[2]

[1] Universiteit van Amsterdam Business School and Hogeschool van Amsterdam Instituut voor Informatica, Weesperzijde 190, 1097 DZ Amsterdam, The Netherlands
d.heinhuis@hva.nl
[2] Universiteit van Amsterdam Business School, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
erik.devries@uva.nl

Abstract. Financial service providers are innovating their distribution strategy into multi-channel strategies. The success of a multi-channel approach and the high investments made in information systems and enterprise architectures depends on the adoption and multi-channel usage behaviour of consumers. We build a model that explains such multi-channel customer behaviour. We take the Technology Acceptance Model as a starting point and extent this model with insights from Expectation Disconfirmation Theory and Customer Choice theory. The model takes into consideration three important elements of multi-channel customer behaviour: trial/adoption, choice and continued use. The model is sophisticated, yet simple and refers to expectations and experiences with concrete channels (which are actually perceived by customers), not to abstract notions like for instance channel configurations.

Keywords: multi-channeling, customer behaviour, Technology Acceptance Model, Expectation Disconfirmation Theory, Consideration set.

1 Introduction

Over the last decades many service providers have been innovating their sales and distribution strategy into multi-channel strategies in which they reach their customers through different channels [4, 30]. Traditional forms of distribution have been supplemented or substituted by direct forms of distribution and electronic distribution channels, like call centers, interorganizational information systems, automatic teller machines, electronic banking, interactive TV, mobile telephony and off course the Internet with its specific applications like extranets, virtual communities, portals and electronic markets. In some service industries, especially in the financial service industry, multi-channel distribution has become the norm [11, 32, 75].

There is a number of reasons for organizations to adopt multi-channel strategies. Most mentioned are customer demand [64], reaching new consumer segments [20,32], cost reduction [84], attracting more profitable customers as multi channel customers seem to spend more [37], improve customer relations [59] and defensive

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 47–63, 2009.

competitive reaction [13]. To reach such goals high investments need to be made in information systems, enterprise architectures, human resources, marketing and communication. Coordination challenges in multi-channeling are high and many information systems and data sources need to be integrated to arrive at a coordinated multi-channel approach [30, 65]. With such high investments and complicated management coordination challenges involved, the big question in multi-channeling is: will customers be reached by the channel configuration and how will they behave in an environment with many channels?

The success of a multi-channel approach depends highly on the adoption of customers of new channels and on multi-channel usage behaviour of consumers [65]. From an information systems perspective this can be seen as a question of multi-channel architecture usage, of the usage of different information systems providing different services to customers through different front office systems and as a question of the adoption and/or substitution of new front office systems. Innovations in the channel configuration by a service provider often need to be followed by influencing customer behaviour regarding the use of channels which is known as the channel migration challenge [64]. Suppliers generally have different reasons to migrate their customers to other channels than they are familiar with. One obvious reason is that they like their customers to migrate to channels with lower costs, but migration to channels with more relationship building or cross/deep selling opportunities might be other reasons. To design effective channel migration strategies we need to understand factors that influence customer channel behaviour. We need to understand the customer's trial and adoption of channels and her choice making between channels.

As new channels are mostly ICT enabled, multi-channel customer behaviour receives growing attention in the information systems field. The adoption of ICT applications, like electronic channels, has been a classical theme in the IS discipline and has been explained with several theories like the Technology Acceptance Model (TAM), (e.g. [29, 38, 54, 56]), Task Technology Fit (TTF) (e.g. [93]) or the Innovation Diffusion Theory (IDT) (e.g. [89]). When it comes to usage and post-adoption behaviour, theories like User Satisfaction (e.g. [26]) or the Expectation Disconfirmation Theory (EDT) (e.g. [8, 9, 10, 63]) are regularly used.

In addition to these theories, several theories on customer behaviour from the marketing literature are used as well. The Theory of Reasoned Action (TRA) and models of the Consumer Decision Making Process are widely used [83]. When it comes to channel usage and post-adoption behaviour, topics like customer retention, customer relationship management and customer satisfaction (e.g. [7]) are studied.

Much of the literature in both disciplines stays within what could be called a *dual-channel* mind-set. General metaphors like 'clicks versus bricks'; 'bricks-and-mortar versus clicks' or 'clicks and mortar' say it all; the emphasis is on the new vis à vis the existing. In papers on click *and* mortar for instance, the click category regularly boils down to the Internet and the mortar category is assumed to represent something like 'traditional retailing'. We should keep in mind however that *mortar* might represent different retail concepts and there is hardly such a thing as *the* Internet (although there are Internet based distribution concepts, like in the financial service industry: home-banking, internet-banking, mobile-banking, financial portals or cybermediaries). Research done from a genuine multi-channel mind-set with research questions on the

level of the multi-channel configuration instead of the level of a (new) channel versus another channel or traditional channels is rather scarce. Up until recently, multi-channeling and its drivers and consequences have received little attention [5], [65] being an exception. The authors provide an overview of multi-channel management challenges emphasizing integration and coordination and explicitly elaborating on customer behaviour in a multi-channel environment as one of the most challenging management tasks. Other examples of recent research on multi-channeling are [31] on knowledge management in multi-channel supply chains and [30] on multi-channel coordination and multi-channel architectures. It can be concluded that genuine multi-channel research is just to begin and that understanding customer behaviour in multi-channel environments is a key research issue. Without knowledge on this issue, managers of financial service institution need to make their decisions on the adoption of multi-channel strategies and the large investments to be made in information systems, enterprise architectures, human resource development and marketing to arrive at it, on unclear expectations on channel adoption and usage.

In this paper, we combine theories from the IS and marketing discipline to arrive at a model that explains multi-channel behaviour in the purchase process of services. Through improving understanding of multi-channel customer behaviour we aim to contribute to the effective adoption of ICT-enabled distribution channels and the design of effective channel migration strategies. We restrict our study to services based on the assumption that by their nature services are more suited for ICT enabled distribution channels than physical products. This is due to the intangible character of service, the information intensiveness of many services and their production/consumption simultaneity in which the co-production of customers requires frequent communication. These characteristics hold especially for financial services. We address the question: which factors influence the use of channels by customers in a multi-channel configuration? To come to an answer, some other questions will be answered first: what causes consumers to try a channel; what causes consumers to continue the use of a channel; and how do customers continuously trade off channels for consumer tasks?

In this paper we provide an overview of theories from the IS and marketing literature that need to be considered in an endeavour to understand multi-channel customer behaviour. Based on these theories we build a conceptual model to explain multi-channel customer behaviour. We split this discussion into three sections: one on understanding trial/adoption of new channels, one on understanding choice between channels and one on continuous use of channels. We end the discussion on the building of the model with an overview of how the complete model works. The next step is defining the constructs of the model based on existing literature. As will be shown this is not possible. The paper ends with some conclusions and a short elaboration on the next steps that will be taken.

2 Multi-channel Customer Behaviour

Multi-channeling comprises three developments [4, 31]: servicing different market segments from different channels (multiple channeling); using different channels to fulfil different distribution functions within the same market segment (composite

Table 1. Theories relevant for the understanding of multi-channel behaviour

THEORY	FOCUS	CONSTRUCTS	MULTI-CHANNEL RELEVANCE
TRIAL/ACCEPTANCE			
Technology Acceptance Model, [23]	Use of information systems	Perceived Ease of Use, Perceived Usefulness	Use of electronic channels
Theory of Reasoned Action (TRA), [36]	Human behaviour in general	Attitude towards behaviour, Subjective norms	Model from psychology, which is the basis for many other models
Theory of Planned Behaviour (TPB), [3]	Behaviour in situations in which the actor has no complete control	TRA constructs, Perceived Behaviour Control	Inclusion of control elements improves use of the model
Innovation Diffusion Theory (IDT), [80]	Adoption of innovations in general	Relative advantage, Compatibility, Complexity, Triability, Observability	Model from social sciences for explaining adoption of innovations
Task Technology Fit (TTF), [39]	Use and success of the use of information systems in organizations	Task Technology Fit, User Performance	Specific model for explaining IS use
Information Richness, [21]	Use of different information channels for different situations in organizations	Uncertainty, Equivocality, Media Richness	The use of information channels combined with the task
CHOICE FROM ALTERNATIVES			
Consideration set (evoked set), [45]	Consumer choice of a brand	Awareness set, Evoked set, Inert set, Inept set	Model for choosing between alternatives
Switching behaviour, [50, 82]	Switching of consumers between service providers	Triggers, Determinants,	Channel switching resembles service switching
SERVQUAL, [72]	Model for measuring service quality	Reliability, responsiveness, assurance, empathy, tangibles	ICT enabled services are by definition services
Multi attribute attitude model [35]	Consumer decision making model	Attributes, weight per attribute	Channel choice can be seen as a decision making process
CONTINUOUS USE			
DeLone & McLean IS Success Model, [25]	Success of an Information System	System quality, Information Quality, Individual impact, Organizational impact	Success of an IS is related to continuous use
Expectancy Disconfirmation Theory (EDT), [68]	The continued use of product/services by consumers	Satisfaction, Expectation, Desire	Satisfaction with channels can explain continuous use

channeling); and adoption of the channel configuration to respond to market dynamics (adaptive channeling). Channels can be complementary, supplementary or a substitute [67]. Multi-channeling can be seen as a coordination problem in which service delivery processes need to be made accessible through multiple channels to different market segments [89]. Several research and management challenges have been recognized, with the issue of how to reach different market segments with different channels and the related issue of understanding customer preferences for channels clearly being one of them [30, 65].

Multi-channel customer behaviour consists of three elements: trial of new channels, continued use of channels and choice making between channels in the context of a set of channels (the channel configuration). Table 1 provides the outcome of our comprehensive literature research for relevant theories regarding trial, continued use and choice of channels in the IS and marketing literature. The first column lists theories that need to be considered in an endeavour to understand multi-channel customer behaviour (with references to the theory's origins). The last column provides reasons for this consideration. Column two and three summarize what the theory aims to explain and what the most important constructs are. The theories are organized around the three elements of customer behaviour that we take into consideration: trial, continued use and choice. Some theories, like for instance TAM and EDT, can be used to explain two of these elements.

In the remainder of this paper we develop a model to explain multi-channel customer behaviour. In the development of this model we consider all theories in table 1. It will become clear that some of these theories look more promising for explaining multi-channel customer behaviour than others.

2.1 Trial of New Channels

In the development of our model we take TAM as a starting point. We do so for a variety of reasons. Originally, TAM has been used to explain user acceptance (both trial and acceptance) of information systems and builds upon the well-established theories about general consumer behaviour, TRA and TPB. TPB is seen as one of the most influential models for the study of human behaviour [2]. Furthermore TAM "has come to be one of the most widely used models in IS, in part because of its understandability and simplicity" [51], has become widely accepted [15, 86] and "is seen as the most influential and widely discussed theory in predicting and explaining the end-user behaviour and system use" [1]. It has "come to occupy a central position in research focused on individual adoption of IT innovations" [61] and some consider it "the only well-recognized theory in IS" [6]. Since the growth of Internet as a distribution channel and the shift in emphasis of many IS research towards electronic commerce issues, TAM has been widely used to explain customer behaviour towards electronic channels, like online banking (e.g. [8, 87]), the internet as a distribution channel in general (e.g. [18, 29, 74]) or mobile commerce [44, 95].

TAM originally includes two variables that cause people to accept or reject an information system:

- perceived usefulness, which is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" [23];
- perceived ease of use, which refers to "the degree to which a person believes that using a particular system would be free of effort" [23].

These two variables are influenced by external variables, like demographic characteristics or personality traits [24], leading to the general model [24, 47], which is used in figure 1 in explaining the attitude towards using.

TAM's wide usage to explain customer behaviour towards electronic channels and it's fundamental origins in psychological theory, makes TAM more suitable as a starting point for research on multi-channel behaviour than other theories on the adoption or use of information systems. Although Innovation Diffusion Theory (IDT) is widely recognized as a general theory explaining the diffusion of all kinds of innovations (and therefore is regularly used in IS as well), there is evidence that IT-related variables have become important (at least as important as traditional IDT factors) in predicting consumer behaviour in relation to ICT enabled channels [65]. Furthermore some factors in IDT resemble factors in TAM. Relative advantage is analogous to the perceived usefulness construct in TAM and IDT's complexity is analogous to TAM's ease of use [49, 90, 95]. A theory like Task Technology Fit (TTF) hasn't received much attention in the IS literature to explain customer behaviour in E-commerce and misses a basis in psychology like TAM has. The Information Richness theory comes to mind as another candidate to explain the usage of a communication channel for customer-provider exchanges. This theory finds its origins in the study of the use of communication media by managers proposing higher usage of 'rich' media for tasks with higher equivocality. The higher the equivocality, the more exchange of subjective views and personal information among managers to define problems and resolve disagreements would be needed [22]. For multi-channeling this would suggest that complex services would be better delivered through personal, non-technological channels. Although the information richness is attractive because of its simple and intuitive construction to use it outside its initial application (e.g. [33, 34, 28, 76]), the empirical results in its own domain have not been favourable [79, 66, 17]. Otondo et al. [71] conclude from a survey among students regarding the choice of the Navy as an employer, that "the notions of media and information richness oversimplify the complex relationships between media, message, and receiver-based communications outcomes. The second is that media richness is a poor predictor of the effects of media type on communications outcomes and media richness, due to its non-monotonic nature across media types, and the weak relationships between media type and media features".

Over the last decade TAM has been used in a large number of studies on E-commerce to explain the use of the Internet as a channel for purchasing goods and services. Several of these studies focused on the trial use of the Internet channel (e.g. [92]). The results indicate that perceived usefulness is the most important factor to explain the intention to use the Internet as a channel. In most of this E-commerce research, TAM constructs have been extended with a number of other constructs, which are almost always related to one of the other theoretical models in table 1, indicating TAM's wide recognition on the one hand and some dissatisfaction with TAM to explain the use of electronic distribution channels on the other hand. Most of TAM's extensions are either related to tasks [19] or to transactions [29, 38, 75]. This is due to the fact that the studied behaviour (E-commerce) is related to the decision making process of consumers and therefore to tasks and transactions. Task characteristic related constructs (like stages of the purchasing process) can be seen as external

variables in the model of figure 1 and transaction related constructs (like trust and perceived risk) can be seen as attitudes towards using the object instead of towards the behaviour [46]. Task and transaction related variables can thus be modeled into the original TAM. Other suggested extensions of TAM are perceived behaviour control (PBC) and the related facilitating conditions and subjective norms [47], which takes TAM back to its roots of the TPB, as has been recommended by Bensabat and Barki [6].

2.2 Choice between Channels

As multi-channeling is a new research area, permanent choice between channels is not widely studied yet. However parallels can be drawn with concepts from customer choice theories from marketing focusing on choice between products, services or brands. One such a concept is the concept of the consideration set (or evoked set) [12, 53]. Consumers limit their choice of brands for instance only to a small number, the consideration set. Because this concept is about choosing between alternatives [16] it is relevant for channel choice as well and has been used for instance in a study on retail selection [85]. The concept of consideration set is intuitively appealing. After all, who considers all possible distribution channels? Is it theoretically conceivable that all customers know all possible distribution channels (given the assumption of bounded rationality) and is it managerial of interest to have all customers perceive all distribution channels? The construct of consideration set means that a channel being not perceived by the customer doesn't belong to the consideration set and therefore will never be used by him/her. We expect the existence of a channel consideration set, which we call the channel choice set, the set from which a customer chooses a channel. We see this set as the set of channels towards which the customer has the behavioural intention to use it. So where TAM originally uses the construct behavioural intention to use, we use the concept channel choice set in our model (see figure 1). Given the introduction of a new channel (e.g. Internet), consumers are influenced by external variables (like marketing efforts by suppliers or personality traits like innovativeness). Based on these influences and the relevant factors for choosing channels (like speed, security), which are translated in TAM as PU and PEOU, the new channel is either added or not added to the channel choice set.

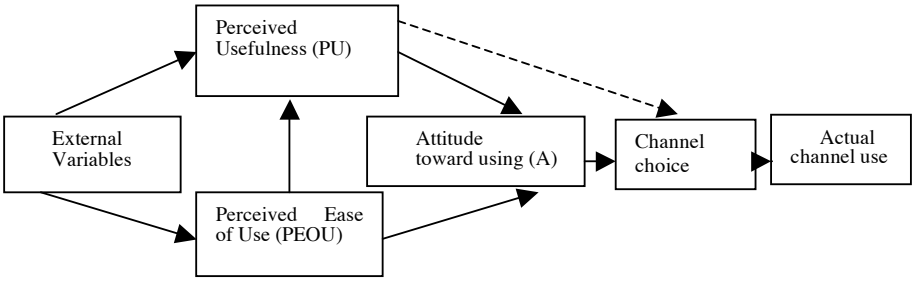


Fig. 1. TAM adjusted with the construct channel choice set

Table 2. The multiattribute attitude model

Attribute	Evaluation	Beliefs		
	(weight)	Channel 1	Channel 2	Channel 3
Risk	+ 3	+ 2	+ 2	+ 1
Advice	+ 1	+ 3	+ 1	- 1
Speed	+ 2	- 1	+ 1	+ 3
Ease of use	+ 4	+ 2	+ 1	+ 1
Total		15	13	12

The concept of channel choice set not only invokes the question of how choice alternatives get in into this set but implies the question of choice between alternatives as well; the question how to evaluate alternatives that are in the choice set. Two basic processes are distinguished [12]: categorization and piecemeal. In a categorization process, the evaluation depends on the category to which an alternative is assigned. (e.g. in choosing drinks for the breakfast, few people consider Coca Cola). In a piecemeal process, possible alternatives are evaluated based on their advantages and disadvantages on important dimensions. Consumers often use cutoffs, which can be seen as a minimal requirement or restriction (e.g. price: if a product exceeds a certain price level, it will not be considered) and may rely on signals (like price, brand name, warranty). In the final step in the piecemeal process the consumer forms an overall evaluation of each alternative's acceptability. In the case of channel choice we expect the evaluation process to be of the latter kind because there are not much categorization opportunities for customers in channel evaluation.

Here, Fishbein's multiattribute attitude model [35] comes to mind. This model is about the determination of an attitude towards an object, which is based on the summed set of beliefs about the object's attributes weighted by the evaluation of these attributes [12, 62]. The model can be explained with an example. Let us assume a consumer wants to buy a life insurance. Her channel set is caused by her experience with channels in former service settings (with other providers or for different services) or as a result of external factors like social influence of her family or marketing campaigns. She considers three channels: telephone, Internet and branch office and weights these against four attributes: the risk of using the channel, the suitability of the channel for provide personal advice, the speed of the channel and its ease of use. Note that such evaluation could be done rather subconsciously. She scores the different channels as in table 2. But how would subsequent use of the channel of her choice influence the next evaluation?

2.3 Continuous Use of Channels

Now that we introduced theories to explain the trail/adoption of channels and the choice making between channels, our next step is to explain the dynamics of channel choice: continuous use and switching.

To explain continuous use (or repeat use) of a channel, past behaviour and satisfaction with that behaviour need to be included [94]. To explain continuous use of an ICT application (like an electronic channel) a first choice might be to use DeLone and McLean's IS Success Model. The model has been adjusted recently to measure E-commerce success as well [27]. One of the shortcomings of the model however is that user satisfaction itself doesn't predict system usage satisfactorily [25, 94]. This is related to the paradox in consumer research in which has been found that satisfied customers may defect [45, 77].

This problem has been addressed in the Expectation-Disconfirmation Theory (EDT), where satisfaction is related to three antecedents: expectation, disconfirmation and perceived performance, EDT has been widely used in consumer behaviour research [8, 14, 68] to study product repurchase and service continuance. It is the basis for research on service quality like the well-known ServQual research [72], which recently has been extended to research on E-service quality [73]. EDT and related models like ServQual have gained recognition in the IS field as well [10] and several studies [8, 9, 46, 63] have used the EDT to explain electronic commerce service continuance. EDT can be seen as a two-stage model with four steps [88]: consumers form an expectation before use; after usage they form perceptions about performance; the perceived performance is assessed against the original expectation and it is determined to what extent the expectation is confirmed. Satisfied customers form a repurchase intention, dissatisfied customers discontinue usage. In our model the (dis)confirmation of expectations effects the channel choice set, which is formed after the actual use of a channel. These changes have been added to the model in figure 1, resulting in our end model in figure 2. When it comes to permanent choice between channels, this model takes into account that confirmation of channel expectations might lead to repeated choice of a channel, while disconfirmation might lead to the opposite.

Figure 2 also shows the construct context specific weight factors (CSW factors). These mediate the relationship between the channel choice set and actual use. Which channel actually is used is not only determined by the outcome of evaluation of attributes in the choice set but by the context of use as well. In a situation of high urgency for example, a channel attribute like speed might be put on a higher weight to than in usual circumstances. Another example might be that a customer decides not the use Internet, although his generally preferred channel, because he perceives the country in which he is in as dangerous. CSW factors could be tasks to be performed (Internet for information gathering, cellular telephone for paying), type of service, time, or location. The context variables are supposed to influence the weighting of the attributes in the consideration set. For instance if the purpose of the use of a channel is information gathering, the attributes information richness, advice, comparisons, etcetera are important, while in a purchasing stage other attributes might be more important. Another example: if one needs cash money late at night, the bank office is not an option, so the attribute availability gets a higher (even an extremely high weighting) and an attribute like safety might be weighted lower than in other circumstances.

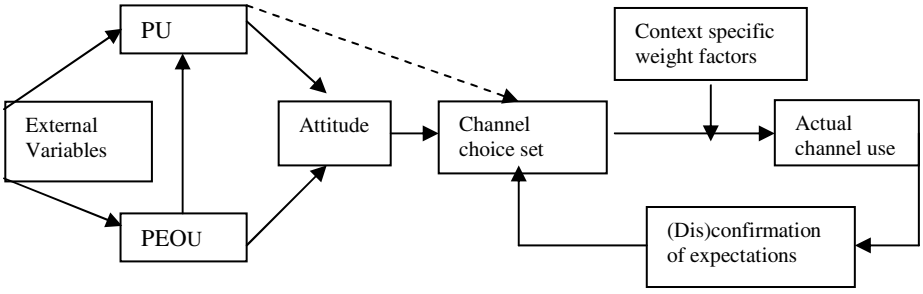


Fig. 2. Model of consumer behaviour in multi-channel service distribution

Alternative models mentioned in table 1 to explain dynamics in channel choice are the ones on switching behaviour. Keaveney [50] for instance comes to eight categories for classifying reasons to switch in service industries (like price, inconvenience, core service failure, response to failure, etc.). Others [43] have found six types of online service failures (delivery problems, Web design problems, payment problems, security problems, low product quality and customer service problems). Although some of these categories and types might apply to channel switching as well, it is hard to tell which ones will be of value and there is no theory of switching behaviour involved. It is just a set of factors which might influence switching behaviour. Furthermore many of these factors resemble disconfirmations with service expectations and therefore confirm the general model of EDT. Roos [82] takes another approach by focusing on the switching process instead of categories of reasons to switch. This approach views the customer relationship into a trigger, a process and a consequence. The trigger gives direction to the switching process and three kinds of switching determinants can be distinguished: the pushing determinant (perceived by the customer as the reason for switching); the swayer, which might speed up or delay the switching decision; and pulling determinants, which cause the customer to return to the service provider. Roos focuses on the complete service experience and switching to other providers, instead of the experience with one channel which is often just a part of the whole service experience and which might lead to switching to just another channel of the service provider instead of switching to a completely other service provider. Therefore also Roos' work doesn't come with explanations convincing us to take another route in understanding dynamics in channel choice than the EDT route.

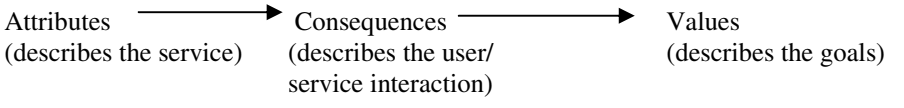
3 Research Approach: Defining the Constructs

In this research the "left part" of the model is taken for granted; that is the channel choice set is taken as given, without researching how the channels became part of this choice set. This means the focus will be first on translating the remaining constructs into a questionnaire, namely the attributes and the weight factors (resulting in the multi-attribute model) and the (dis)confirmation of expectations. The most simple and often used approach is using existing research. Based on existing research it will be researched whether this can lead to a definition of the constructs, which leads to a questionnaire.

3.1 The Attributes

In the literature a large number of attributes, advantages and reasons for using a channel has been mentioned. These items represent different levels of reasons to use. Some are real attributes (characteristics) of the channel (e.g. speed), while others are the results of those attributes (e.g. time saving). And in some cases the mentioned attribute is more or less a desire of the consumer (e.g. efficiency).

This has been captured in a theoretical framework: means-end theory, where the more or less concrete attributes are the means, which lead to more abstract consequences (the perceived benefits or costs), which in their turn lead to the highly abstract personal values, the ends [52]. Values can be defined as "an enduring belief that a particular mode of conduct or that a particular end-state of existence is personally and socially preferable to alternative modes of conduct or end-states of existence" [81]. Where the multi-attribute model focuses on *if* and to *what* degree attributes are important, the means-end framework focuses on *why* and *how* the attributes are important [52]. It can be modelled as follows [40, 78] :



Although the literature on the use of ICT enabled channels has expanded rapidly since the introduction of the Internet, the literature review has made it clear that there is not a clear understanding of the relevant attributes in the choice of a channel. This is partly due to the fact that the reasons for using a specific channel are defined on different levels, which leads to different levels of analysis. When the reasons are defined on attribute level, many of the reasons are related to the specific used channel as can be seen in the Internet related research. However, in understanding consumer behaviour regarding channels the desired level of analysis is the consequences level. The reasons for using or not using a channel should be measured on this level; then it is possible to draw conclusions that go beyond the use of the individual channel and that reach the desired multi- channel level.

3.2 Context Variables

Together with the attributes the context specific weight factors make up the multi-attribute attitude model. The context variables are supposed to influence the weighting of the attributes: what are the most important attributes given a specific situation. For instance if the purpose of the use of a channel is information gathering, the attributes information richness, advice, comparisons et cetera are important, while in a purchasing stage other attributes are more important. Another example: if one needs cash money late at night, the bank office is not an option, so the attribute availability gets a higher (even a very high weighting) and for instance the attribute safety plays no role at all. This is in accordance with consumer decision strategy models in which several strategies are mentioned (e.g. compensatory versus non compensatory strategies, cut-off strategies) and in which the context is seen as important.

Given the fact that in the research only one service in a (in time) similar context will analysed, it can be argued that context variables are of no importance in this research.

3.3 Expectancy, Disconfirmation and Desire

The EDT model explains the satisfaction (and therewith continued use) with the use of the channel by evaluating the actual performance with the initial expectations. These initial expectations are formed by expectations and perceived performance. There is some debate whether disconfirmation should be measured by subtracting expectation from perceived performance or whether it should be measured directly as an independent construct [63]. In research using EDT with regard to an online channel, disconfirmation is usually measured by asking whether the experience was better or worse than expected. This is in line with the origins of the model [68] and more recent publications by the author [69, 70]. In a number of surveys (e.g. [10, 63]) this is done on the attribute level.

4 Defining the Attributes: Laddering

The method that will be used for finding the attributes and the related consequences is laddering. This interviewing format has been used as the interview technique related to the already discussed means-end theory (e.g. [41, 42, 52, 55, 58, 59, 60, 96]).

A total of 30 interviews will be conducted. Respondents are asked about the channels they have used during the last month for conducting financial affairs like paying bills and transferring money. For the used channels they are asked the advantages and disadvantages for using this channel. Based on the mentioned reasons the laddering starts conform the standard procedures [78]. Given the fact that the research interest is at the consequence level, as many different attributes as possible will be used in the laddering (opposed to the much used method of asking respondents which attributes they find most important and use these for laddering).

5 Conclusion

In this paper we developed a model to understand multi-channel customer behaviour by reviewing theories from IS and marketing regarding the most important elements of that behaviour (table 1 provides an overview). With TAM as a starting point for such theory development, the model in figure 2 is strongly anchored in the IS literature. Our literature review shows that extending TAM to new developments general to the field or to particular circumstances has become common practice. Our extensions to understand multi-channel behaviour have resulted in a model that:

- takes into consideration three important elements of multi-channel customer behaviour: trial, continued use/adoption and choice between channels;
- acknowledges the importance of (dis)confirmation of expectations in actual performance of channels, a notion central to EDT;

- reflects the general notion that customers search among limited alternatives in the concept of channel consideration sets;
- shows that channel consideration sets are formed during earlier service experiences or by external factors like social influence or marketing campaigns;
- explains several well known channel migration methods;
- still refers to expectations (like in PU and PEOU) and experiences with concrete channels, not to abstract notions like channel configurations;
- is sophisticated, yet simple.

The next step will be a laddering survey in which the attributes are generated. This will lead to a pilot survey in which the conversion of the constructs into a questionnaire will be tested and if necessary adjusted. The model will be tested in a research among clients of a large international banking and insurance company in the Netherlands. The subject will be the use of mobile banking. The results are analysed with SPSS using the most common statistical methods.

References

1. Ahn, T., Ruy, S., Han, I.: The impact of the online and offline features on the user acceptance of Internet shopping malls. Electronic Commerce Research and Applications 3, 405–420 (2004)
2. Ajzen, I.: Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. Journal of Applied Social Psychology 32, 1–20 (2002)
3. Ajzen, I., Madden, T.J.: Prediction of Goal-Directed Behavior: Attitudes, Intentions, and Perceived Behavioral Control. Journal of Experimental Social Psychology 22, 453–474 (1986)
4. Anderson, E., Day, G.S., Rangan, V.K.: Strategic channel design. Sloan Management Review 38(4), 59–69 (1997)
5. Balasubramanian, S., Raghunathan, R., Mahajan, V.: Consumers in a multi-channel environment: product utility, process utitlity and channel choice. Journal of Interactive Marketing 19(2), 12–30 (2005)
6. Benbasat, I., Barki, H.: Qua Vadis, TAM? Journal of the Association for Information Systems 8(4), 211–218 (2007)
7. Bendoly, E., Blocher, J.D., Bretthauer, K.M., Krishnan, S., Venkataramanan, M.A.: Online/In-Store Integration and Customer Retention. Journal of Service Research 7(4), 313–327 (2005)
8. Bhattacherjee, A.: Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly 25(3), 351–370 (2001)
9. Bhattacherjee, A.: An empirical analysis of the antecedents of electronic commerce service continuance. Decision Support Systems 32, 201–214 (2001)
10. Bhattacherjee, A., Premkumar, G.: Understanding Changes in Belief and Attitude Toward Information Technology Usage: A Theoretical Model and Longitudinal Test. MIS Quarterly 28(2), 229–254 (2004)
11. Black, N.J., Lockett, A., Ennew, C., Winklhofer, H., McKechnie, S.: Modelling consumer choice of distribution channels: an illustration from financial services. International Journal of Bank Marketing 20(4), 161–173 (2002)
12. Blackwell, R.D., Miniard, P.W., Engel, J.F.: Consumer Behavior, 9th edn. South-Western, division of Thomson Learning, Mason, Ohio (2001)

13. Bradley, L., Stewart, K.: The Diffusion of Online Banking. Journal of Marketing Management 19, 1087–1109 (2003)
14. Cadotte, E.R., Woodruff, R.B., Jenkins, R.L.: Expectations and Norms in Models of Consumer Satisfaction. Journal of Marketing Research XXIV, 305–314 (1987)
15. Calantone, R.J., Griffith, D.A., Yalcinkaya, G.: An Empirical Examination of a Technology Adoption Model for the Context of China. Journal of International Marketing 14(4), 1–27 (2006)
16. Chakravarti, A., Janiszewski, C.: The Influence of Macro-Level Motives on Consideration Set Composition in Novel Purchase Situations. Journal of Consumer Research 30, 244–258 (2003)
17. Carlson, P.J., Davis, G.B.: An Investigation of Media Selection Among Directors and Managers: From "Self" to "Other" Orientation. MIS Quarterly, 335 – 362 (September 1998)
18. Cheng, J.M., Sheen, G.-J., Lou, G.-C.: Consumer acceptance of the internet as a channel of distribution in Taiwan – a channel function perspective. Technovation 26, 856–864 (2006)
19. Childers, T.L., Carr, C.L., Peck, J., Carson, S.: Hedonic and utilitarian motivations for online retail shopping behvior. Journal of Retailing 77, 511–535 (2001)
20. Coelho, F.J., Easingwood, C.: Multiple Channel Systems in Services: Pros, Cons and Issues. The Service Industries Journal 24(5), 1–29 (2004)
21. Daft, R.L., Lengel, R.H.: Organizational Information Requirements. Media Richness And Structural Design 32(5), 554–571 (1986)
22. Daft, R.L., Lengel, R.H., Trevino, L.K.: Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. MIS Quarterly, 355 – 366 (September 1987)
23. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 318 – 340 (September 1989)
24. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35(8), 982–1003 (1989)
25. DeLone, W.H., McLean, E.R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3(1), 60–95 (1992)
26. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. Journal of Management Information Systems 19(4), 9–30 (2003)
27. DeLone, W.H., McLean, E.R.: Measuring e-Commerce Success: Applying the DeLone & McLean Information Systems Success Model. International Journal of Electronic Commerce 9(1), 31–47 (2004)
28. Dennis, A.R., Kinney, S.T.: Testing Media Richness Theory in the New Media: The Effects of Cues, Feedback, and Task Equivocality. Information Systems Research 9(3), 256–274 (1998)
29. Devaraj, S., Fan, M., Kohli, R.: Antecedents of B2C Channel Satisfaction and Preference: Validating e-Commerce Metrics. Information Systems Research 13(3), 316–333 (2002)
30. De Vries, E.J.: Multi-Channeling and Front, Mid and Back Office Architectures in the Financial Service Industry. In: IEEE Second International Workshop on Enterprise Applications and Services in the Finance Industry at the 13th European Conference of Information Systems, Regenburg, May 25-28 (2005)
31. De Vries, E.J., Brijder, H.G.: Knowledge Management in Hybrid Supply Channels: a Case Study. International Journal of Technology Management 20(5/6/7/8), 569–587 (2000)
32. Easingwood, C., Storey, C.: The Value of Multi-Channel Distribution Systems in the Financial Services Sector. The Service Industries Journal 16(2), 223–241 (1996)

33. Evans, P.B., Wurster, T.S.: Strategy and the new economics of information. Harvard Business Review, 71–82 (September - October 1997)
34. Evans, P.B., Wurster, T.S.: Getting Real About Virtual Commerce. Harvard Business Review, 85–94 (November – December 1999)
35. Fishbein, M.: An Investigation of the Relationships between Beliefs about an Object and the Attitude toward That Object. Human Relations 16, 233–240 (1963)
36. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention And Behavior. Addison-Wesley Publishing Company, Reading (1975)
37. Ganesh, J., Padmabhuni, S., Moitra, D.: Web Services and Multi-Channel Integration: A Proposed Framework. In: Procedings of the IEEE International Conference on Web Services (ICWS 2004). IEEE, Los Alamitos (2004)
38. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in Online Shopping: An Integrated Model. MIS Quarterly 27(1), 51–90 (2003)
39. Goodhue, D.L., Thompson, R.L.: Task-Technology Fit and Individual Performance. MIS Quarterly, 213–236 (June 1995)
40. Gutman, J.: A Means-End Chain Model Based on Consumer Categorization Processes. Journal of Marketing 46, 60–72 (Spring, 1982)
41. Gutman, J.: Analyzing Consumer Orientations Toward Beverages Through Means-End Chain Analysis. Psychology & Marketing 1(34), 23–43 (1984)
42. Gutman, J.: Means-End Chains as Goal Hierachies. Psychology & Marketing 14(6), 545–560 (1997)
43. Holloway, B.B., Beatty, S.E.: Service Failure in Online Retailing: A Recovery Opportunity. Journal of Service Research 6(1), 92–105 (2003)
44. Hong, S., Thong, J.Y.L., Tam, K.Y.: Understanding continued information technology usage behavior: A comparison of three models in the context of mobile internet. Decision Support Systems 42, 1819–1834 (2006)
45. Howard, J.A.: Marketing Management: Analysis and Planning, Irwin, Illinois (1963)
46. Hsu, M.-H., Chiu, C.-M.: Predicting electronic service continuance with a decomposed theory of planned behaviour. Behaviour & Information Technology 23(5), 359–373 (2004)
47. Jiang, J.J., Hsu, M.K., Klein, G., Lin, B.: E-commerce user behavior model: an empirical study. Human Systems Management 19, 265–276 (2000)
48. Jones, T.O., Sasser, W.E.: Why Satisfied Customers Defect. Harvard Business Review, 88–99 (November/December 1995)
49. Karahanna, E., Straub, D.W.: The psychological origins of perceived usefulness and ease-of-use. Information & Management 35, 237–250 (1999)
50. Keaveney, S.M.: Customer Switching Behavior in Service Industries: An Exploratory Study. Journal of Marketing 59, 71–82 (1995)
51. King, W.R., He, J.: A meta-analysis of the technology acceptance model. Information & Management 43, 740–755 (2006)
52. Klenosky, D.B., Gengler, C.E., Mulvey, M.S.: Understanding the Factors Influencing Ski Destination Choice: A Means-End Analytic Approach. Journal of Leisure Research 25(4), 362–379 (1993)
53. Kotler, P.: Marketing Management, 4th edn. Prentice-Hall, Englewood Cliffs (1980)
54. Koufaris: MApplying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. Information Systems Research 13(2), 205–223 (2002)
55. Kuisma, T., Laukkanen, T., Hiltunen, M.: Mapping the reasons for resistance to Internet banking: A means-end approach. International Journal of Information Management 27, 75–85 (2007)

56. Lai, V.S., Li, H.: Technology acceptance model for internet banking: an invariance analysis. Information & Management 42, 373–386 (2005)
57. Laukkanen, T.: Customer preferred channel attributes in multi-channel electronic banking. International Journal of Retail & Distribution Management 35(5), 393–412 (2007)
58. Laukkanen, T.: Internet vs mobile banking: comparing customer value perceptions. Business Process Management Journal 13(6), 788–797 (2007)
59. Lederer, A.L., Mirchandani, D.A., Sims, K.: The Search for Strategic Advantage from the World Wide Web. International Journal of Electronic Commerce 5(4), 117–133 (2001)
60. Leitner, M., Wolkerstorfer, P., Sefelin, R., Tscheligi, M.: Mobile Multimedia: Identifying User Values Using the Means-End Theory. In: MobileHCI 2008. ACM, New York (2008)
61. Lucas Jr., H.C., Swanson, E.B., Zmud, R.W.: Implementation, Innovation, and Related Themes Over The Years In Information Systems. Journal of the Association for Information Systems 8(4), 206–210 (2007)
62. Lutz, R.J.: Changing Brand Attitudes Through Modification of Cognitive Structure. Journal of Consumer Research 1, 49–59 (1975)
63. McKinney, V., Yoon, K., Zahedi, F.: The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. Information Systems Research 13(3), 296–315 (2002)
64. Myers, J.B., Pickersgill, A.D., Van Metre, E.S.: Steering customers to the right channels. McKinsey Quarterly (4), 36–47 (2004)
65. Neslin, S.A., Grewal, D., Leghorn, R., Shankar, V., Teerling, M.L., Thomas, J.S., Verhoef, P.C.: Challenges and Opportunities in Multi-channel Customer Management. Journal of Service Research 9(2), 95–112 (2006)
66. Ngwenyama, O.K., Lee, A.S.: Communication Richness in Electronic Mail: Critical Social Theory and the Contextuality of Meaning. MIS Quarterly, 145 – 167 (June 1997)
67. Nysveen, H., Pedersen, P.E., Thorbjørnsen, Berthon, P.: Mobilizing the Brand: The Effects of Mobile Services on Brand Relationship and Main Channel Use. Journal of Service Research 7(3), 257–276 (2005)
68. Oliver, R.L.: A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. Journal of Marketing Research XVII, 460–469 (1980)
69. Oliver, R.L.: Cognitive, Affective, and Attribute Bases of the Satisfaction Response. Journal of Consumer Research 20, 418–430 (1993)
70. Oliver, R.L., Rust, R.T., Varki, S.: Customer Delight: Foundations, Findings, and Managerial Insight. Journal of Retailing 73(3), 311–336 (1997)
71. Otondo, R.F., Van Scotter, J.R., Allen, D.G., Palvia, P.: The complexity of richness: Media, message, and communication outcomes. Information & Management 45, 21–30 (2008)
72. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing 49, 41–50 (Fall, 1985)
73. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: A Multi-Item Scale for Assessing Electronic Service Quality. Journal of Service Research 7(3), 213–233 (2005)
74. Pavlou, P.A.: Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. International Journal of Electronic Commerce 7(3), 101–134 (2003)
75. Pavlou, P.A., Fygenson, M.: Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior. MIS Quarterly 30(1), 115–143 (2006)
76. Qiu, L., Benbasat, I.: An Investigation into the Effects of Text-to-Speech Voice and 3D Avatars on the Perception of Presence and Flow of Live Help in Electronic Commerce. ACM Transactions on Computer-Human Interaction 12(4), 329–355 (2005)

77. Reichheld, F.F.: Learning from Customer Defections. Harvard Business Review, 56 – 69 (March/April 1996)
78. Reynolds, T.J., Gutman, J.: Laddering Theory, Method, Analysis, And Interpretation. Journal of Advertising Research, 11–31 (February/March 1988)
79. Rice, R.E., Shook, D.E.: Relationships of job categories and organizational levels of use of communication channels, including electronic mail: a meta-analysis and extension. Journal of Management Studies 27(2), 195–229 (1990)
80. Rogers, E.M.: Diffusion of innovation. The Free Press, New York (1962)
81. Rokeach, M.: The Role Of Values In Public Opinion Research. Public Opinion Quarterly 32(4), 547–559 (1968)
82. Roos, I.: Switching Processes in Customer Relationships. Journal of Service Research 2(1), 68–85 (1999)
83. Schiffman, L.G., Kanuk, L.L.: Consumer behavior. Pearson Education Inc., Upper Saddle River (2007)
84. Sharma, A., Krishan, R.: Clicks Only, Clicks and Bricks, and Bricks Only: Are Retail Salespeople an Important Factor in Choice. Journal of Marketing Management 18, 317–336 (2002)
85. Spiggle, S., Sewall, M.A.: A Choice Sets Model of Retail Selection. Journal of Marketing 51, 97–111 (1987)
86. Sun, H., Zhang, P.: A Methodological Analysis of User Technology Acceptance. In: Proceedings of the 37th Hawaii International Conference on System Sciences. IEEE, Los Alamitos (2004)
87. Sundarraj, R.P., Wu, J.: Using information-systems constructs to study online- and telephone-banking technologies. Electronic Commerce Research and Applications 4, 427–443 (2005)
88. Swan, J.E., Trawick, I.F.: Disconfirmation of Expectations and Satisfaction with a Retail Service. Journal of Retailing 57(3), 49–67 (1981)
89. Tan, M., Teo, T.S.H.: Factors Influencing the Adoption of Internet Banking. Journal of the Association for Information Systems 1(5), 1–42 (2000)
90. Taylor, S., Todd, P.A.: Understanding Information Technology Usage: A Test of Competing Models. Information Systems Research 6(2), 144–176 (1995)
91. Walley, P., Amin, V.: Automation in a customer contact environment. International Journal of Operations & Production Management 14(5), 86–100 (1994)
92. Wang, W., Benbasat, I.: Trust in and Adoption of Online Recommendation Agents. Journal of the Association for Information Systems 6(3), 72–101 (2005)
93. Wells, J.D., Sarker, S., Urbaczewski, A., Sarker, S.: Studying Customer Evaluations of Electronic Commerce Applications: A Review and Adaptation of the Task-Technology Fit Perspective. In: Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS 2003). IEEE, Los Alamitos (2003)
94. Wixom, B.H., Todd, P.A.: A Theoretical Integration of User Satisfaction and Technology Acceptance. Information Systems Research 16(1), 85–102 (2005)
95. Wu, J.-H., Wang, S.-C.: What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. Information & Management 42, 719–729 (2005)
96. Zeithaml, V.A.: Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. Journal of Marketing 52, 2–22 (1988)

A User-Driven Environment for Financial Market Data Analysis

Fethi A. Rabhi[1], Omer F. Rana[2], Adnene Guabtni[1], and Boualem Benatallah[3]

[1] School of Information Systems, Technology and Management, UNSW, Sydney 2052, Australia
f.rabhi@unsw.edu.au
[2] School of Computer Science, Cardiff University, Cardiff CF24 3AA, UK
o.f.rana@cs.cardiff.ac.uk
[3] School of Computer Science and Engineering, UNSW, Sydney 2052, Australia
boualem@cse.unsw.edu.au

Abstract. This paper proposes a software development environment which facilitates the analysis of large financial datasets by end-users. This environment is based on an event-based data model that gives a coherent representation of market activities, particularly high-fequency market and news data. The model makes it possible to define software components and Web services to manipulate entities in the model. The paper also describes a prototype implementation which allows domain experts to compose components and services to build an application. This prototype uses the Triana scientific workflow system to define workflows of existing software components and Web Services. This approach is demonstrated on a realistic case study related to processing both news and financial market data.

Keywords: financial data, Triana, scientific workflow, timeseries analysis, news analysis, event model, composition framework, Web Services.

1 Motivation and Introduction

With the dramatic increase in the speed and availability of computer networks, a significant proportion of all economic activities are now conducted electronically. In particular, the field of financial trading has seen an unprecedented increase in the number of participants and the volumes of trades conducted via electronic markets [27]. As a result, high frequency data has become increasingly available for historical analysis by researchers in fields as diverse as econometrics, finance and accounting. Datasets are often stored in a format suitable for viewing as a spreadsheet. A row usually corresponds to a timestamped piece of information such as the occurrence of a trade, a variation in an instrument's price or an index, the publication of a news story or a market announcement etc. Our research focuses on datasets originating from financial exchanges made available by third party information providers to the research community. For example, SIRCA's Taqtic system [17] provides users with a Web interface for downloading market

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 64–77, 2009.

data according to search criteria such as type of instrument, exchange, time period or frequency (e.g. intraday or interday data). There are several similar portals offering users the possibility to download financial market data (e.g. WRDS [25]).

Analysing such datasets requires expert domain knowledge (e.g. in finance and microeconomics), experience and IT skills [26]. Besides being able to identify suitable data sources and specify the right search criteria, users must be able to perform a wide range of analysis functions (statistical, data mining, language processing) and present results in a suitable form (e.g. through visualisation or report creation). Analysis processes cannot be determined in advance as users tend to perform tasks in a piecemeal fashion: they use a dataset to generate some results, they combine results to build new datsets and the process may be repeated iteratively with datasets obtained with different search criteria (e.g. a different time period). Users also tend to use a variety of tools such as Excel and Matlab to store results and perform routine calculations. When the type of analysis is complex, users spend a lot of effort cleaning, reading and interpreting the data, converting datasets from one format to another, copying some results from one file to another, merging datasets with different semantics, which increases analysis time and the risk of errors.

Building software tools to support such analysis processes is a very challenging task for a number of reasons. Firstly, it requires a deep understanding of financial market operations which skilled software engineers do not possess. Secondly, as data formats vary considerably and little standardisation has taken place in this area, an automatic tool that guarantees information consistency between the different stages of the analysis process would be difficult to develop. Thirdly, it is always preferable to let domain expert users be in control of the analysis process. According to David Shirref, "humans can understand intuitively a great deal about markets. They can scan the complex factors governing a market far more efficiently than any computer" [16].

This paper describes an approach which facilitates the definition and execution of analysis processes by proposing a conceptual model which unifies different views of the datasets, then offering high-level services for manipulating datasets according to these concepts, and finally giving the user the possibility to compose such services in a way that addresses their requirements. It is organised as follows. In section 2 we describe an event-based data model to give a coherent representation of market activities. In particular, we have developed a model that adequately represents high-fequency market and news data. The model makes it possible to define software components and Web services to manipulate entities in the model. Some services will be responsible for querying market and news data from existing repositories (thus acting as event sources), some will implement event processing functions (e.g. filtering, aggregation), some will determine relationships between events (i.e. event patterns) etc. Section 3 contains a case study in which complex analysis business processes can be decomposed into a hierarchy of services. In section 4 we describe a prototype implementation consisting of a user-driven composition framework that allows domain experts to

combine components and services to build an application. The prototype allows for incremental development, thereby giving users the possibility to refine their application in an iterative fashion.

2 A Conceptual Model for Analysing Financial Market Data

In our conceptual model[9], electronic markets (e-markets) can be thought of as distributed event systems, consisting of market events – which capture data attributes such as bid/offer, types of products being traded, volume/number of products, etc; and news events - which capture data related to particular news stories – as published by news organizations such as Reuters. An e-market system may operate in different phases – pre/post-trade and trading. The pre-trade phase involves the submission of buy/sell offers to the market, with the market generating trades during the trading phase. The post-trade phase involves analysing the trades that have taken place, and understanding whether market rules have been followed, and signalling any illegal behaviour. Each of these phases generates different types of market events, and produce different event patterns. In this paper, we are only considering events originating from exchanges, as they represent the type of datasets that are widely available for researchers.

2.1 Basic Event Concepts

In our conceptual model[1], an event of type E which occurs at time t is denoted $e^{(t)} : E$. Each event has a collection of attributes associated with it, hence $e_j = \{a_1, ..., a_m\}$, where each a_i is an attribute with one or more values. An event may be uniquely characterized by its attributes and the time at which it has been recorded. We also consider that the occurrence of such an event is non-deterministic, with no pre-assumed distribution associated with when an event is generated. Our event detection/recording mechanism starts at some time t_0, and then a sequence of events (of type $< E >$) at a given time point $s^{(t)}$ may be specified as:

$$s^{(t)} = e_1 \triangleleft e_2 ... \triangleleft e_n :< E >$$

where e_1 represents the first event that was recorded when the event detection process began, similarly e_n is the last event that was recorded before the current time point t_n. The relation $\triangleleft$ represents an ordering on the recorded events. We make no assumption about when the event was generated, and ordering between events is based on the detection/recording of these events by our system – i.e. when the event was detected. In this context, two events e_i and e_j may have the same attributes and values, but will be treated as different events if they are recorded at different times. Any time stamp on the event itself (as part of its

[1] $x : T$ denotes that an object x is of type T.

attributes) is not used to order events – as time mechanisms (clocks) on external systems that generated the events cannot be guaranteed to be synchronized. In the simple case, we may assume that all events are recorded by the same system, and therefore the time stamp on these events may be used to order them. However, when events may come from different systems, it may not be possible to assume any pre-existing event ordering. Given this representation, we define a grouping of events as:

$$\{e_1 \triangleleft ... \triangleleft e_k\} : \{< E >\}, k \leq n$$

where each group is represented by a sequence of events that has a particular semantics. For instance, each sequence in a group may represent those events within the original sequence that correspond to a "price jump". As an another example, one may consider a group to represent the changes that take place in a particular news story – where all events in such a sequence have some common attributes defined through some ontology. More information about the types of events and their attributes in a financial market context can be found in [9].

2.2 Event-Based Services

Assuming these basic concepts, any type of financial data processing capability can be thought of as a service that operates on events. Frequently used services fall under the following categories[2]:

- **Event Provision Services**: given a query object (of type Q), and the location of an event source (of type S), these services extract a sequence of events that match the constraints identified in the query:

$$EventQueryST \ = \ L \times Q \ \rightarrow \ < Event >$$

- **Event Processing Services**: given a sequence of events, process them to generate another sequence of events. Examples of processing are adding/ removing events, enriching existing events, aggregating events, etc. Some possible services in this category are:
  - *Event enrichment services*: given a sequence of events, enriches the attributes of these events with new information (either supplied from an external source or derived from other attributes)

$$EventEnrichST \ = \ < Event > \ \rightarrow \ < Event >$$

for each event e_j (for all j in the input event sequence) that has attributes $\{a_1, ..., a_m\}$, such a service extends this to $\{a_1, ..., a_m\} \bigoplus \{a_1^{'}, ..., a_m^{'}\}$. The $\bigoplus$ operator involves extending the values associated with one or more existing attributes, and does not involve the addition of new attributes to the existing set $\{a_1, ..., a_m\}$.

[2] $T_1 \times \ldots \times T_n \ \rightarrow \ R$ denotes the type of a function with n arguments of type $T_1 \ldots T_n$ and whose result type is R.

- *Event pattern detection services*: given a sequence of events, identify groups of events that match some user-defined characteristics.

$$EventPatternDetectionST \ = \ < Event > \ \rightarrow \ \{< Event >\}$$

  each group produced as a result is an event pattern instance i.e. an event sequence that matches the pattern being detected. Also, if e is an input event sequence, and p is a pattern instance, we have the property $length(p) < length(e)$.
- *Event classification services*: such a service groups a sequence of events in several classes according to some classification criteria.

$$EventClassificationST \ = \ < Event > \ \rightarrow \ \{< Event >\}$$

  each group produced as a result contains events that belong to the same class according to the classification method selected.

– **Reporting services**: refer to services used to convert event streams or event patterns into a format suitable for text display, download or visualisation.

$$PresentationST = \{< Event >\} \ \rightarrow \ (String \mid HTML \mid Graph)$$

A service implementation can be supported by a local component (i.e. the executable code and data are held on the same machine), or can be remotely invoked – using a Web Service interface. Remote execution would involve transferring object instances from our model to a remote machine. In most cases, existing software modules/packages or Web Services can be adapted through the use of service wrappers.

3 Case Study

This case study demonstrates how researchers have the possibility to define their own analysis processes from a community of services that have been designed according to the conceptual model described in the previous section. At the highest level, researchers are interested in understanding how market events may be influenced by news events, although the exact relationship between them may not be known *apriori*. Market events may be limited in their type, whereas news events may be categorized based on the particular domain of interest to a user.

In this case study, we assume that the researcher is interested in one particular asset over a fixed period and that all market and news data related to this period is available. The analysis process is conducted in two distinct stages respectively called *price jump detection* and *news summarization*, each stage containing several variations. In the rest of this section, we describe each analysis stage (and its variations) as an application of different types of services.

3.1 Price Jump Detection

In the first stage, asset prices are analysed over the period of interest, looking at *exceptional* fluctuations in prices or *price jumps*. Being able to distinguish between jumps and continuous sample path price movements is important as it has implications for risk management and asset allocation [3]. Determining when a price jump has occurred is influenced by the volatility of the market, and the type of asset being considered. It is possible to utilize statistical approaches, based on a time window of recorded asset prices (using the quadratic variation process [1]), to calculate jumps in asset prices. Much of this work relies on the assumption of a continuous price process – whereas in our work we are primarily considering a sampling of this process through market events. Our approach therefore approximates this continuous market process through a discrete time process – generally through ordered events.

Price jumps can be measured in a number of possible ways including measuring sums of squared returns over some time interval [1], use of central limit theorem for realized variances in prices, use of parametric models, use of non-parametric, Markovian analysis and Monte Carlo based finite sampling. Researchers select an appropriate technique depending on the state of the market at a particular time and their particular expertise. In some cases, researchers need to experiment with different techniques before finding a suitable one. Our model supports "plug-and-play" experimentation as each technique is supported by a component of type $EventPatternDetectionST$ (thereby allowing a user to analyse the same data using different algorithms concurrently):

$$priceJumpDetectionService : EventPatternDetectionST$$

Each sequence of events returned by the service corresponds to a price jump determined according to the technique implemented. Each technique may have additional parameters specified by the user.

3.2 News Summarization

Price jumps are, in the first instance, only used to identify particular time periods of *interest* that require further analysis. In the second stage of the analysis process, the researcher tries to identify which categories of news have had an impact on prices. There are many techniques that can be used to analyse and group news stories. In this case study, we assume researchers want to use summarisation techniques [23]. Given news events that are tagged with specific keywords, event summaries are created by grouping together events with similar tags according to a user-supplied ontology. In our model, a summarization technique is supported by a component of type $EventClassificationST$:

$$summarizationService : EventClassificationST$$

Each sequence of events returned by the service represents news events with "similar" tags. The summarization service relies on an ontology – therefore the choice of an ontology impacts the obtained results.

To be effective, the summarization service assumes that appropriate tags and ontologies have been defined and applied to the news events sequence supplied as input. One key theme in the current project is also to investigate how ontologies can be used to provide news summaries at different levels of granularity. Significant work already exists about tagging news events, such as NewsML [11], Semantic Web related efforts such as "Calais" (part of the Sekt project) [15] and work by Sanchez-Fernandez et al. [14]. There are also existing efforts that try to identify connectivity between news stories, and to visualize the outcome [12]. Our model captures these variations in the form of event enrichment services. A technique to add tags to news events is supported by a component of type *EventEnrichST*:

$$computeNewsTagsService : EventEnrichST$$

Therefore, there are many variations in the way news summaries are determined. Based on the output, researchers have the choice between using different ontologies, tag generation schemes and summarization techniques. In the implementation described next, we illustrate one possible analysis scenario.

4 Implementation

In this section we describe a prototype implementation using the Triana [20] scientific workflow system. To address the requirements of the case study, a number of software components have been developed and assembled in a way that expresses various analysis processes through Triana. We first provide a general introduction to Triana followed by a description of each supported process.

4.1 Using Triana as a Service Composition Tool

Triana was initially developed by scientists in GEO 600 [7] to help in the *flexible* analysis of data sets generated from scientific instruments, and therefore contains many of the core tools needed for one-dimensional data analysis, along with many other toolboxes that contain units for areas such as signal processing (e.g. FFT, Spectral analysis) and text processing (e.g. string comparison). All in all, there are around 500 units within Triana covering a broad range of applications. Further, Triana is able to choreograph distributed resources, such as Web Services, to extend its range of functionality. The Web Service-based algorithms have also been added to Triana to support data mining [18]. Triana may be used by applications and end-users alike in a number of different ways [19]. For example, it can be used as a: graphical workflow composition system; a data analysis environment for image, signal or text processing; as an application designer tool, creating stand-alone applications from a composition of components/units; and through the use of its pluggable workflow representation architecture, allowing third party tool and workflow representation such as WSDL and BPEL4WS. A workflow graph in Triana is encoded in XML, and can be mapped to a BPEL representation, for instance.

The Triana user interface consists of a collection of toolboxes containing the current set of Triana components and a work surface where users graphically choreograph the required behaviour. The modules are later bound to the services that they represent to create a highly dynamic programming environment. Triana has many of the key programming constructs such as looping (do, while, repeat until etc.) and logic (if, then etc.) units that can be used to graphically control the dataflow, just as a programmer would control the flow within a conventional program using specific instructions. Programming units (i.e. tools) include information about which data-type objects they can receive and which ones they output, and Triana uses this information to perform design-time type checking on requested connections to ensure data compatibility between components; this serves the same purpose as the compilation of a program for compatibility of function calls.

Triana has a modularized architecture that consists of a cooperating collection of interacting components. Briefly, the thin-client Triana GUI connects to a Triana engine (Triana Controlling Service, TCS) either locally or via the network. Under a typical usage scenario, clients may log into a TCS, remotely compose and run a Triana application and then visualize the result locally – even though the visualization unit itself is run remotely. Alternatively, clients may log off during an application run and periodically log back on to check the status of the application. In this mode, the Triana TCS and GUI act as a portal for running an application, whether distributed or in single execution mode.

4.2 Specifying a Price Jump Detection Process

The first process developed consists of detecting price jumps. To extract market data, we developed a Web Service called `MarketQueryService` that returns market data (e.g. trades, quote and market depth) according to a number of criteria specified in a (`MarketQuery` : Q) object:

$$\texttt{MarketQueryService} : \mathit{EventQueryST}$$

Generating a query object through the graphical user interface is achieved through the `MarketQueryGen` component. (`DataSetLocation` : L) is a component that supplies the source of market data but this is restricted to SIRCA's Taqtic system in this prototype. The market query service returns a stream of events from the specified source that correspond to the query search criteria.

We also developed a `PriceJumpDetection` module in Java which implements a simple time series analysis based on conditional forecasting. The module makes use of the times series properties of the price series based on the estimation of an autoregressive model. For an autoregressive polynomial of order 1, the process reduces to a simple first order autoregression in the form,

$$y_t = c + \phi y_{t-1} + \epsilon_t$$

ϵ_t is the part of the dynamics not captured by the information available to the user, i.e., all past values of y and other variables. In this model ϕ, c are

unknown to the researcher. We estimate these parameters using either ordinary least squares or maximum likelihood estimation[8]. Given these estimates, we define a *Jump* as any price value above a number of standard deviations from the process long run estimated mean $\widehat{\mu}$. This is similar to constructing confidence intervals around the mean and characterizing the prices outside this confidence region as potential jumps. The number of standard deviations is determined by the probability tails of a *standard normal* distribution. For example, a 95% confidence interval corresponds to 1.96 standard deviations. We used both 90% and 95% levels as they are standard values among researchers. However, the user can vary this parameter and the module can be adapted to internally calculate the length of the confidence interval. The algorithm based on the unconditional moments is as follows, for $t = 1, .., n$

$$if\ y_t < \widehat{\mu} - z_{\alpha/2}\widehat{\sigma}_y\ or\ price_t > \widehat{\mu} + z_{\alpha/2}\widehat{\sigma}_y\ then\ Jump$$

where $z_{\alpha/2}$ is the number of standard deviations associated with $(1-\alpha)100\%$ confidence level. We performed simulation where we generated a series with an autoregressive structure of known order. We then injected into the series a number of jumps in the price. The module did detect all the jumps that are outside the confidence intervals and outside the forecast intervals.

$$\texttt{PriceJumpDetection} : EventPatternDetectionST$$

Different algorithms can be built to account for different time series dynamics. Models that account for unobservable price changes using *ordered Probit model* specifically cater for high frequency data. Duration models are more concerned with the time interval between trades to account for intraday activity that is missed in the standard low frequency time series modeling (see Tsay [24] for a thorough exposition).

Finally, the components `PJString` and $\texttt{PJVisualiser} - \texttt{HTML}$ allow the results to be visualised in comma separated values (CSV) or HTML format respectively. Figure 1 illustrates the final price jump detection process expressed in Triana.

4.3 Specifying a News Analysis Process

Before news can be analysed, they have to be extracted from the archive. We developed `NewsQueryService` as a Web service that extracts news according to a number of criteria specified in a $(\texttt{NewsQuery} : Q)$ object.

$$\texttt{NewsQueryService} : EventQueryST$$

As previously mentioned, this service takes a $(\texttt{DataSetLocation} : L)$ as an argument but this is fixed in our prototype (SIRCA's news archive only). The news search query is constructed from the results of the previous workflow using the `PJ2NewsQuery` software component. The user is offered different options in filtering the price jumps that are of interest through the `PJFilteringOptions` component.

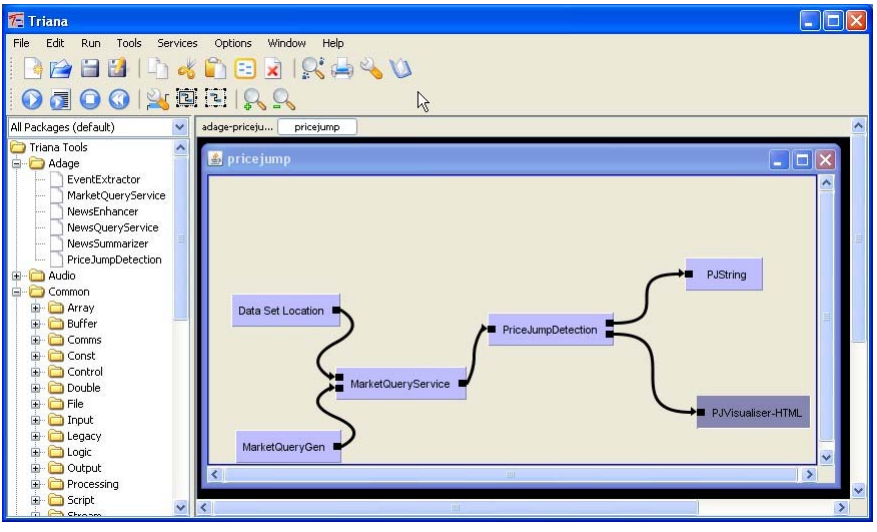


Fig. 1. Price Jump Detection Workflow

The news query service returns news events with very limited tags (Reuters topic codes only). For this reason, we developed a `NewsEnrichment` module which extends the tags of each news event by new tags extracted from the news story. This is based on a simple text analysis technique. In the future, we plan to integrate external Web services such as Thomson-Reuters' Calais [21] to provide a richer (and more appropriate) set of tags.

$$\texttt{NewsEnrichment} : \mathit{EventEnrichST}$$

The `NewsSummarizer` module builds news summaries according a technique proposed in [13]. The ontology is specified in an `OntologyLocation` object which can be modified by the user through an external tool (Protégé [22] in this prototype).

$$\texttt{NewsSummarizer} : \mathit{EventClassificationST}$$

Finally we developed some event presentation modules to visualise streams of events or event patterns:

$$\begin{aligned}
&\texttt{PJVisualiser} - \texttt{HTML} &&: \mathit{PresentationST}\\
&\texttt{PJString} &&: \mathit{PresentationST}\\
&\texttt{NewsSummariesCSV} &&: \mathit{PresentationST}\\
&\texttt{NewsSummariesViewer} - \texttt{HTML} &&: \mathit{PresentationST}
\end{aligned}$$

Figure 2 show the previous process with the news summarization process added to it.

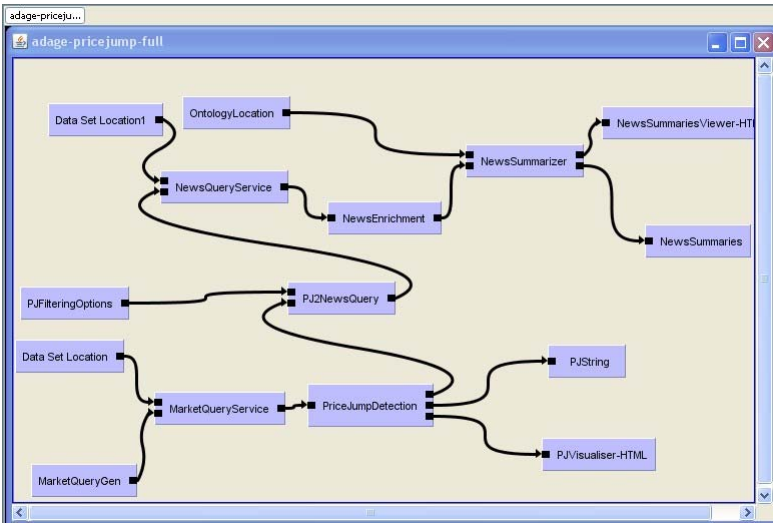


Fig. 2. Price Jump Detection Workflow with News Summarisation – showing all the services

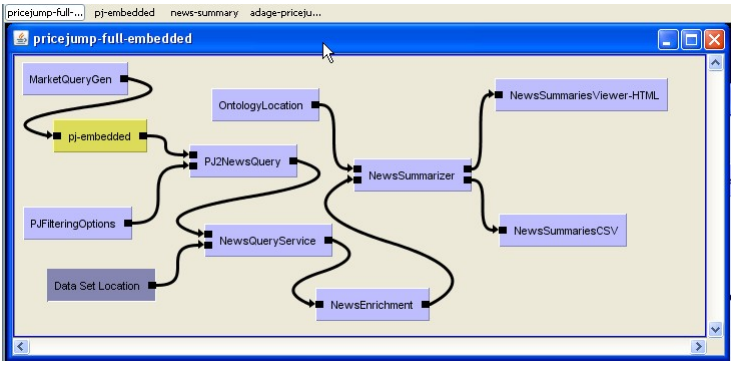


Fig. 3. Price Jump Detection Workflow with News Summarisation – with embedded components – showing a hierarchy

4.4 Specifying the Complete Process with an Embedded Component

As analysis processes can become large and complex, it is possible to abstract out certain parts of the process as embedded components. The ability to decompose complex processes into a hierarchy of components improves readability and gives researchers the ability to make changes more easily. Figure 3 shows the previous workflow with price jump workflow abstracted out as an embedded component. This technique of combining components into a single one also provides the

ability to utilize additional third party libraries which offer a similar capability – provided the interface of the component does not change.

5 Related Work

Related work falls under two categories. The first category comprises a wealth of specialised software tools for analysing financial data available both publicly and commercially. Our aim is not to provide yet another set of tools. Instead, our approach provides an opportunity to integrate several tools working together to support user-defined analysis processes. Through our prototype implementation, we demonstrated that it is possible to empower financial analysts to build complex processes through the reuse and composition of services. The prototype is also built in an *open* manner, allowing many variations and optimisations. On one hand, all components are only accessible through service interfaces meaning that any data source, software component or packaged tool can be integrated into the system. On the other hand, the definition of the Triana workflows is in XML meaning they can be converted to other representations like BPEL.

The second category of related work includes approaches for composing Web services [10], particularly those focusing on user-oriented composition tools [6,5]. The main difference with our work is that such frameworks are too general to be of practical use in such a specialised domain. Our event-based model and categorisation of services provides users with high-level domain-specific abstractions that facilitate the definition of complex analysis processes. The component composition approach used here, also shares similarities with previous work on "Problem Solving Environments" (PSEs) in computational science. In many ways, a PSE is seen as a mechanism to integrate different software construction and management tools, and application specific libraries, within a particular problem domain. One can therefore have a PSE for financial markets [2], for gas turbine engines [4], etc. Focus on implementing PSEs is based on the observation that scientists who used computational methods had to write and manage all of their own computer programs. However computational scientists must now use libraries and packages from a variety of sources, and those packages might be written in many different computer languages. Due to the wide choice of libraries available, navigating this large design space has become its own challenge. In the same way, our approach focuses on the development of an environment similar to a PSE for analysing financial data and relating this to news events.

6 Conclusion

In this paper, we described an event-based conceptual model and an environment for integrated analysis of financial market and news data. So far, this work has focused on building a proof-of-concept prototype that demonstrates the feasibility of the proposed approach. The initial case study illustrated the high level nature of analysis processes defined as workflows and demonstrated that it is possible to easily modify or adapt these workflows to handle different requirements. The

main limitation is that validation has been limited to testing the "plumbing" between the different components of the workflow. Current work-in-progress involves expanding the case study and validating the proposed prototype with finance researchers interested in finding correlations between news and financial data. As part of this effort, the event-based model will be extended and new components and services will be developed thus increasing the system's applicability and effectiveness. The long term goal is to release these components and services to the user community for public use.

Acknowledgements

The work described in this paper is a part of a large project called Ad-hoc DAta Grid Enviroments (ADAGE) which is funded by the Australian Government's DEST Innovation Science Linkage (ISL) Scheme. We are grateful to the Securities Industry Research Centre of Asia-Pacific (SIRCA) for providing the financial data used in this research. We also would like to thank Maurice Peat, Aim Mangkorntong, Rachida Ouysse, Kader Lattab, Quang-Khai Pham , Fletcher Cole and Hairong Yu for helping with the definition of the case study and the implementation of the prototype.

References

1. Back, K.: Asset Pricing for General Processes. Journal of Mathematical Economics 20, 371–395 (1999)
2. Bunnin, F.O.: Design of problem-solving environment for contingent claim valuation. In: Sakellariou, R., Keane, J.A., Gurd, J.R., Freeman, L. (eds.) Euro-Par 2001. LNCS, vol. 2150, pp. 935–938. Springer, Heidelberg (2001)
3. Barndorff-Nielsen, O.E., Shephard, N.: Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. Journal of Financial Econometrics 4(1), 1–30 (2006)
4. Fleeter, S., Houstis, E., Rice, J., Zhou, C., Catlin, A.: GasTurbnLab: A Problem Solving Environment for Simulating Gas Turbines. In: Proceedings of 16th IMACS World Congress, pp. 104–105 (2000)
5. El-Gayyar, M.M., Alda, S.J., Cremers, A.B.: Towards a User-Oriented Environment for Web Services Composition. In: Proc. Fourth Workshop on End-User Software Engineering, WEUSE IV (In conjunction with ICSE 2008), pp. 81–85 (May 2008)
6. Gavran, I., Milanovic, A., Srbljic, S.: End-User Programming Language for Service-Oriented Integration. In: Proc. of 7th Workshop on Distributed Data and Structures (WDAS 2006), CA, USA (2006)
7. EO 600 Gravitational Wave Project, http://www.geo600.uni-hannover.de/ (last Accessed April 2008)
8. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)
9. Mangkorntong, P., Rabhi, F.A.: A domain-driven approach for detecting event patterns in E-markets: A case study in financial market surveillance. In: Benatallah, B., Casati, F., Georgakopoulos, D., Bartolini, C., Sadiq, W., Godart, C. (eds.) WISE 2007. LNCS, vol. 4831, pp. 147–158. Springer, Heidelberg (2007)

10. Milanovic, N., Malek, M.: Current Solutions for Web Service Composition. IEEE Internet Computing, 51–59 (November-December 2004)
11. IPTC Web, NewsML (2008), http://www.newsml.org/ (last accessed May 5, 2008)
12. News.com ontology viewer (2008), http://infosthetics.com/archives/2005/10/cnet_newscom_ontology_viewer.html (last accessed May 2008)
13. Pham, Q.-K., Saint-Paul, R., Benatallah, B.: Time-Aware Content Summarization of Data Streams., Technical Report UNSW-CSE-TR-0722 (December 2007)
14. Sanchez-Fernandez, L., Fernandez-Garca, N., Bernardi, A., Zapf, L., Penas, A., Fuentes, M.: An experience with Semantic Web technologies in the news domain. In: Proc. 4th Int. Semantic Web Conference (ISWC 2005), Workshop Semantic Web Case Studies and Best Practices for eBusiness, Galway, Ireland (2005)
15. The Sekt Project (2008), http://www.sekt-project.com/ (last accessed May 2008)
16. Shirreff, D.: The Human Factor. Euromoney (321), 30–35 (January 1996)
17. The Securities Industry Research Centre of Asia-Pacific (SIRCA), Taqtic On-line System (2008), https://taqtic.sirca.org.au/ (last accessed August 2008)
18. Rana, O.F., Ali, A.S., Taylor, I.J.: Web Services Composition for Distributed Data Mining. In: Katz, D. (ed.) Proc. of ICPP, Workshop on Web and Grid Services for Scientific Data Analysis, Oslo, Norway, June 14 (2005)
19. Taylor, I., Shields, M., Wang, I., Philp, R.: Grid Enabling Applications using Triana. In: Proc. Workshop on Grid Applications and Programming Tools. In: Conjunction with GGF8. Organized by: GGF Applications and Testbeds Research Group (APPS-RG) and GGF User Program Development Tools Research Group, UPDT-RG (2003)
20. Taylor, I., Wang, I., Shields, M., Majithia, S.: Distributed computing with Triana on the Grid: Research Articles. Concurrency and Computation: Practice & Experience 17(9), 1197–1214 (2005)
21. Reuters, T.: Calais Web Service (2008), http://www.opencalais.com/ (last accessed August 2008)
22. The Protégé Ontology Editor and Knowledge-Base Framework (2008), http://protege.stanford.edu/ (last accessed August 2008)
23. Saint-Paul, R., Raschia, G., Mouaddib, N.: General purpose database summarization. In: Proc. International Conference on Very Large Databases (VLDB 2005), pp. 733–744 (August 2005)
24. Tsay, R.S.: Analysis of Financial Time Series: Financial Econometrics. Wiley Series in Probability and Statistics (2002)
25. Wharton Business School, Wharton Research Data Services (2008), http://wrds.wharton.upenn.edu/ (last accessed August 2008)
26. Woodroof, J.: How to Link Web Data. Journal of Accountancy, 55–58 (March 1999)
27. Zwick, S.: Not Your Father's Trading Technology. Futures, 72–75 (March 2005)

Proposing the Relationship between IT Business Alignment and the Business Value of Service-Oriented Architectures in Financial Firms

Daniel Beimborn and Nils Joachim

University of Bamberg, Chair for Information Systems and Services,
Feldkirchenstr. 21, 96045 Bamberg, Germany
{daniel.beimborn,nils.joachim}@uni-bamberg.de

Abstract. What is the business value of Service-Oriented Architectures (SOA) and how can we achieve it? This paper represents a conceptual piece of research which focuses on the impact of IT Business Alignment (ITBA) on the successful implementation of SOA, in terms of its business value. The contribution of this model is predominant in proposing a threefold effect of ITBA on achieving a successful implementation of SOA through the specific strategic needs defined by a particular firm in a specific industry, i.e., the banking industry. As a result, we show that the business strategy moderates the impact of SOA's general potentials on its actual business value and claim that this relationship is further moderated by ITBA, which must be thoroughly considered by practitioners deciding on introducing SOA in their firm.

Keywords: SOA, Service-Oriented Architecture, Business Value, IT Business Alignment, IT Value, Alignment, ITBA.

1 Introduction

Although many firms have decided or are currently considering migrating towards a Service-Oriented Architecture (SOA), as flexible and adaptive fundamentals of their business applications, there is still no answer to one of the most essential questions associated with this new architecture paradigm: *What is the business value of SOA and how can it be achieved?*

Since introducing SOA is a fundamental architectural change for the firm, evaluating its benefits is both critical and difficult. The typical arguments of increased flexibility, reusability etc. primarily lead to benefits, from a long-term strategic perspective. Therefore, the strategic orientation of the firm is a primary decision factor when evaluating the potential business value of implementing an SOA in a particular firm. This, in turn, leads to the requirement of having reached a high degree of alignment of business and IT strategy (i.e., strategic business IT alignment [1]) in the firm before realizing such a significant architectural change.

Consequently, in this paper we focus on the question: *What is the impact of IT business alignment on the successful implementation of SOA?*

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 78–93, 2009.

In this paper, we develop – as the first step of a research project on the business value of SOA – a research model which maps the general potentials, often discussed in the context of SOA, towards the actual strategic business needs of a particular industry and, therefore, examines the role of IT Business Alignment for achieving SOA business value. The objective of the overall research project is to obtain a causal model and to conduct a subsequent empirical study that examines the effect of IT Business Alignment on the successful implementation of SOA in banks and on achieving maximum business benefits from SOA. In this first step, we draw our propositions from previous research in order to conceptualize our research model, which will be the foundation for case studies and a survey-based analysis in subsequent steps.

While it is obvious that the efforts and risks related to introducing SOA have to be overcome with the positive outcomes, the main contribution of our work – and, thus, the differentiator against other quantitative and qualitative works on the SOA business value (like [2-6]) – is the inclusion of a bank's strategic requirements on its IT architecture in order to identify and increase SOA's potential benefits while reducing the related risks and efforts, i.e. the incorporation of ITBA. This approach is consistent with the idea that "the most effective way to cut through the hype surrounding SOA is to consider it in the context of clear and specific cases where generalities are replaced with specific business goals" [7].

Consequently, we focus our conceptualization on a particular industry with unique strategic demands regarding IT. We chose the banking industry for various reasons. First, banks "produce" virtual products, i.e., the IT infrastructure represents the bank's production facility and, consequently, is highly affected by the bank's business strategy. Second, banks are currently facing high levels of competition and, thus, have to flexibly react and adapt both business strategy and IT infrastructure. Third, the banking industry is highly regulated, which leads, e.g., to quite unique requirements in operational risk management and subsequent specific IT demands. Moreover, frequent changes of the regulatory requirements demand a flexible IT infrastructure, again.

The reason why we investigate SOA is that its inherent characteristics, in terms of flexibility are seen to "make SOA-based software far superior to both the customized software supporting proprietary processes and so-called 'off-the-shelf' enterprise software packages" [8]. Therefore, more insights into the relationship of ITBA on the business value of an SOA will help to cope with implementation issues and concerns. Further, strong adoption and implementation trends in the financial industry justify this object of research and promise huge opportunities for empirical research[1].

[1] Schulte et al. did a survey in 2006 asking Germany's 1020 largest banks about their plans to implement SOA [9]. Their results show that "more than 31% of the examined companies are planning an implementation, the implementation is in progress or already finished. Further 23% deem an SOA implementation as interesting" [9].

2 Service-Oriented Architectures and Related Research

"Service-oriented architectures (SOA) is an emerging approach that addresses the requirements of loosely coupled, standards-based, and protocol-independent distributed computing." [10]. Within an SOA, the business functionality, which was formerly present in applications, is now used through the invocation of different components. In order to integrate these components, an enterprise service bus (ESB) is used to facilitate the communication among them. Therefore, "SOA is based on six assumptions: applications are loosely coupled; interface transactions are stateless; interface follows the RPC (remote procedure call) model; interface is message-based; messages use XML data [encoding]; and interfaces may support both synchronous and asynchronous transactions" [11]. In connection with SOA, web services are often used protect the investment in legacy systems of an organization, as the use of relatively simple interfaces and the incorporation of standards, such as SOAP and XML, to deliver standard messaging formats increases the cost-effective reuse of information assets [10].

From a business point of view, SOA promises that "well-executed SOA implementations will bridge the gap between enterprise architecture and business strategy, as companies achieve a closer alignment of IT and the business and, in parallel, implement the robust reuse of existing technology and application code with unprecedented agility and cost effectiveness" [12]. Thus, the SOA paradigm creates a view of IT from a business process perspective, which is contrary to other architectures. Here, alignment is seen as an outcome rather than as a success factor of introducing SOA. This multi-faceted and mutual relationship between ITBA and SOA represents the main motivation for our research.

Various authors emphasize that "services" within an SOA encapsulate *business* functionalities (e.g. [7, 11]) and that "technical services" are just complementary but do not constitute an SOA. Therefore, services are used to create composite applications which support particular business processes. To allow a flexible combination of different services, the services are explicitly defined by their interfaces, which are independent of their concrete implementation. The loosely coupled services are invoked through their communication protocols, which promote location transparency and interoperability. Services are self-characterizing and encapsulate reusable tasks in order to support a fast and cost-saving composition to underpin new or changed business processes [13-15].

SOA had become evident some years ago in academic research. However, most of this research deals with the technical issues, which offers possibilities for business-oriented questions about the business value of SOA [4, 16, 17]. The sparse business-oriented literature mainly dealing with very specific potentials of SOA, such as, exposing information sources as services [18], transitioning large-scale distributed healthcare enterprises to SOA [19], examining the potential of SOA as extensible organizational architecture [16] or during mergers and acquisitions [20], investigating the ability of SOA for organizational integration [21], examining the impact of SOA on supply chains [2-4, 22], assessing the potentials of Web Services as a particular implementation technology [23, 24], or presenting best practices how to introduce SOA [25].

In contrast to these focused research scopes, Müller et al. present "a model that describes the sources of the economic potential of" [17] SOA. Comparing our model with this research model, we do not simply extend by including the efforts and risks associated with SOA, but also show how banks are able to transform SOA's characteristics into economic potentials through ITBA.

3 Model Development

The objective of this paper is to conceptualize a model which qualitatively determines the relationship between SOA and ITBA, i.e. how can ITBA improve the effectiveness (i.e., achieving a business value) and efficiency of an SOA implementation and how does – vice versa – the successful implementation of SOA affect ITBA?

The business value of SOA can be determined along two dimensions: (1) improved business agility, and (2) cost reduction [5]. The first dimension results, for example, from an increased flexibility to adapt to changes in the competitive environment, easier integration of external functionality, better support of a firm by IT [20], or an increased information quality in terms of more accurate or complete real-time information. The second dimension arises from reduced development and maintenance cost due to reusability of functionality, increased scalability of the architecture, as well as easier integration of internal and external systems.

The development of our research model is grouped into four subsections in order to increase the level of complexity with each additional step. The first starts with the efforts and risks associated with the introduction of SOA. After presenting these negative aspects, we continue to enhance the model with SOA's *general* potential to contribute to the business value. The third subsection draws on the foundation of the second subsection discussing the reasons for the *specific* requirements which banks have for their IT architecture. The last section combines the derived requirements of bank's business strategy on their IT architecture (subsection three) with the general potential of SOA (subsection two). Finally, the last section proposes the influence of ITBA on all three before-mentioned sub models in order to further improve the success of the implementation of SOA.

3.1 Efforts and Risks Associated with SOA

Our literature review reveals a couple of issues which an organization has to overcome in order to implement SOA successfully and to achieve the maximum business value of SOA. Therefore, this subsection starts with the downsides of SOA and groups the identified factors into three main categories of efforts and risks: organizational efforts, governance, and technology risks. As the literature demonstrates, we can expect that these aspects have a negative impact on the successful implementation of SOA and the resulting business value of SOA ("SOA success", in short). This is, for example, the case if organizations do not perform necessary organizational changes, develop a sufficient governance, or evaluate the specific technology risks for their own company in advance.

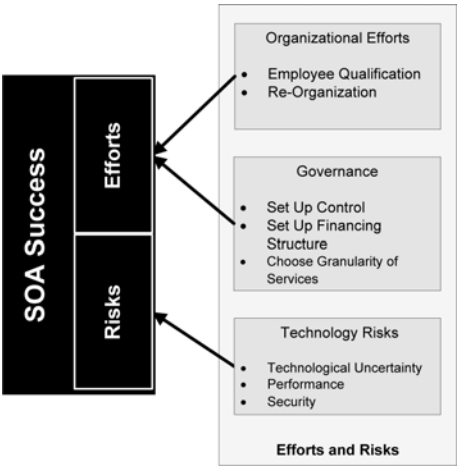


Fig. 1. Efforts and risks and their impact on SOA success

Organizational Efforts. Efforts arising from organizational change can be ascribed to two main aspects: employee qualification and re-organization.

SOA combines new aspects, such as service choreography and service repositories, with existing principles, e. g. encapsulation or modularization, which have to be altogether understood by the system architects requiring additional *qualifications,* such as "the mapping of process tasks to compositions of software-based services" [26]. These qualifications are not classified as risks, as they only represent efforts, which can be overcome with further training and education [23]. Moreover, an adjustment of the organization is necessary to achieve an effective usage of the IT infrastructure [27]. As *re-organization* depends on the employees, it can take a long time until SOA's principles are internalized by the employees and established within the company [28]. One problem, which is associated with this organizational change, is the cultural change an organization has to facilitate in order to establish a new organizational structure [25, 29]. One the one hand, as SOA enables the business units to define processes "IT mangers are worried about losing their influence and position" [25]. However, on the other hand, service orientation may also add new responsibilities to the IT: instead of maintaining applications the IT now has to integrate different service interfaces to satisfy the business needs [23]. Therefore, an organization needs intermediates, who have business knowledge as well as IT skills, in order to coordinate service development. IT can only serve as a promoter for competitive advantage, if this organizational prerequisite has been realized [30]. For example, "the organizational model must be transformed to create differentiated and flexible team-based services" [31]. Consequently, the "new model optimizes cross-business unit operations to deliver objectives, eliminates costly duplication, and flattens management chains. The resulting structure is flexible, agile, and well-orchestrated" [31].

Governance. Establishing an SOA governance is a key factor of successfully implementing SOA [5, 7, 12], but leads to significant *control* efforts (Who is allowed to develop new services?) [12]. Furthermore, the governance requires the specification and periodic adaptation of a *financing and budgeting structure* (Who pays for the infrastructure (implementing and running the SOA itself)? Who pays for developing and setting up a service? Who pays for running the service?) [25, 31].

In order to prevent a decline in governance, the right level of *granularity of services* has to be determined. If services are designed to be too fine-grained, management complexity will increase and inter-service communication can lead to network bottlenecks and to a shortage of parser capacities. By contrast, if services are too coarse-grained, the flexibility of service reusability will be reduced [16]. In addition, Bieberstein et al. point out that for "efficient SOA deployments, it is critical to streamline SOA-related project controls to the bare essentials and promote service reuse" [31] making the level of granularity – as an enabler of service reuse – a key IT governance topic. Additionally, Fricko reports the importance of a governing body with new guidelines, in order "to reinforce the importance of reusable artifacts and ensure widespread participation"[25].

Technology Risks. The technology risks consist of three major aspects: technological uncertainty, performance, and security.

SOA's most often used implementation vehicles are web service standards, such as SOAP and WSDL [18, 32]. Although web services are widely used, the W3C still classifies them as immature [33], which leads to *uncertainty* about the future technical advancement of web services within organizations and possible lack of competence of the employees [16]. Therefore, "the immaturity of some Web service related standards […] is a key concern" [23] for many organizations according to the case studies from Ciganek et al..

Due to the use of web services, in contrast to a method call, additional tasks must be performed: network communication, creation and parsing of XML documents, and packaging of the XML documents into a transport protocol. These additional tasks influence the *performance* of the entire IT infrastructure negatively [34].

Setting up the necessary level of *security* can also be a challenging task [23, 24]. For example, mature security standards, as well as vertical industry payload standards, are lacking [23, 24].

3.2 Contribution of SOA's General Potentials to Business Value

For the assessment of SOA's contribution to the business value we draw on Yoon et al. [5], who have performed a multiple case study in order to discover the benefits of SOA. These benefits contribute to the main aspects of improved business agility and cost reduction.

According to Yoon and Carter [5], the improved business agility consists of: easier integration of systems, better alignment of IT and business, and a quicker response to market change or customer demand (shorter time-to-market), whereas cost reduction consists of lower application development efforts, reuse of existing functions (applications), and lower maintenance costs.

However, what are SOA's potentials which will lead to these business benefits? Adopting and adapting own previous work [35], we propose that the SOA characteristics lead to two basic general potentials: (1) increased support and enhanced management of business processes and (2) opportunities on the technical layer.

Business Process Support. Business processes can be supported more effectively and efficiently due to concepts such as modularization, reuse of functionality, reduced complexity of interfaces, and SOA's integration potential.

Modularization. As one of the basic principles of service orientation, the encapsulation of business functionality in services, modularization is a key potential of SOA. Breaking down business processes into smaller parts, i.e. services, enables higher business agility in the case that business processes have to be changed [36]. Furthermore, due to the increased ability of selective outsourcing, the costs of production can be reduced.

Reuse of Functionality. Due to the reuse of functionality, which is encapsulated in services, a functionality has to be implemented only once but can be used in different business processes [16]. As a result, new applications can be developed in shorter time at lower costs. If considering the high rate of new products in banks, which nevertheless are very interrelated, within their segments (e.g. credit products or investment products), there are huge opportunities for re-using particular functionalities in order to deliver/perform an end-customer product/service.

Moreover, the maintenance costs of the entire system landscape can be reduced, because redundant functions are implemented and thus maintained only once. Additionally, through the use of web services it is possible to implement functionality by a single service, which "could be adapted and modified without compromising the functionality and the stability of already developed/deployed 'consumers' components" [16]. Altogether, the reuse of functionality can lead to an improved business agility and cost reduction.

Reduced Complexity of Interfaces. Compared with an architecture in which all application systems are connected to each other, SOA reduces the number of interfaces as each service has only an interface to the ESB [37]. Due to the reduction in interface complexity, it is possible to maintain and test the existing interfaces more intensely at lower costs. Moreover, changes in business processes can be supported quickly as each service encapsulates a specific part of a business process.

Integration Potential. SOA offers the possibility of programming a service-oriented facade around existing (e.g. legacy) systems in order to make them compatible with the new service-oriented paradigm [16]. This feature allows a firm to support its existing business processes with their existing systems while moving towards SOA. Due to the possibility of integrating systems with service-oriented facades, SOA offers higher protection of invested capital [38], which in turn reduces costs, if information systems of different firms have to work together in order to support a (partly) outsourced business process.

New functions can be more easily integrated with less complex interfaces, promoting easier changes in business processes. Furthermore, new technologies can be adopted

faster, offering enhanced flexibility in order to support the business processes with that technology, which is most appropriate, e.g. in terms of a shorter time-to-market.

Technological Opportunities. SOA offers technological opportunities such as virtualization and grid computing as well as platform independence.

Virtualization and Grid Computing. Despite the fact that practitioners and major software vendors, such as IBM or Microsoft, promote the natural fit between these two topics and SOA [39-41], the academic literature has not investigated this area. Nevertheless, we expect that SOA can leverage the benefits of virtualization and grid computing, due to easier rerouting of services in order to reduce capacity costs. For example, business processes such as securities processing do not cause a balanced workload over time, since there are usually ups and downs in market activity, which could result in performance bottlenecks. Due to the increased scalability resulting from virtualization, the opportunity of rerouting services within an SOA, these bottlenecks can be overcome while reducing fixed costs [41]. Moreover virtualization helps to enhance governance as "service and policy configuration and management are done centrally for each application server, enterprise service bus, or orchestration engine" [39].

Platform Independence. Through the common use of web service standards, one of SOA's features is its independence in programming languages and technical platforms which enables compatibility with other existing technologies [10, 37, 38]. This allows it to offer the possibility of choosing the technology which is best suited in terms of quality and cost for a given problem.

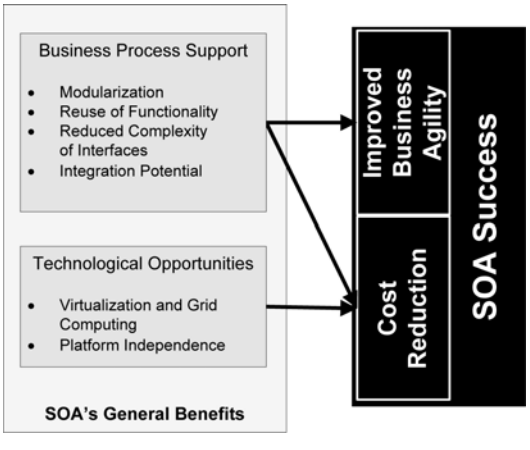


Fig. 2. SOA's contribution to business value

3.3 Deriving Business' Demands on the IT Architecture

General benefits delivered by a new architectural paradigm do not necessarily result in a concrete business value for every firm, e. g., not every technological potential, such as grid computing, will be relevant and beneficial for many parts of the retail banking business.

As the IT infrastructure should support the bank's business strategy, the actual needs fulfilled by the IT infrastructure have to be derived from the latter. For banks, this is even more important since, in contrast to physical goods, financial services are purely virtual. Their design, distribution, and clearing rely completely on information systems. Therefore, the IT infrastructure represents the bank's "production facility" and, consequently, has to adapt significantly to changes in the bank's strategy [42].

Generally, a firm's business strategy is highly affected by its competitive environment and the resulting market opportunities. In many countries, such as the US, Germany, etc., the banking industry shows strong dynamics and high competitive pressure, resulting from globalization and an increasing disintermediation, as more and more direct sellers enter the market [43, 44]. The resulting increase in necessary customer orientation leads both to high flexibility requirements and cost pressures, which, in turn, claim a higher degree of automation, modularization, and outsourcing. Furthermore, the pressure forces banks to engage in mergers and acquisitions which, in turn, lead to unique requirements regarding the IT infrastructures.

In addition, the financial services industry faces a strict and dynamic regulatory environment (e. g. Basel II), which has to be met by the IT infrastructure. Changes in reporting structures, but also the requirements of precisely determining operational risks of singular components within the banks' IT infrastructure, raise the need for a transparent and flexible IT architecture.

As a consequence, typical business strategies in the banking industry – which can be derived from generic types of strategy, such as cost leadership vs. differentiation [45] or prospector vs. analyzer vs. defender [46] – are increased customer orientation (in terms of increased service quality and product customization) [47], industrialization (in terms of increasing automation and modularization of services) [48], outsourcing and shared services [49], enhanced risk policy [50], and mergers and acquisitions [51].

From these strategy groups, we can derive six major requirements regarding the IT infrastructure: (1) since banks manage sensitive data and, therefore, are subject to high regulations, they have comparatively high security requirements for being immune, amongst others, against confidentiality leakages and service delivery interruptions. (2) Banks usually run large legacy systems which cannot be easily substituted. Consequently, an IT infrastructure will always have to ensure their proper integration. (3) Moreover, since banks are increasingly engaged in outsourcing of business activities, such as payments and securities processing, loans management etc. [52], the tight integration of the sourcing providers' systems is necessary in order to reach straight-through processing. This also requires the IT infrastructure to support modular and granular business functions which can be selectively outsourced. (4) On the other side, bundling of processing volumes ("cooperative sourcing"), including those from external firms (i.e. insourcing) is increasingly demanded by banks' business strategies [53, 54]. (5) Customer orientation leads to two primary requirements. First, since banking products are "produced to order", competition leads to high real-time requirements. Even a credit application should ideally be responded to immediately. (6) Furthermore, although banking products are quite homogenous, banks try to diversify their product portfolios and issue new products very frequently. For example, large banks offer more than one thousand new products per year, which all have to be implemented by the underlying information systems. Thus, the IT infrastructure is

required to enable fast support of new products and business processes in order to decrease time-to-market.

Figure 3 shows the resulting relationships between business strategy and the derived strategic IT need.

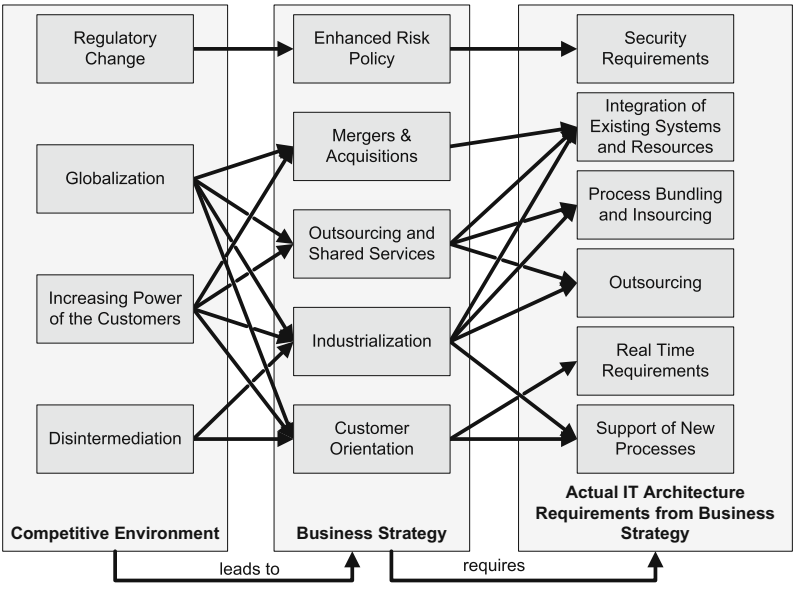


Fig. 3. The competitive environment influences the requirements on the IT architecture

3.4 IT Business Alignment (ITBA)

In order to be able to implement an IT infrastructure which effectively and efficiently supports the business, the bank has to establish a sufficient degree of alignment between the IT and the business on all levels: strategy, people, projects, and structure [55]. During the last two decades, the research community has developed and refined a multi-dimensional concept of IT business alignment in order to consider all aspects of alignment relevant for firm success. Henderson and Venkatraman distinguish between the alignment of the strategic and the structural level of the firm. Both business and IT strategies but also business and IT structures (in terms of processes, skills, routines etc.), have to be aligned in order to generate value from IS. Reich and Benbasat [56] further distinguish between an intellectual and social dimension, the first covering the congruence of IT and business strategies and plans, while the second focuses on a shared mindset and mutual understanding of business and IT managers. Combining the social dimension and Henderson and Venkatraman's alignment on the structural level, subsequent works develop a concept of operational alignment targeting the project and operations level [57, 58]. These authors further distinguish between a cognitive dimension (mutual trust and respect between business and IT people) [59], high communication intensity and quality, and high shared knowledge [60].

Based on the previous section, we will now develop propositions on how this multi-dimensional concept of alignment is interrelated with SOA implementation success and the resulting business value (i.e. SOA success), in order to derive a nomological model which will be used in later empirical work.

We expect four different relationships between ITBA and SOA success, which are illustrated in figure 4.

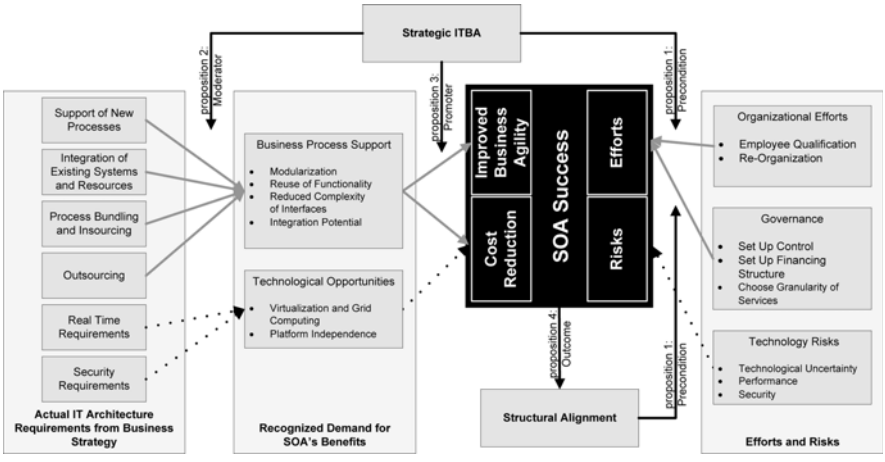


Fig. 4. Moderating impact of strategic ITBA on SOA success. (ITBA influences solid grey arrows, while dotted arrows are assumed to remain unaffected by ITBA.)

IT Business Alignment as Precondition. As explained earlier, implementing an SOA comes along with major risks and efforts (such as organizational re-structuring, setting up an SOA governance, determining optimal service granularity, etc.). It is quite obvious that the implementation goals can only be efficiently and effectively achieved in the case of strong collaboration between business and IT [61] since major parts of the firm (across business and IT) are affected and since major planning problems will be probable if the business gets not involved at a very early stage. This is a typical reason for the failure of large IT projects, in general. Consequently, planning and implementing SOA in a firm requires high alignment on the strategic level and the structural level [55], and from an intellectual as well as from an social perspective [56]. For example, a close relationship between business and IT will also result in higher commitment toward performing the necessary re-organization of the business. In addition, shared knowledge and collaboration between IT and business experts is expected to significantly improve the effectiveness of determining appropriate service granularity [25]. Therefore, ITBA enables efficient SOA implementation, i.e. helps minimizing avoidable efforts and implementation risks.

Proposition 1: All (strategic and structural) dimensions of IT Business Alignment represent a necessary precondition for successfully managing the implementation of SOA and thus reduce the resulting efforts.

IT Business Alignment as Moderator. Strategic ITBA represents an important moderator during the SOA evaluation process, which determines whether and how SOA will have the potential to deliver a business value to the particular bank. Strategic alignment is expected to strongly moderate the process of recognizing the demand for SOA's benefits resulting from an organization's business strategy. For example, in case of high alignment on the strategic level, where the SOA potential is evaluated and implementation decisions are made, the bank can more effectively decide whether and where the general SOA benefits, such as reusability, modularization, and integration potential, would deliver to the firm's specific situation and demands defined by the business strategy. In case of good strategic alignment, SOA's potential will be mapped to the bank's strategy regarding the support of new processes, integration of new systems due to mergers and acquisitions, process bundling, or regarding in- and outsourcing, leading to an effective SOA decision by the business and IT executives [62].

Proposition 2: Strategic IT Business Alignment helps to determine the actual potential for SOA business value in a specific bank.

ITBA as Promoter for Business Value. After deciding to implement SOA in an organization, strategic ITBA will also promote the successful implementation leading to business value from IT. Even if strategic ITBA has helped to identify that SOA is a desirable solution for the particular bank, we suppose that ITBA is also helpful in the implementation phase, as certain aspects, such as shared knowledge between IT and business, will significantly improve the effectiveness of the implemented SOA. As a result, ITBA helps to transform the identified potential benefits into actual business value, because the specific SOA realization of a bank is better aligned with the needs and requirements of the specific business.

Proposition 3: IT Business Alignment promotes the implementation success of SOA in order to achieve maximum IT business value.

IT Business Alignment as Outcome. We hypothesize that the relationship between ITBA and a successful implementation of SOA does not end after the SOA has been established. We believe that a well-developed SOA will positively affect the alignment on the structural level, facilitating collaboration between business and IT on the project and operations level [22, 60, 63]. For example, the development of a new service or the implementation of a new banking product requires an inter-departmental point of view which makes new project management techniques and new – more close – forms of collaboration between business and IT necessary [27]. Yoon and Carter [5] found evidence, in five case studies, that SOA leads to better alignment of IT and business, being explicitly mentioned as a realized benefit. Since both sides now talk about "services", the mutual understanding between business and IT will increase and lead to more effective communication during projects. However, the case studies of Henningsson et al. show inconsistent results: "even though the respondents agreed to that SOA in some cases lead to better communication between business and IS departments the majority of them described the current situation as lacking from such communication" [20]. This counter-argument is also supported by

the results we found in one of our own case studies: most of the interviewed IT managers argued, after an SOA was introduced, that "business people do not think in services" but rather in whole business processes or products.

Proposition 4: Implementing and using a service-oriented architecture will increase the IT Business Alignment on the structural level.

4 Conclusion and Outlook

Our findings propose a multi-faceted relationship between ITBA and achieving business value from an SOA. As this paper only develops a conceptual model, we are going to empirically evaluate the work presented here in a subsequent step. If our propositions can be supported by empirical work, the findings will sensitize practitioners for the critical importance of IT Business Alignment on both the strategic and the structural level for SOA implementation plans and projects. If there is no strategic alignment, SOA's potential benefits might be either overseen (proposition 2) or overestimated, or, if SOA is recognized as a solution, a lack of strategic alignment may result in a suboptimal implementation which does not utilize the full potential business agility and cost reductions (proposition 3), as well as faces higher efforts as necessary (proposition 1). But, if a successful SOA implementation can be realized, alignment on the structural level will in turn improve (proposition 4).

Based on this conceptual model, we will conduct case studies to find evidence for our propositions and to develop a stronger understanding of the differential impact of the different dimensions and levels of alignment on SOA success. In a next step, we will derive a measurement model consisting of indicators for the different constructs in order to formally describe the relations between potential benefits and strategic needs mapped to an SOA context. Finally, we will validate the relevance of the different moderating effects by a quantitative study with the financial services industry to quantitatively validate our model.

References

1. Henderson, J., Venkatraman, N.: Strategic alignment: a model for organizational transformation through information technology. In: Kochan, T., Unseem, M. (eds.) Transforming organizations, pp. 97–117. Oxford University Press, New York (1992)
2. Kumar, S.: Impact of Service-Oriented Architecture Adoption on Electronic Supply Chain Performance. In: Americas Conference on Information Systems 2007, Keystone, Colorado, USA (2007)
3. Kumar, S., Dakshinamoorthy, V., Krishnan, M.S.: Does SOA Improve the Supply Chain? An Empirical Analysis of the Impact of SOA Adoption on Electronic Supply Chain Performance. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, pp. 1530–1605. IEEE Computer Society Press, Los Alamitos (2007)
4. Kumar, S., Dakshinamoorthy, V., Krishnan, M.S.: SOA and Information Sharing in Supply Chain: "How" Information is Shared Matters. In: Twenty Eighth International Conference on Information Systems, Montreal, Canada (2007)

5. Yoon, T., Carter, P.E.: Investigating the Antecedents and Benefits of SOA Implementation: A Multi-Case Study Approach. In: Americas Conference on Information Systems 2007, Keystone, Colorado, USA, pp. 1–11 (2007)
6. Oh, L.-B., Leong, Y.-X., Teo, H.-H., Ravichandran, T.: Service-oriented Architecture and Organizational Integration: An Empirical Study of IT-enabled Sustained Competitive Advantage. In: 28th International Conference on Information Systems (ICIS), Montreal (2007)
7. Tews, R.: Beyond IT: The business value of SOA. In: AIIM E-DOC (21), p. 5 (2007)
8. Merrifield, R., Calhoun, J., Stevens, D.: The Next Revolution in Productivity. Harvard Business Review 86(6), 72–80 (2008)
9. Schulte, S., Repp, N., Berbner, R., Steinmetz, R., Schaarschmidt, R.: Service-Oriented Architecture Paradigm: Major Trend or Hype for the German Banking Industry? In: Americas Conference on Information Systems 2007, Keystone, Colorado, USA (2007)
10. Papazoglou, M.P., Heuvel, W.-J.: Service oriented architectures: approaches, technologies and research issues. The VLDB Journal 16(3), 389–415 (2007)
11. Brandl, D.: SOA explained. Control Engineering 54(8), 22–22 (2007)
12. Laurent, W.: The Importance of SOA Governance. DM Review 17(8), 38–38 (2007)
13. Keen, M., Acharya, A., Bishop, S., Hopkins, A., Milinski, S., Nott, C., Robinson, R., Adams, J., Verschueren, P.: Patterns: Implementing an SOA Using an Enterprise Service Bus. IBM Redbooks (2004)
14. OASIS: Reference Model for Service Oriented Architecture 1.0 (2006) (cited 06/15/2008), http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf
15. Papazoglou, M.P., Georgakopoulos, D.: Service-Oriented Computing. Communications of the ACM 46(10), 25–28 (2003)
16. Baskerville, R., Cavallari, M., Hjort-Madsen, K., Pries-Heje, J., Sorrentino, M., Virili, F.: Extensible Architectures: The Strategic Value of Service-Oriented Architecture in Banking. In: 13th European Conference on Information Systems, Regensburg, Germany (2005)
17. Müller, B., Viering, G., Ahlemann, F., Riempp, G.: Towards Understanding the Sources of the Economic Potential of Service-Oriented Architecture: Findings from the Automotive and Banking Industry. In: 15th European Conference on Information Systems, St. Gallen (2007)
18. Patrick, P.: Impact of SOA on enterprise information architectures. In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data, Baltimore, Maryland, pp. 844–848. ACM Press, New York (2005)
19. Vasilescu, E., Mun, S.K.: Service Oriented Architecture (SOA) Implications for Large Scale Distributed Health Care Enterprises. In: Proceedings of the 1st Distributed Diagnosis and Home Healthcare (D2H2) Conference, Arlington, Virginia, USA, pp. 91–94 (2006)
20. Henningsson, S., Svensson, C., Vallen, L.: Mastering the Integration Chaos Following Frequent M&As: IS Integration with SOA Technology. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences. IEEE Computer Society Press, Big Island (2007)
21. Oh, L.-B., Leong, Y.-X., Teo, H.-H., Ravichandran, T.: Service-oriented Architecture and Organizational Integration: An Empirical Study of IT-Enabled Sustained Competitive Advantage. In: Twenty Eighth International Conference on Information Systems, Montreal, ON, Canada (2007)
22. Vitharana, P., Bhaskaran, K., Jain, H., Wang, H.J., Zhao, J.L.: Service-Oriented Enterprises and Architectures: State of the Art and Research Opportunities. In: Americas Conference on Information Systems 2007, Keystone, Colorado, USA (2007)
23. Ciganek, A.P., Haines, M.N., Haseman, W.D.: Challenges of Adopting Web Services: Experiences from the Financial Industry. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS 2005). IEEE Computer Society Press, Big Island (2005)

24. Ciganek, A.P., Haines, M.N., Haseman, W.D.: Horizontal and Vertical Factors Influencing the Adoption of Web Services. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences. IEEE Computer Society, Kauai (2006)
25. Fricko, A.: SOAs Require Culture Change And Service Reuse. Business Communications Review 36(5), 58–64 (2006)
26. Zhao, J.L., Goul, M., Purao, S., Vitharana, P., Wang, H.J.: Impact of Service-Centric Computing on Business and Education. Communications of the Association for Information Systems 22 (2008)
27. Zhu, K., Kraemer, K.L., Gurbaxani, V., Xin Xu, S.: Migration to Open-Standard Interorganizational Systems: Network Effects, Switching Costs, and Path Dependency. MIS Quarterly 30(Special Issue August 2006), 515–539 (2006)
28. Wong-Bushby, I., Egan, R., Isaacson, C.: A Case Study in SOA and Re-Architecture at Company ABC. In: 39th Hawaii International Conference on System Sciences, Kauai, HI, USA, pp. 1–8 (2006)
29. Bose, S., Walker, L., Lynch, A.: Impact of Service-Oriented Architecture on Enterprise Systems, Organizational Structures, and Individuals. IBM Systems Journal 44(4), 691–708 (2005)
30. Sambamurthy, V., Bharadwaj, A., Grover, V.: Shaping Agility Through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms. MIS Quarterly 27(2), 237–263 (2003)
31. Bieberstein, N., Bose, S., Walker, L., Lynch, A.: Impact of Service-Oriented Architecture on Enterprise Systems, Organizational Structures, and Individuals. IBM Systems Journal 44(4), 691–708 (2005)
32. Zhang, Y., Tanniru, M.: Business Flexibility and Operational Efficiency – Making Trade-Offs in Service Oriented Architecture. In: 11th Americas Conference on Information Systems, Omaha, Nebraska, USA, pp. 2265–2270 (2005)
33. Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., Orchard, D.: Web Services Architecture (2004) (cited 06/15/2008), http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/
34. Stiemerling, O.: Web-Services als Basis für evolvierbare Softwaresysteme. Wirtschaftsinformatik 44(5), 435–445 (2002)
35. Beimborn, D., Joachim, N., Weitzel, T.: Proposing an Instrument for Evaluating the Business Value of Service-Oriented Architectures. In: 3rd International Workshop on Enterprise Applications and Services in the Finance Industry (FinanceCom 2007), Montreal, ON, Canada (2007)
36. Sanchez, R.: Creating modular platforms for strategic flexibility. Design Management Review 15(1), 58–67 (2004)
37. Channabasavaiah, K., Holley, K., Tuggle, E.M.: Migrating to a Service-Oriented Architecture. IBM DeveloperWorks (2004)
38. Lyytinen, K.J., Ren, M.: Building Enterprise Architecture Agility and Sustenance with SOA. Communications of the Association for Information Systems 22, 75–86 (2008)
39. Krill, P.: Tibco Turns to Virtualization for SOA. InfoWorld 28(49), 12–13 (2006)
40. Brodkin, J.: IBM merges virtualization and SOA to ease rollout. Network World 24(11), 18 (2007)
41. Taft, D.K.: Experts See Link Between Virtualization and SOA, Baseline, p. 10 (2007)
42. Winter, R.: Unternehmensarchitektur und Integrationsmanagement. In: Sokolovsky, Z., Löschenkohl, S. (eds.) Industrialisierung der Finanzwirtschaft: Strategien, Management und Methoden für die Bank der Zukunft, Gabler, Wiesbaden, pp. 575–599 (2007)
43. Dombret, A.R., Kern, H.J.: European retail banks - An endangered species? Monitor Group, Rothschild. John Wiley & Sons, Chichester (2003)

44. Hamoir, O., McCamish, C., Nierderkorn, M., Thiersch, C.: Europe's banks: verging on merging. McKinsey Quarterly, pp. 116–125 (2002)
45. Porter, M.E.: Competitive advantage. In: Creating and sustaining superior performance. Free Press, New York (1985)
46. Miles, R.E., Snow, C.C.: Organizational Strategy, Structure, and Process. MacGraw-Hill, New York (1978)
47. Engstler, M., Vocke, C.: Bank&Zukunft 2004-2005. Fraunhofer Institut für Arbeitswirtschaft und Organisation, Frankfurt am Main, Stuttgart, Germany (2004)
48. Lamberti, H.-J., Pöhler, A.: Die Industrialisierung des Backoffice am Beispiel der Etb. In: Lamberti, H.-J., Marlière, A., Pöhler, A. (eds.) Management Von Transaktionsbanken, pp. 3–38. Springer, Heidelberg (2004)
49. Lammers, M., Löhndorf, N., Weitzel, T.: Strategic Sourcing in Banking - a Framework. In: 12th European Conference on Information Systems (ECIS), Turku, Finland (2004)
50. Schierenbeck, H.: Zukunft der Banken – Banken der Zukunft? Industrialisierung der Finanzwirtschaft: Strategien. In: Sokolovsky, Z., Löschenkohl, S. (eds.) Management und Methoden für die Bank der Zukunft, pp. 785–807. Gabler-Verlag, Wiesbaden (2007)
51. Berger, A.N., Demsetz, R.S., Strahan, P.E.: The consolidation of the financial services industry: causes, consequences, and implications for the future. Journal of Banking and Finance 23(2-3), 637–653 (1999)
52. Gewald, H., Franke, J.: The Risks of Business Process Outsourcing: A Two-Fold Assessment in the German Banking Industry. International Journal of Eletronic Finance 1(4), 420–441 (2007)
53. Focke, H., Kremlicka, R., Freudenstein, G., Gröflin, J., Pratz, A., Röckemann, C., West, A.: Tendenz steigend: Transaction Banking auf dem Weg zu Service und Innovation. A.T. Kearney Transaction-Banking-Studie, A.T. Kearney, Frankfurt, Munich (2004)
54. Beimborn, D.: Cooperative Sourcing - Simulation Studies and Empirical Data on Outsourcing Coalitions in the Banking Industry. Gabler, Wiesbaden (2008)
55. Henderson, B.D., Venkatraman, N.: Strategic Alignment: Leveraging Information Technology for Transforming Organizations. IBM Systems Journal 32(1), 4–16 (1993)
56. Reich, B.H., Benbasat, I.: Factors that influence the social dimension of alignment between business and information technology objectives. MIS Quarterly 24(1), 81–113 (2000)
57. Franke, J., Wagner, H.-T., Weitzel, T.: The role of IT business alignment for value creation: a multiple case study among German banks. In: 26th International Conference on Information Systems (ICIS). Las Vegas, NV (2005)
58. Wagner, H.-T., Beimborn, D., Franke, J., Weitzel, T.: IT business alignment and IT usage in operational processes: a retail banking case. In: 39th Hawaii International Conference on System Sciences (HICSS), Koloa, Kauai, HI (2006)
59. Tiwana, A., Bharadwaj, A., Sambamurthy, V.: The antecedents of information systems development capability in firms: a knowledge integration perspective. In: 24th International Conference on Information Systems (ICIS), Seattle (WA), USA, pp. 246–258 (2003)
60. Reich, B.H., Benbasat, I.: Measuring the Linkage between Business and Information Technology Objectives. MIS Quarterly 20(1), 55–81 (1996)
61. Chan, Y.E., Reich, B.H.: It Alignment: What Have We Learned? Journal of Information Technology 22, 297–315 (2007)
62. Lawler, J., Anderson, D., Howell-Barber, H., Hill, J., Javed, N., Li, Z.: A Study of Web Services Strategy in the Financial Services Industry. Information Systems Education Journal 3(3) (2005)
63. Beimborn, D., Franke, J., Wagner, H.-T., Weitzel, T.: The Impact of Operational Alignment on It Flexibility - Empirical Evidence from a Survey in the German Banking Industry. In: 13th Americas Conference on Information Systems (AMCIS), Keystone (CO), USA (2007)

Building the Business Case for SOA: A Study of the Business Drivers for Technology Infrastructure Supporting Financial Service Institutions

Haresh Luthria and Fethi A. Rabhi

Information Systems, Technology & Management, The Australian School of Business
The University of New South Wales, Sydney, Australia
{h.luthria,f.rabhi}@unsw.edu.au

Abstract. Financial service institutions are pursuing organizational agility in the face of an increasingly competitive marketplace, and are consequently looking infrastructure technologies that enable process and infrastructure agility. Service-oriented computing (SOC) appears to provide flexibility and agility, not just in systems development but also in business process management. This paper empirically examines the decision to adopt SOC as an enterprise strategy across fifteen firms, and investigates the business drivers that influence the enterprise adoption of SOA. In doing so, this paper adds crucial empirical evidence to the formal academic literature about the business case for SOA as an enterprise strategy, and lays the groundwork for future work on SOA alignment with business strategy.

Keywords: Service Oriented, SOA, SOC, Business Drivers, Technology Adoption.

1 Introduction

In response to dynamically changing market conditions, financial institutions are increasingly looking for avenues of organizational agility [1]. By virtue of being the underlying enabler of the core business processes, information technology is very critical to achieving this agility [17]. Technology infrastructures built on service oriented computing principles appear to facilitate business process and, subsequently, organizational agility [23]. The paradigm of Service Oriented Computing (SOC) views whole business functions as modular, standards-based software services. The associated Service Oriented Architecture (SOA) establishes a defined relationship between such services offering discrete business functions and the consumers of these services, independently of the underlying technology implementation [28].

There is a great deal of enthusiasm in the industry about this concept but the adoption of SOA by end-user organizations is still in a relatively early stage [32]. Therefore, there is a scarcity of critical research addressing the ability of organizations to realize business value from the adoption of SOA. From a pragmatic perspective, there is widespread recognition of the fact that various organizational issues need to be addressed for the successful implementation of any information technology [19]. What is

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 94–107, 2009.

needed beyond the current research on the technology implementations of SOA, is a focus on the study of the real-world adoption of SOA across the enterprise and the factors that aid or impede such adoptions. This understanding becomes even more critical in the context of financial services institutions since the strategic impact of information technology is very high for financial institutions, and the industry sector is at the leading edge of the adoption curve for innovative technology solutions [25].

This paper empirically examines the use of SOA across fifteen firms – a mix of banks, insurance firms, and service providers - and as part of a broader study, specifically investigates the business drivers for choosing to adopt SOA as the technology infrastructure. Having identified the business requirements that drive organizations to choose SOA as the enterprise infrastructure, the paper then examines the value proposition of SOA that could fulfill these business requirements. In doing this, the paper helps build a business case for the adoption of SOA by financial institutions. The results of the empirical study also provide valuable insights into the factors that influence the real-world adoption of SOA, thus filling a crucial gap in academic literature.

Section 2 provides some background on the concept of service orientation. Section 3 describes the empirical study of SOA adoption across fifteen firms including the data collection and analysis processes. Section 4 describes the results of the empirical study, while Section 5 examines the value proposition of SOA along each of the business requirements identified by the study and articulates the business case for SOA. Section 6 identifies related work in this area, and the final section, Section 7, concludes the paper with a discussion of the contributions and possibilities for future research.

2 Service Oriented Computing

As organizations have evolved with ever-changing technology, and natural organic and acquisition-based growth, the complexity of their infrastructures has increased dramatically, requiring more innovative distributed computing techniques to address their needs [7]. With the increasing diversity of platforms, protocols, and development environments, the need for a higher level of abstraction was recognized as being imperative for the efficient use of existing heterogeneous and/or geographically distributed resources [6]. This need was compounded by the growing business need for communications across diverse domains – for example, across partner or customer systems - for increased business value through strategic partnerships [8]. This gave rise to the concept of services, functional entities whose location and implementation are abstracted from the client or user, to allow the integration and communication of diverse and distributed technology domains.

2.1 What Constitutes a Service?

A service is a business function implemented in software, wrapped with a formal, documented interface that is well known, does not depend on the internal workings of other services, and can be located and accessed by any software agent using standards-based communication mechanisms [28]. These services could be simple services performing basic granular functions such as order tracking or composite services that assemble simple or other composite services to accomplish a broader

modular business task such as a specialized product billing application. As an example, a business flow, such as an online book retail service, could be built using services across multiple service providers pulling together, say, billing services from a partner, and warehousing services from another partner. At a lower level, this could also potentially work for an individual business application say, the ordering of a book being built from tying together simple services such as a book search feature and customer verification.

2.2 Service Oriented Architecture (SOA)

While services manifest business functionality in the service-based computing model, a Service-Oriented Architecture (SOA) provides a framework for the infrastructure to facilitate the interactions and communications between services [28]. An SOA is as an interconnected set of services which in its basic form is a message-based interaction between software agents, each accessible through standard interfaces and messaging protocols. These agents can be service providers or service requesters (clients) interacting with service discovery agencies, and the services in the SOA should be technology neutral, loosely coupled (not tightly integrated into the requester's process), and support location transparency.

SOAs can be thought of as both an architecture and a programming model, more a way of thinking about building software than a software development technique [7]. According to widely accepted definitions of services [11], [28], [36], SOA-compliant architectures exhibit the following four properties:

- Modularity. The services in the architecture are developed as independent modules of functionality, offering well-defined interfaces to their users. The services represent a logical view of discrete business level operations (e.g., customer verification) and are relatively granular or coarse-grained in scope.
- Loose Coupling. This is facilitated by encapsulation of the underlying functionality so that the implementation is logically decoupled from the invoking entity. Services may encapsulate functionality at various levels – from components within an application to components or sub-systems communicating across enterprises – as long as they represent discrete meaningful business functions. This facilitates the composition of these services into complex services and applications.
- Technology neutrality. Services are universally usable by any requester, and communication between services is message based, with the message format being standards-based and platform-neutral.
- Location transparency. The services are self-describing in that they have formal documented interfaces that are well known, and they are easily locatable and accessible over a network. According to [28], the service interface is "known where to be found not only by agents who designed the services but also by agents who do not know about how the service has been designed and yet want to access and use it".

In the business world, SOAs may be viewed as application architectures "within which all functions are defined as independent services with well-defined invokable interfaces, which can be called in defined sequences to form business processes" [7]. Decomposing this definition, services can represent simple business capabilities (e.g., address validation), complex business transactions built from simple business

capabilities (e.g., placing purchase orders), or broader system functions (e.g., user authentication). In addition to this essential attribute of granularity, services are "independent" in that they meet the requested need but their internal implementations are irrelevant to the business process. Services are "invokable" in that they can be used from within or across enterprise boundaries by users across diverse platforms.

Service consumers can, ostensibly, weave together business services, with no knowledge of the underlying technical service implementations, changing existing business models where the business process are invariably tightly tied to specific technology solutions [34] . The pulling together of simple, basic, core services to form a complex business process in the SOA world is called service composition [29]. To achieve this seemingly easy equilibrium of the composition of services into higher level applications, however, the underlying technical infrastructure needs to provide stateless or context independent technical services with relevant metadata that describes what the service does and how to interact with it. These stateless technical services can then be bound together on demand to form business services, using generic communications infrastructure and the contextual metadata [26], [35].

2.3 Technology Implementations of SOA

SOAs may be implemented using any appropriate technology as long as the services in the SOA framework support the basic principles of service-oriented computing - modularity, loose coupling, technology neutrality, and location transparency.

Hub-centric message driven systems used by enterprises today are widely viewed as precursors to what is now called SOA [12]. Message oriented middleware (MOM) allowed systems to build modules that communicated over a messaging infrastructure, forming a loosely coupled system and allowing for a level of abstraction. Component based software programming models, such as DCOM, CORBA, and Enterprise Java Beans (EJB), are also earlier attempts at building loosely-coupled object-based systems [20]. These system models that were precursors to the service concept continue to be viable options of implementing SOA in certain situations [2]. When it comes to wider use across organizational boundaries, however, the use of these models are hampered by the lack of uniform standards and support from major software vendors [5].

The next step on the implementation chain was made possible by the ubiquitous information channel - the Internet. Web services are essentially the deployment of a service-based computing model over the Internet, and unlike other earlier technology implementations, leverage open Internet standards to facilitate diverse inter-enterprise communication [28], garnering relatively unanimous industry vendor support [20].

2.4 Current State of Industry Adoption

There are a variety of statistics available from various trade magazines and technology analysts relating to the adoption of SOA in the industry, all generally indicating the widespread acceptance of SOA. As early as 2003, Gartner [27] had predicted that over time lack of SOA would become a competitive disadvantage for most enterprises. When it comes to adoption of SOA in the industry today, there are a variety of statistics available from various trade magazines and technology analysts, all generally indicating the adoption of SOA as widespread. Surveys of professionals

worldwide [9], [31], [32], indicate that knowledge and awareness of SOA amongst the IT professional community is "significant", with most companies "doing something related to SOA". The associated reports conclude that the spread of SOA is "almost inevitable". This mirrors the general optimism in trade journals and magazines, indicating that SOA, and specifically Web Services, is the popular choice for businesses looking for flexible systems development.

The major technology vendors appear to have invested significant effort in promoting SOA, building supporting products and tools, and even publishing related research, with IBM seemingly in the forefront. Its research initiative (Service Science, Engineering, and Management or SSME for short) is a collaborative effort with various universities worldwide to promote multi-disciplinary research in service-oriented computing. HP has recently introduced the Business Technology Optimization (BTO) for SOA, a set of software and services for service management. Microsoft and BEA are also updating their product suites and infrastructures to include service-oriented concepts. In addition to specific products and solutions, software vendors, large and small, have thrown their support behind SOA, working on various cross-vendor initiatives to promote the growth of SOA adoption. The efforts are too numerous to list here, but industry trade journals have an abundance of information on vendor products, and ongoing collaboration efforts across various vendors to promote standards and interoperability for enterprise service infrastructures.

3 The Empirical Study

A case study approach was chosen as the research methodology to study the alignment and adoption of SOA across the enterprise because, according to Benbasat *et al* [3], case studies are "well-suited to capturing the knowledge of practitioners and developing theories from it".

Fifteen firms – a mix of both financial service institutions in the banking and insurance sectors, and software service providers that had a significant number of clients in the financial services industry - were approached to understand their position on SOA. Most of these firms were chosen based on their involvement in industry conferences on SOA which was an indication of their interest in adopting SOA. A few, however, were chosen on an opportunistic basis leveraging a network of contacts. Table 1 describes the industry sector and profile of the firms interviewed as well as the designation of the interviewees.

Semi-structured interviews were conducted with business managers, enterprise architects, and CIOs/CTOs of 13 (thirteen) of these firms. A broad set of questions addressing specific areas of discussion (technology strategy, business drivers for the technology infrastructure, implementation details, challenges and concerns, benefits realized, and lessons learned) was used to guide the interviews. Wherever possible, the interview data was augmented by documents provided by the interviewees. Each of the individual interviews lasted an hour with the exception of the interview with Firm 5, which lasted 30 minutes.

Table 1. Summary of Firms Interviewed

Firm	Sector	Interviewee	Profile
1	Bank	Head of Strategy	Large Australasian private bank
2	Bank	Business development executive; Technical Architect	Large U.K. based bank
3	Bank	Business development executive	Large Europe based bank
4	Bank	CIO	India's second largest private bank
5	Bank	Enterprise Architect	Mid-sized Australasian public-sector bank
6	Bank	Enterprise Architect	Large Australasian private bank
7	Insurance	2 x Technology manager / Architect	Mid-sized Indian private general insurance firm
8	Insurance	CTO	Large Indian public sector general insurance firm
9	Insurance	CIO	Large Australasian insurance firm
10	Insurance	Enterprise Architect	Large Australasian public sector insurance firm
11	Product & Services	CTO; VP of Strategic Accounts	Small India-based ERP solutions firm
12	Product & Services	Technical architect	Large European ERP solutions firm
13	Product & Services	Technical architect	Large U.S. based software and services firm
14	Services	2 x Technical architect; Product manager	Large India-based software services and consulting firm
15	Services	Principal	Large US-based multi-national consulting firm

Communications with Firms 10 and 15 were limited to electronic communication. Firm 10 indicated that their firm did not have an explicit SOA strategy, but they were pursuing SOA practices at a technical level by "following reasonable SOA practices in terms of trying to keep things abstracted through the use of messaging middleware and a messaging portal". Firm 15 was able to supply documents describing its SOA strategy at the business and technical levels, and provide specifics of a case study of a large financial services firm. Both these firms are included in the analysis not as primary data but more as an emphasis to the findings from the data gathered in the interviews with the other firms.

Fourteen of the firms interviewed were in various stages of implementing SOA, either for themselves or their clients, most of them already having migrated targeted business functions to a service based deployment. The firms were able to provide some insight into the anticipated and observed benefits of the migration to a service-oriented approach. Firm 6 did not have an SOA strategy and had tried unsuccessfully to migrate to a service based infrastructure. The interview provided a valuable insight

into the challenges of building a business case for SOA adoption. The product and software service providers were able to provide an insight not only into the business drivers for their product offerings but also their perception of the business drivers for their clients.

The role of cultural influences was an obvious concern in interviewing these firms across various geographical domains. Many large banks generally have a mix of internal and external (either outsourced or increasingly off-shored) development teams. The approaches to technology strategy in these environments are a blend of the various geographically diverse teams that comprise the corporation. Europe-based bank Firm 3 has its development team operating out of South Asia, while the U.K. based Firm 2 has two development centers – one in the Far East and the other in South Asia. Representatives from these and some of the other firms in this analysis indicated that the larger corporate culture generally seemed to minimize the impact of geographically diverse operating environments. Cultural variation, hence, was treated as a controlled variable in this analysis since the current global environment of large corporations appears to mitigate the impact of cultural variations.

Transcripts of the individual interview data were analyzed using a two-pass method. The first pass of the analysis used thematic coding to identify broad categories of organizational issues. The second pass of analysis was performed using axial coding and major factors were identified using meta-codes. The meta-codes were then used to identify similar patterns across the data from the multiple firms interviewed. The following section details the results of the data analysis identifying the major business drivers for the choice of a service-based infrastructure solution..

4 Understanding the Business Drivers for SOA

From a strategic perspective, a business requirement that was common across most of the firms (Firms 2-8, 12, 14) interviewed was the delivery of a standardized set of products, both internal and those provided by strategic partners, over a unified *service delivery platform*. Firms envisage a seamless customer experience across all of their offerings and are consequently looking to implement a single service delivery platform for their customers.

At a tactical level, businesses recognize the need for process and infrastructure *agility* according to Firms 2-6, 8, 9, and 11-14. Firms are looking for the ability to integrate new products and third-party services into their product offerings for increased customer value – a flexible plug-and-play approach to facilitate the use of best-of-breed products transparently over their service delivery platform.

Finally, at an operational level, firms are pursuing opportunities for *efficiency gains*, looking to optimize their business processes and reduce costs, as indicated by Firms 1, 2, 3, 6, 9, 11, 13, and 14. These business drivers, it appears from the cross-firm data, are pushing the supporting IT teams to implement services based solutions that allow the seamless integration of internal and external resources, promise infrastructure agility to support business process agility, and provide process and development efficiencies [23].

These findings are summarized in Table 2 along with a few typical quotes from the interview transcripts.

Table 2. Summary of Business Drivers for SOA

Business Driver	Firms	Typical Quotes
Service Delivery Platform	2, 3, 4, 5, 6, 7, 8, 12, 14	[We are] looking for one solid plug-and-play platform…for seamless user interface and back-end integration. Drivers for consumer banking [are] a service platform to provide a single view of back-end systems in Phase 1. Phase 2 will provide a single customer view to agents. Business drivers for SOA [is] to be able to give consistency and standardization across systems.
Process and Infrastructure Flexibility	2, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14	Flexibility is a core requirement. This is achieved purely by the technical infrastructure. Flexibility is paramount, to be able to change the business process. We have services…If someone comes along and says I want to do this we can.
Efficiency Gains	1, 2, 3, 4, 9, 11, 13, 14	What matters to us is process optimization. [Our infrastructure] addresses operational and efficiency problems. We didn't realize this is SOA, but looking at it, it has evolved into SOA. Using SOA increases IT value by making it reusable and extensible.

5 Articulating the Business Case for SOA

Existing literature indicates that SOAs can potentially offer corporations increased business value. This literature is now examined to understand how SOA may satisfy the strategic, tactical, and operational business drivers identified from the analysis of the cross-firm data.

5.1 Strategic Value Proposition

The ability to seamlessly integrate external resources from strategic partners and internal resources, including legacy systems, enables a single *service delivery platform*, while the reuse of existing assets promotes standardization [14], [23]. Firms can take advantage of this seamless integration of internal and external resources and pull together disparate best-of-breed products and services to create a single service delivery platform, independent of the physical implementations of each individual component.

Moving to a service based approach also allows existing and proven legacy system functions on a diverse set of hardware and software platforms to be encapsulated as services on a new standards-based integration platform [7], [8], [10] and delivered to a broader customer base. The reuse of existing components, while enabling rapid product development and cost efficiencies, allows for a standardized implementation

of business functionality and consequently, a more consistent customer experience across product offerings [14].

The location transparency and technology neutrality of services allows businesses to include services from third-party providers or business partners into their own processes as value-added service offerings [23]. This opens up new avenues of strategic partnerships with suppliers, partners, and customers beyond traditional organizational boundaries [30], allowing a new business model – a re-bundling of intra- and inter-enterprise business processes as seamless services [13].

5.2 Tactical Value Proposition

Service-oriented systems enable *agility* in business processes by virtue of modularity and loose coupling, and allow for a flexible plug-and-play approach to business and infrastructure functions by abstracting the underlying service implementations [23].

Effective SOAs tend to be well-defined process-centric architectures facilitating better process visibility and process knowledge resulting in easier design, automation, monitoring, and most significantly, modification of business processes – i.e., resulting in improved process flexibility [7], [14], [27].

In addition, as the service paradigm permeates organizations, the services themselves can be virtualized from the underlying hardware platform. The underlying technology platform can be potentially substituted with ease, allowing for the best choice of platform for the services. This allows the business to focus on the core services while the infrastructure used to run the services become more of a commodity, to be leased or purchased from the provider of choice [7], [13], [33]. Organizations can focus on the efficient orchestration of services to form a product and shed the burden of owning resources [7].

From an implementation perspective, SOA's modular approach also means that companies need not plan to take on a high-risk all-or-nothing approach to its rollout [13]. They can adopt a phased migration to service-orientation, and leverage this approach to focus initially on opportunities that meet immediate customer requirements.

5.3 Operational Value Proposition

SOAs can potentially offer corporations the opportunity to realize process and development *efficiency gains* while mitigating the overall change and technology related risks of the corporation.

Existing architecture frameworks tend to be program-centric with business flow or process knowledge often spread across individual system components. This hampers the consolidation of information relevant to clearly understanding business flows. Well-designed service architectures allow for better process knowledge and facilitate the potential for continuous process improvement [7], [14], [27], [33].

Service orientation enables monitoring of services from a business perspective rather than systems perspective, allowing for better process visibility [7], [14], [27], [33]. Business services that are the core competence of the organization can be clearly identified, and the non-core services can then be candidates for substitution by those provided by vendors with the relevant expertise [13].

Table 3. Summary of how SOA meets the identified business drivers

Business Driver	Business Case for SOA
Service Delivery Platform	1. Seamless integration, facilitates the leveraging of internal legacy systems and enabling strategic partnerships 2. Reuse of services enables consistent customer experience
Process and Infrastructure Flexibility	1. Process centric architecture enables process flexibility. 2. Virtualization facilitates portability across infrastructures. 3. Modular architecture enables phased rollout across the enterprise.
Efficiency Gains	1. Increased business process visibility allows for process improvement. 2. Process visibility also enables the identification of services that the firm has a comparative advantage in, allowing: a. Firms to outsource non-core services and b. Firms to offer core services as new products 3. Reuse could result in reduced development and testing costs. 4. Reuse also allows for rapid product development.

In addition to improving the existing business models, service based architectures provide a framework for corporations to offer their core competencies as services to other companies [13], [14], [23], focusing on areas of comparative advantage while buying or leasing services in which they lack superior expertise from other service providers [8]. SOA, thus, potentially allows for the creation of new products generating additional revenue streams.

The reuse of existing components, while providing potential savings in operational costs could also reduce risk in more ways than one [7], [10]. The enhanced business process incurs no new potential points of failure, and the maintenance of the supporting infrastructure continues to remain unaltered. In addition to the reduction in development and testing costs brought about by modularity and re-usability of service modules [15], the learning curve of the development or assembly team could potentially reduce over time due to familiarity with existing services [7], [14], [15], [23], [27], [38]. Gains may also be realized in terms of development and maintenance cost savings by purchasing services from reliable providers with a comparative advantage in developing the services [13].

Over time, the developed services become a core asset of the organization – a library of tested, ready to use, compatible components [7]. This promotes rapid product development, reducing the time to pull together well-design tested functionality to meet new and changing market needs [14], [23].

6 Related Work

Of the studies that evaluate the strategic value proposition of SOA, three studies stand out as stage-setters for future research and are discussed in the following paragraphs.

The first two examine the potential value of Web Services specifically, while the third looks at SOA in general.

The first of these studies, an analytical study by Huang and Hu [14], investigates the link between Web services and competitive strategy using a popular strategic management tool, the Balanced Score Card [18]. The authors use the scorecard's four dimensions to establish propositions about how Web services could support or improve the following perspectives - learning and innovation, internal business process, customer, and financial. The cases to support the validity of these propositions are drawn from existing literature, industry reports, and vendor analyses from IBM and Microsoft.

The second study [16] argues the advantages of Web Services, a popular technology implementation of SOA, as an enabler of dynamic business networks using a popular stakeholder model for IS architecture [37]. The authors extend Zachman's architecture model to build a stakeholder model for Web Services along the dimensions of the owner, architect, builder, and end-user, and argue the benefits of Web Services along these dimensions to be ease of sourcing the IS implementation, modularity, IS integration, and ease of access respectively.

The last of these studies links SOA to the concept of dynamic capabilities, a concept in strategic management that research scholars indicate may help firms gain competitive advantage in rapidly changing market environments [23]. The concept of dynamic capabilities is a widely accepted approach to understanding the competitiveness of organizations. The authors explain the attributes of SOA that may make it amenable to creating these dynamic capabilities – i.e., integration of internal resources, integration of external resources, rapid product development, learning, and the creation of technological assets.

These frameworks offer possible means to study the business drivers and the real-world benefits of SOA implementations across the business and technology domains, but the real-world business drivers for SOA remain largely unexplored. The examination of the real-world drivers in the financial services domain is crucial given the rapid adoption of SOA by financial service firms.

7 Contribution and Future Research

Research in the area of information technology diffusion indicates that the successful adoption of new technology requires organizations to take an integrated approach to organizational and technical changes caused by the adoption of the new technology [24]. There is a growing understanding of the organizational processes and characteristics that influence the adoption and implementation of technology [4], [19], but there is little understanding of the business drivers that influence the organizational adoption of SOA [22].

This study adds to current knowledge by empirically investigating the business drivers for SOA adoption using data gathered from semi-structured interviews across multiple firms with a fairly broad representation within the financial services industry – banks and insurance firms, which researchers have identified as having high dependence on technology [17], [25], and service providers with clients in the financial services industry. The interview data was thematically coded to glean what challenges

the firms faced in the process of implementing SOA, and our findings were fairly consistent across the firms interviewed. Having identified the business drivers for the technology infrastructure in financial service institutions, this study then examines the value proposition of SOA vis-à-vis these business drivers and builds a business case for the use of SOA by financial service institutions.

In investigating the business drivers influencing the adoption of service-orientation, this study

(i) fills a crucial knowledge gap because there is little empirical evidence of the practical enterprise business drivers for the adoption of SOA,

(ii) builds the business case for the adoption of SOA by examining the value propositions of SOA that fulfill the identified business drivers, and

(iii) lays the groundwork for future research for understanding the actual business value that may be realized from the adoption of SOA.

The findings of this paper are part of a larger research effort to leverage the data from the fifteen firms interviewed to understand how the enterprise SOA strategy can be aligned with the organizational strategy. Augmenting the findings of this study, the real-world challenges faced by organizations adopting a service-based approach may be found in a related study by the authors of this paper [21]. The next phase in this research effort involves a continued analysis of the data to develop and validate a framework for SOA implementations.

Acknowledgements

The generous scholarship provided by the DEST-funded project ADAGE is gratefully acknowledged.

References

[1] Beidleman, C., Ray, M.: The agility revolution. In: Cortada, J.W., Woods, J.A. (eds.) The Quality Yearbook 1998 (1998)

[2] Bell, A.: DOA with SOA. ACM Queue 5(1), 56–58 (2007)

[3] Benbasat, I., Goldstein, D.K., Mead, M.: The Case Research Strategy in Studies of Information Systems. MIS Quarterly 11(3), 369–386 (1987)

[4] Broadbent, M., Weill, P.: Improving business and information strategy alignment: learning from the banking industry. IBM Systems Journal 32(1), 162–179 (1993)

[5] Brodie, M.L.: Illuminating the Dark Side of Web Services. In: 29th Very Large Data Base (VLDB) Conference, Berlin, Germany (2003)

[6] Campbell, D.: Service Oriented Database Architecture: App Server-Lite? In: SIGMOD 2005. ACM, Baltimore (2005)

[7] Channabasavaiah, K., Holley, K., Tuggle, E.M.J.: Migrating to a service-oriented architecture. On demand operating environment solutions (2004)

[8] Curbera, F., et al.: The Next Step in Web Services. Communications of the ACM 46(10), 29–34 (2003)

[9] Best Practices Council - SOA Survey Results 2008. Data_Strategy_Journal (2008)

[10] Datz, T.: What You Need to Know About Service-Oriented Architecture. In: CIO 2004 (2004)
[11] Erl, T.: Service-Oriented Architecture - Concepts, Technology, and Design, p. 792. Prentice Hall - Pearson Eduction, Inc. (2005)
[12] Fricko, A.: SOAs Require Culture Change and Service Reuse. Business Communications Review, 58–64 (2006)
[13] Hagel, J.I., Brown, J.S.: Your Next IT Strategy. Harvard Business Review, 105–113 (2001)
[14] Huang, D.C., Hu, Q.: Integrating Web Services With Competitive Strategies: The Balanced Scorecard Approach. Communications of the Association of Information Systems 13, 57–80 (2004)
[15] Huhns, M.N., Singh, M.P.: Service-Oriented Computing: Key Concepts and Principles. IEEE Internet Computing 9(1), 75–81 (2005)
[16] Iyer, B., et al.: Web Service: Enabling Dynamic Business Networks. Communications of the Association of Information Systems 11, 525–554 (2003)
[17] Jarvenpaa, S.L., Ives, B.: Information technology and corporate strategy: a view from the top. Information Systems Research 1(4), 351–376 (1990)
[18] Kaplan, R.S., Norton, D.P.: The Balanced Scorecard: Measures That Drive Performance. Harvard Business Review 70(1), 71–79 (1992)
[19] Lai, V.S., Guynes, J.L.: An assessment of the influence of organizational characteristics on information technology adoption decision: a discriminative approach. IEEE Transactions on Engineering Management 44(2), 146–157 (1997)
[20] Lim, B., Wen, H.J.: Web Services: An Analysis of the Technology, Its Benefits, and Implementation Difficulties. Information Systems Management 20(2), 49–57 (2003)
[21] Luthria, H., Rabhi, F.A.: Organizational Constraints to Realizing Business Value from Service Oriented Architectures: An Empirical Study of Financial Service Institutions. In: 6th International Conference on Service Oriented Computing (ICSOC 2008). University of New South Wales, Sydney (2008)
[22] Luthria, H., Rabhi, F.A.: Service Oriented Computing in Practice - An Agenda for Research into the Factors Influencing the Organizational Adoption of Service Oriented Architectures. Journal of Theoretical and Applied Electronic Commerce Research (to appear, 2008)
[23] Luthria, H., Rabhi, A., Briers, M.: Investigating the Potential of Service Oriented Architectures to Realize Dynamic Capabilities. In: IEEE Asia-Pacific Conference on Services Computing, 2007 (APSCC 2007). IEEE Computer Society/ Conference Publishing Services, Tsukuba Science City/ Japan (2007)
[24] Margaria, T., Steffen, B.: Service Engineering: Linking Business and IT. In: Computer (IEEE), p. 45 (2006)
[25] McFarlan, F.W.: Information technology changes the way you compete. Harvard Business Review 62(3), 98–103 (1984)
[26] Mukhi, N.K., Konuru, R., Curbera, F.: Cooperative Middleware Specialization for Service Oriented Architectures. In: International World Wide Web Conference. ACM, New York (2004)
[27] Natis, Y.V.: Service-Oriented Architecture Scenario. In: Gartner 2003, Gartner, Inc. (2003)
[28] Papazoglou, M.P.: Service-Oriented Computing: Concepts, Characteristics and Directions. In: Fourth International Conference on Web information Systems Engineering (WISE) (2003)

[29] Papazoglou, M.P., Georgakopoulos, G.: Service-Oriented Computing. Communications of the ACM 46(10), 25–28 (2003)
[30] Papazoglou, M.P., Van Den Heuvel, W.J.: Service-oriented design and development methodology. International Journal of Web Engineering and Technology 2(4), 412–442 (2006)
[31] Progress_Actional, SOA Survey Overview: The Current State of SOA Governance 2006 (2006)
[32] Quocirca, SOA: Substance or Hype? 2005, Quocirca Ltd. (2005)
[33] Sprott, D.: The Business Case for Service Oriented Architecture. CBDI Journal (2004)
[34] Sprott, D.: Service Oriented Architecture: An Introduction for Managers. CBDI Journal (2004)
[35] Turner, M., Budgen, D., Brereton, P.: Turning Software into a Service. Computer, 38–44 (2003)
[36] W3C, W3C Working Group Note 11 February 2004, W3C (2004)
[37] Zachman, J.A.: A Framework for Information Systems Architecture. IBM Systems Journal 26(3), 276–292 (1987)
[38] Zhang, J., Chung, J.-Y., Chang, C.K.: Migration to Web Service Oriented Architecture. In: SAC 2004. ACM, Nicosia (2004)

Model Based IT-Governance Compliance Analysis

Jan-Helge Deutscher and Carsten Felden

TU-Bergakademie Freiberg, Lessingstraße 45,
09599 Freiberg, Germany
Tel.: +49 (0) 37 31 - 39 - 26 11
Fax: +49 (0) 37 31 - 39 – 36 90
{jan-helge.deutscher,carsten.felden}@bwl.tu-freiberg.de

Abstract. This paper presents a metamodel that allows constructions of specific model instances to support the measurement of successfully tracking IT-governance related objectives. This aim is accomplished by a model-based support of the operationalization and the analysis of defined objectives in the manner of determining compliance degrees for companies. The resulting metamodel is achieved by applying an established construction method.

Keywords: Model, Governance, Compliance, Measurement, Analysis, Operationalization, Metamodel, Construction, Objectives, Tracking.

1 Introduction

The financial industry, including banking, insurance, brokerage, and other services, is challenged by a competitive environment and an increasing number of regulations. Since frequent mergers and acquisitions cause an ongoing consolidation, the adaption to changing markets is crucial to economic success. Moreover, this sector is constrained by ensuring compliance to regulations and auditability of their systems. A significant contribution to meet these challenges can be made by IT [9; 25; 7]. However, this requires IT to provide capabilities of use to the organization [5; 14; 23]. This can be achieved by aligning the IT-strategy to the organization's strategy [20; 12, 27], which means to adjust priorities, goals, and objectives of the IT-strategy to the priorities, goals, and objectives of the firm's business strategy [10; 11; 28]. Proper alignment involves both strategic and operative aspects as proposed by Henderson's and Venkatraman's alignment model [17]. Hence, an IT-strategy requires proper implementation to be able to provide capabilities of use to the organization. The tasks regarding IT-strategy implementation are within the scope of IT-governance. It comprises the following categories that are derived from several definitions [18, 167; 33, 2, 8; 13]: first, the definition of objectives inferred from the IT-strategy. Second, managing the process of operationalizing abstract objectives. Third, tracking the attainment of objectives by analyzing implementations' compliance to their IT-strategy [19, 22].

However, compliance cannot be measured directly, since it is based on strategic objectives that are insufficiently operationalized. Hence, there is neither a possibility to analyze the degree of compliance accurately nor courses of action that are clearly

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 108–125, 2009.

identified in attaining higher levels of compliance. Thus, the objectives lack operationalization. Moreover, the progress in attaining objectives as assessable expenditures needs to be analyzable in the meaning of measuring progress, since investments in IT are to be justified.

For supporting the operationalization of IT-governance objectives in this paper an approach is taken based on a basic algorithmic design principle: divide and conquer. This means to break down a problem (objective) into sub-problems (several objectives). These sub-problems are easier to solve in terms of finding measures and approaches to implementation. The solution of the entire problem is the combination of all sub-problems solutions – and this principle can be applied recursively to sub-problems.

For this papers field of interest the idea of divide and conquer is adopted in the anticipation of being able to split intangible objects into analyzable parts. This is done as follows:

The goal of compliance can be decomposed into multiple partial goals. This leads to the assumption that the achievement of all partial goals means that the total goal is achieved and thus compliance is attained. Therefore, the pursued approach is the splitting of goals into operationalized partial goals, whose successful conversion can be seized quantitatively. With the pursuit of this approach, a fundamental component is the quantitative analysis of the current degree of compliance and comparison with the desired situation.

This paper introduces a metamodel to support the following aspects:

- The metamodel permits the acquisition of all targets which form the total goal of IT-governance compliance. Goals, which are not directly operationalizable, are split until seize-ability.
- This splitting process is arbitrary and ends, if a sufficient operationalization is achieved. A hierarchical structure is developed, whereby the partial goals are measurable in its degree of completion on the lowest level.
- Furthermore, the partial goals can be weighted, in order to be able to model their impact on the total goal, in relation to the other goals.

The quantitative orientation of the metamodel permits the application of the economic principles and thus a clear comparison of compliance developments and the associated expenditures by measuring the progress.

The remainder of this paper is organized as follows: The construction of the metamodel is conducted according to the design method presented in section 2. Afterwards the results of the construction process are documented in section 3, comprising the steps problem formulation, requirement specification, model concept, concept validation, implementation, verification, and validation. The validation is based on a prototype application that is developed in cooperation with an industry partner. Finally, in section 4 the results are summarized and further research directions are pointed out.

2 Structural Design Method

Literature studies and discussions with several enterprises during research projects lead to the following procedure:[1]

[1] [2, 150; 3, 621; 4, 16; 8, 52; 16, 607; 21, 11, 38, 42, 46, 226].

1. Formulation of the problem definition in the initial situation which serves as basis for the derivative.
2. This is enhanced by the definition of the requirement specification which defines the goal that should be achieved, and a manifestation of the desired solution. Further requirements can be assigned to the objective, for example restrictions, which exclude certain solution types.
3. The next step is the Construction of the model concept which embodies the actual idea of the problem solution in form of a solution hypothesis.
4. A concept validation takes place, in order to examine the correctness of the artifacts provided in the preceding steps.
5. Subsequently, the model concept is converted (implementation) which usually means a specification by formalization up to simulation-able models.
6. The subject of the verification is the examination of the formal model regarding the correct conversion of the model concept. Therefore it is examined whether the model corresponds to the agreements (specification) and the transformation of the model concept into a model according to the rules. Due to this, the term of specification refers not only to the requirement specification in step 2, but covers all the model affecting artifacts (agreements, assumptions, restrictions, etc.) which can be provided.
7. It is examined in the final validation step whether the provided model is suited for the solution of the assigned tasks. A substantial fitness criterion is the utility of the result accuracy due to the application of a model within its intended field of operation. Additionally, credibility is a further criterion, in particular with complex models.

Addressing the issue of possible alternative modeling procedures it can be stated, no available procedure model could be identified as an improved and useful solution. The search for available alternate modeling procedure is limited and not pursuable in extenso, but a quite relevant criterion is to choose a modeling procedure that is accepted by all project stakeholders, because common sense is important for such a measurement approach.

3 Metamodel Construction

This section specifies the discussion and the results of the model construction. The procedure is according to the structural design method previously chosen.

3.1 Problem Formulation

The description of the initial situation already took place in the introduction: A principal purpose of IT-Governance is the development of targets, whose successful conversion leads to an optimal implementation of IT-strategies within organizations.

As stated before, IT-Governance cannot be attained directly, thus operationalization of its goals is required to be able to analyze the degree of compliance on a quantitative basis. The operationalization leads to multiple possible partial goals and transparency regarding each goals' purpose is to be ensured. Furthermore, the partial goals are set in relation to each other, regarding to their impact on the total goal of

compliance. This is done by emphasizing the partial goals. Decomposing the goal of total compliance leads to problems with the partial goal weighting and the maintenance of the goals.

Additionally, the determination of successfully attaining the total goal is based on meaningful analysis results. This means to express the successful conversion for each partial goal within hierarchies of partial goals (whereby each level has to be summarized).

3.2 Requirement Specification

The preceding problem formulation emphasizes which expectations are addressed to the model. The requirements can be summarized as follows:

- The model provides the permitted set of all partial goals. It is negligible, if the goal is operationalized. The only condition is that the goal can be operationalized directly or the goal can be divided into partial goals, whereby the division must be finite and ends in an operationalization. This leads to the result that the level of attainment is measurable.
- Furthermore, there has to be a feasible weighting of all partial goals.
- For each partial goal, an analyzed achievement of the objective degree should be available. Therefore, partial results of underlying levels have to be aggregated.
- The analytical work within the model has to have a quantitative nature, so that for example the economic principle can be applied.
- The model should be formalizable in such a manner that a simulation is possible.

3.3 Model Concept

The philosophy of the solution is already addressed, i.e. separating a complex, not understandable total requirement to gain improved analyzable partial requirements. Since the total requirement cannot be formulated precisely, it is tried to define the partial requirements as precisely as possible. Assuming idealized conditions, the total problem is solved, if all sub-problems are solved (in accordance with the principle of divide and conquer) [30, 78, 90].

Each group is going to seize partial goals as so-called aspects. An aspect represents a group of partial goals which realizes a super ordinate goal in their fulfillment. Any number of aspects forms a multidimensional system of objectives of requirements which can be fulfilled. Within the entire model system weightings are possible, in order to be able to consider the balances of weights of the individual analysis results.

The weighting on the aspect level is a control instrument. Conflicts between aims can be solved by weighting the means of requirements prioritization [32, 55]. This approach is comparable to the approach of defining the maximum utility by accurate achievement of objectives with minimum expenditures. Therefore, the usage of the IT-Governance is increased, if all nominal values are fulfilled with minimum expenditures [1, 104; 6, 332; 26, 29; 35, 1].

The fulfillment of a goal within the model is determined by the comparison of a target, characterized by a nominal value, with a current value. The current value is the result of measuring things in the real world. The target level reflects necessary developments of one or more criteria, in order to ensure compliance. The current level is the actually measured compliance level in the respective situation.

Furthermore, the approach enables the examination of each measure (investment) regarding its influence on the compliance. Due to quantifications in the model, the economic principle is reflected by the application. If several investments are competing for a defined budget, the most economical can be selected. The model is conceived in such a manner that it can be formalized. This has the advantage that e.g. the effects of taken measures can be simulated and examined a priori. In summary the idea of the model concept is as follows:

1. The system of objectives is operationalized, while aspects and contained sub goals are formed. Methods of analysis (procedure measurements) are specified on the lowest hierarchy level for the determination of the degrees of completion of the individual partial goals.
2. The execution of analyses consists of the following steps for each measurement procedure:

- The definition of a nominal value which represents the goal which is to be achieved.
- The definition of a minimum or initial value that is the lowest possible respective worst value. The difference between target and initial value is set in relationship to the difference between current and initial value. This relationship is expressed in percent and reflects the degree of attainment completion.
- The computation of the current actual value, which is computed in the respective moment on basis of the available input data.

3. The summary of the analysis results consists of meaningful numbers representing the degree of compliance on each hierarchy level.
4. The user-fair presentation of the summarized analysis results to serve as basis for decision making

Finally, the role of the model during the conversion process in favor of an optimal compliance has to be clarified. Figure 1 shows that the given goal optimal compliance is represented by a block.

An eventual situation is the current compliance. The achievement of objective degrees is determined in the context of the analysis (Degrees of compliance). These results are brought to decision makers in the context of the presentation of analysis

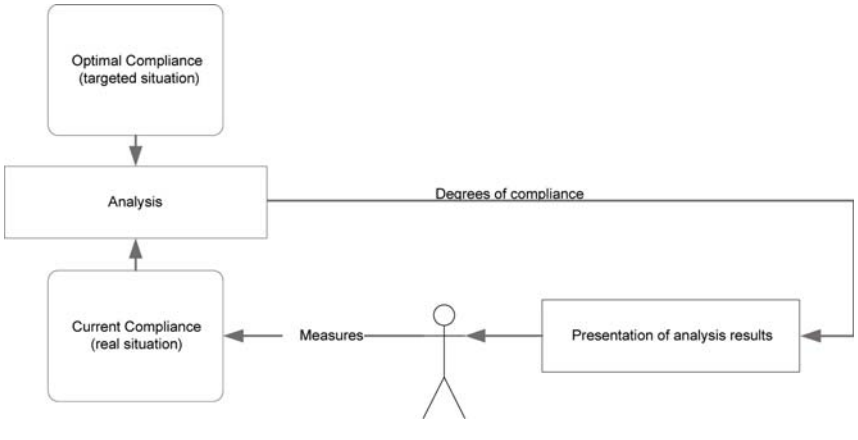


Fig. 1. Process of tracking objectives for optimal compliance

results as decision basis. The decision maker (represented by the stickman in illustration) seizes measures in order to reduce the deviations. These steps are repeatedly conducted to achieve the targeted situation.

3.4 Concept Validation

The concept validation is in form of a structured walk through [22, 83]. In addition, the results of all preceding modeling steps were discussed in a meeting with the key project personnel.

There are five participants in the meeting held for the structured walk through. Two of the university, i.e. the modelers, and three of the corporation, one innovation manager, one project responsible employee, and one power user. The walk through is based on a document about which parts comply with the type of assumption document described by Law and Kelton [22, 276]. In addition, the document contains all artifacts of the modeling process at a certain point in time. In detail, the validation document comprises the following aspects:

- Methods used for validation
- Justifications and implications of the methods' chosen
- Artifacts considered as input
- Expected result artifacts
- Statements regarding the reasonableness of input- and output-data
- Exercises to be conducted for obtaining expected results
- Documentation of the exercises
- Evaluation of the results
- Initial impressions of the concept's ability to solve the problem
- Conclusions, whether the concept is valid or further validation exercise is required

The result of the walk through is that the involved stakeholders appreciated the solution.

3.5 Implementation

This section presents results of the modeling process. Additionally, the interacting metamodel components are presented. This leads to the result of a metamodel with components and construction rules. This is the basis that allows the design of explicit instances. The representation of the metamodel takes place according to the rules of the Meta Object Facility (MOF) of the Object Management Group [27]. The model is based explicitly on EMOF which is a sufficient part of the entire MOF [27, 31]. The necessary data types defined for the metamodel are presented in Figure 2.

- Base type of all types used in the model is *Skalar*. The type *double* is chosen for implementation. The type *NavResult* contains three values of the type *Skalar*. The invariant *Constraint1* assures that one *NavResult* contains one nominal, one current and one initial value (*Xnominal, Xcurrent, Xinitial*).
- The type *PW* is derived from *Skalar* with an additional constraint (*Constraint2*): All values of *PW* must be equal or greater zero.
- The type *KPW* is derived from *PW* with an additional constraint (*Constraint3)*: All values of *KPW* must be equal or greater zero and lower or equal to 100. This type represents the percent values of degrees of compliance.

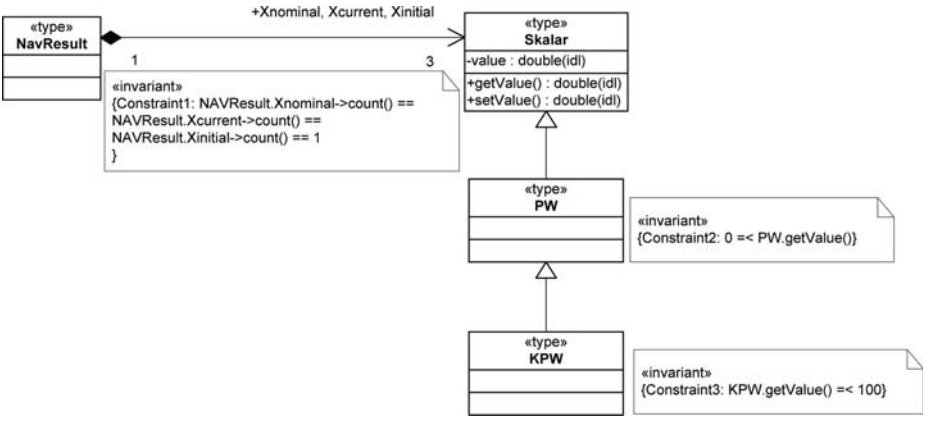


Fig. 2. Data types defined in the metamodel

The metaclasses of the metamodel are as follows (see Figure 3):

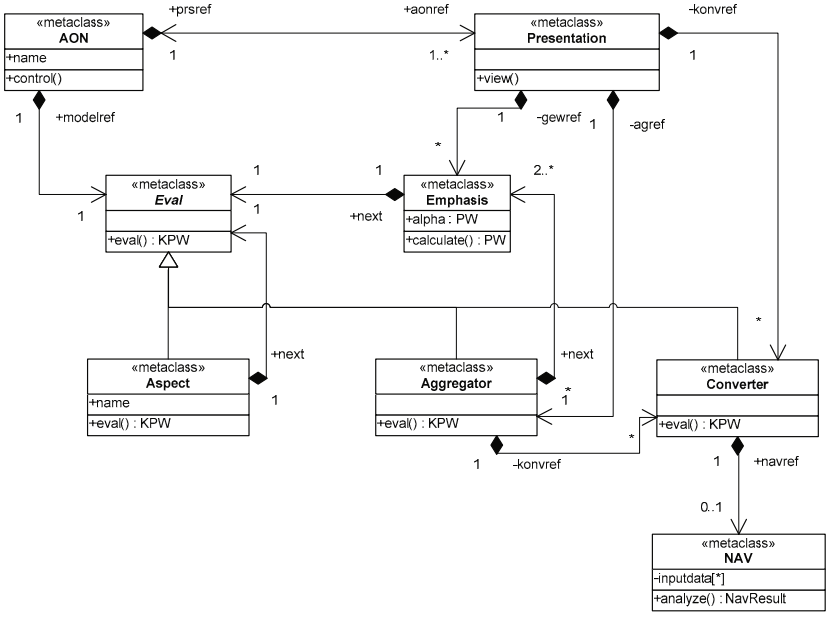


Fig. 3. Classes of the metamodel

- Each model (*AON*) holds one or many presentations of analysis results (*metaclass presentation)* and has a reference to the analysis model (*modelref*). Each model hat a unique name for identification purposes (*AON.name*). The metaclass *AON* contains a method *control* which stands for the stereotype role of the metaclass in a model view controller architecture (MVC) [15, 4], i.e. the role of the controller. The reference *modelref* is used to access the analysis model part.

- The metaclass *Presentation* is for appropriate user-oriented presentations of analysis results. The analysis results are generated in the analysis model part and presented by classes of *Presentation.* Therefore, the metaclass *Presentation* has access to the entire model structure. For realizing the task of user-fair presentations, metaclasses of the type *Presentation* may utilize the metaclasses *Emphasis*, *Aggregator* and *Converter.* The stereotype role of *Presentation* in the MVC- architecture is the role of the *view* component.
- The metaclass *Eval* represents the model in the MVC-architecture, i.e. the analysis part of the model. Furthermore, the metaclass *Eval* defines a command (*Eval.eval*) and is in conjunction with the metaclasses *Aspect*, *Aggregator* and *Converter* a command-pattern [15, 233]. This pattern allows pointing out a method call and initiates the calculation of all analysis results. The return value of the method *eval* is the type *KPW.* Thus, by calling the method *eval* the compliance degrees within all levels of the model can be calculated.
- In conjunction with the metaclasses *Aspect* (role: composite), *Emphasis* (role: composite in conjunction with *Aggregator*) and *Converter* (role: Leaf) the metaclass *Eval* is a composite pattern [15, 163]. This enables the realization of the recursive model idea. Due to this reason, a total requirement can be divided into many desired partial requirements, grouped into aspects and weighted results of analysis. An aspect can consist also of aspects, if it is necessary for a specific problem.
- The metaclass *Aspect* represents an aspect which is identified by a name (*Aspect.name*). The method *eval* returns a degree of compliance which represents an aggregated degree of compliance of all subordinated goals of the aspect. The subordinated components can be accessed via the reference (*Aspect.next*).
- The metaclass *Aggregator* aggregates two or more weighted results (metaclass *Emphasis*, access with *next)* and returns one aggregated result value with the method *eval.* Metaclasses of this type may use its own instances of *Converter.*
- The metaclass *Emphasis weights* a value received with the reference *Emphasis.next.eval* by multiplying it with an emphasis factor (*alpha*). The result is returned with the method *calculate.* The result is of the type *PW.*
- The metaclass *Converte*r converts a value of the type *NavResult* to a value of the type *KPW.* The method to be called for conversion is *eval.* Metaclasses of this type may be owner of analysis procedures (metaclass *NAV).*
- The metaclass *NAV* represents all analysis procedures which serve for the transfer of measured values from the real world into the model. The input data is arbitrary. A result of the type *NavResult* is returned by calling the method *analyze*.

In order to allow tests of the metamodel its classes have been implemented in the tool *Matlab Simulink* as follows (see Figure 4):

- The metaclass NAV is implemented with the block *measurement procedure*. The block contains the analysis operation (method eval) and places the results as tripels of the type NavResult with the outputs: *nominal value*, *current value*, and *initial value*.
- The metaclass converter is implemented with the block *converter*. The input of the block is a triple of the type NavResult and the output is a value of the type KPW.

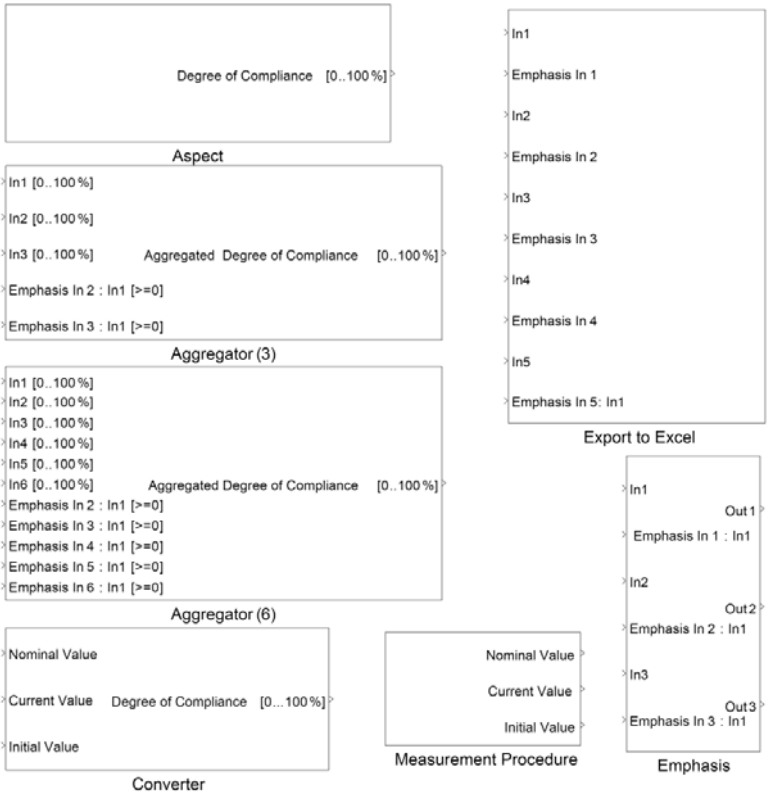


Fig. 4. Library components that implement the metamodel in *Matlab Simulink*

- The metaclass aggregator is implemented with the block *aggregator*. The number in braces indicates the number of the aggregator's inputs. It can be any positive number greater one. A weighting is already integrated, so that the metaclass emphasis is contained. The weighting factors indicate emphasis in relation to the values of the entrance *1* of the type KPW. The block performs an aggregation and outputs a degree of compliance of the type KPW. For this purpose the block uses the block *converter* internally.
- The metaclass emphasis is implemented with the block *emphasis*. Each entrance (*In1*, *In2*, *In3*) calculates a weighting by multiplication with a weighting factor (emphasis). This block is not directly necessary for the modeling, since it is already integrated in the *aggregator* block.
- The metaclass aspect is implemented with the block *aspect*. The block outputs a degree of compliance with values of the type KPW.
- The metaclass eval is abstract and not directly implemented. The blocks *aspect*, *aggregator*, and *converter* are implementations of eval.
- The metaclass presentation is implemented by already existing *Matlab Simulink* blocks. In the models for example, the display component is part of the standard

library. The block *export to MS Excel* belongs also to the class of presentations and has been specifically developed.

- The metaclass AON aggregates all components of a model instance and is the entire simulation file of a project.
- An explicit conversion of the data types is not necessary, because these intervals are defined as natural numbers. In order to the adherence of the ranges of values, which are defined in the metamodel, constraints and tests are integrated. These tests are performed at run time. Thus, the simulation interrupts, if an inadmissible value excess is recognized. In such a case a message is displayed with references to error localization.

3.6 Verification

According to Sommerville, the verification can be performed by testing the solution [31, 97]. This was conducted by developing interactive tests in M*atlab Simulink* for all components. Even though these tests run automatically, they allow investigating each components behavior. This feature was used to introduce our cooperation partner in the verification process.

The general set up of the component tests is depicted in Figure 5.

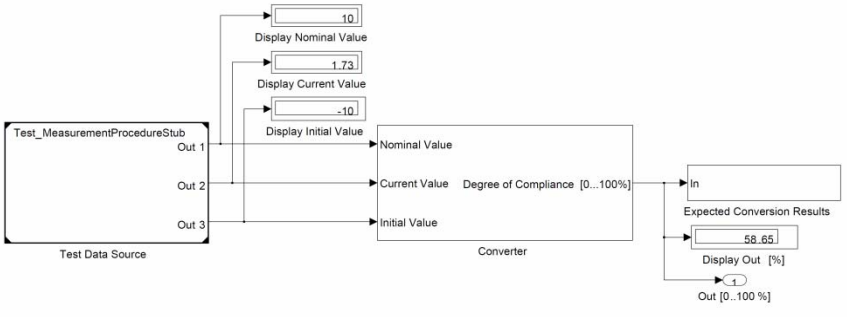


Fig. 5. General setup of the metamodels' component tests implemented in *Matlab Simulink*

On the left hand side the component providing test data is depicted. In the center the component under test is located. At the right a component checking for expected conversion results is located. In order to allow user interaction in terms of understanding the components' behavior several gauge-components display input and output values. The output of each tested component is propagated. This allows designing cascaded tests, which means the input is taken from another test and this test's output can be used as source of the input for the aggregator component test.

The adherence of all components to the specification during run-time is ensured by included constraint-controls. They are based on assertions. If one fails, the simulation is interrupted and further information is provided for error localization.

The verification results prove that all components meet their respective specification.

3.7 Validation On the Basis of an Example

The model is validated with the help of an application, which is developed in cooperation with an external capital provider of the utility industry.[2] The venture is motivated by the need of utilities to comply with new legislative regulations (unbundling), which are the result of Germany's approach to energy market liberalization. Moreover, as a consequence of market liberalization, utilities are forced to cope with the advent of competition. Even though the application is developed in cooperation with a utility firm, the results regarding the metamodel's validity should equally apply for any organization that relies on proper IT-strategy implementation with the help of IT-governance.

Our capital provider decided to take this challenge by developing and implementing new strategies. In this change process IT is considered as key provider of capabilities that are required for strategy implementation. Therefore, the IT-Strategy is to be aligned to the organization's strategy and properly implemented. This process of ensuring proper IT-strategy implementation is a major task of IT-governance, as defined in the introduction.

The approach taken by our capital funder is supported by the model presented in this paper. It was used to develop a specific instance that covers the needs of our project partner as follows: Initial points for the structure of the model are the tasks of the IT-governance identified in co-operation with the partner, which are modeled as aspects in the model. These aspects are IT related objectives, whose fulfillment is analyzed with the help of the model. For that purpose the model's top level is designed according to the following aspects:

1. The aspect *IT value management* covers requirements from the range of the IT investment governance [18, 167]. This includes questions of cost accounting and transparent billing for services consumed by contractors.
2. The aspect of *IT risk management* comprises requirement regarding security related issues. According to [18, 167]: „IT related risks must be mitigated". Therefore, appropriate actions need to be taken.
3. The aspect *IT organization management* covers requirements regarding the dispatching of tasks, authority, and responsibility. This request starts with the definition of the tasks of IT-governance [33, 2, 8]. The requirements formulated in this aspect point further to a high congruence to the tasks from corporate governance according to Schmidt and Weiss [29, 108]. Due to this reason, it is to be derived that also requirements can be specified that regard the observance of behavior recommendations. For example, the German corporate governance codex consists of approx. 40 percent behavior recommendations [24, 741].
4. The aspect *IT integration management* bundles requirements ensuring that the goals of the IT-governance fit the goals of the enterprise. This is necessary, because the aim of IT-governance is to guide IT-strategy implementation. Thus, they way it is implemented has to be compliant to the corporate governance rules [34, 11].
5. The aspect *IT implementation management* groups requirements referring to the way of the conversion of requirements. For example, our capital provider defines that a regular project is to be realized according to defaults of *Prince2* and must not endure more than twenty months. Consequently, this dimension holds requirements that apply within the other dimensions when requirements are implemented.

[2] The model is developed in cooperation with Stadtwerke Düsseldorf (SWD), contact person: Dr. Ivonne Servaes.

The model is realized in *Matlab Simulink* to allow validation. Details regarding the implementation are subsequently provided (see Figure 6).

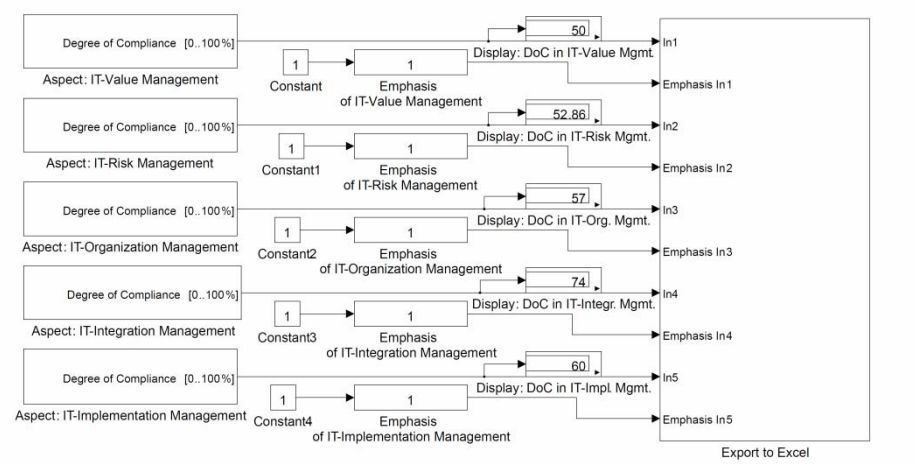


Fig. 6. Top level of the IT-governance model implemented in *Matlab Simulink*

Figure 6 shows the top level of the model: the recognized aspects are shown on the left hand side. The weightings and announcements of the intermediate results are placed in the middle display space. Since the model is implemented in a simulation software environment (*Matlab Simulink*) the analysis can be performed in real-time. This allows analyzing the influence of partial goals to the overall goal. For instance, the influence of changing weights can be explored by simulation.

The results at the model's top level are exported to MS Excel as shown in Figure 7:

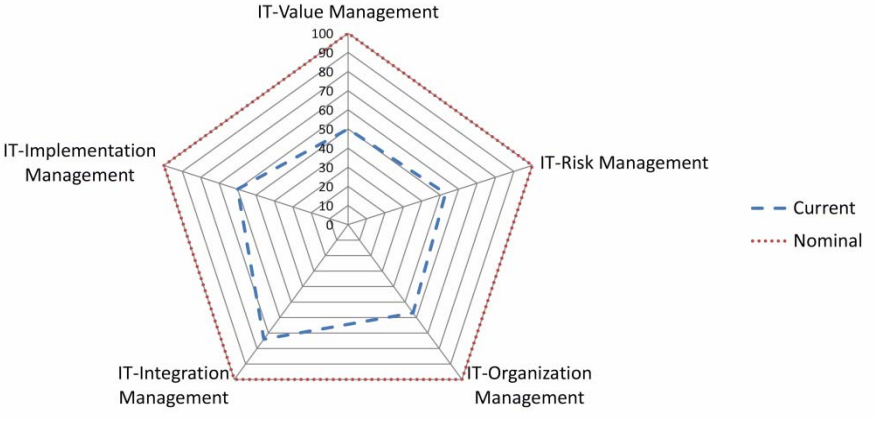


Fig. 7. Analysis results of the IT-governance model, automatically exported to Microsoft Excel. The outer dotted line represents the nominal values. The inner dashed line represents the current analysis results.

A radar diagram arranges the aspects. The outside line presents the targets and the inner line gives the current levels of compliance. The model structure repeats itself on all levels, until sufficient operationalization of the partial goals / requirements is reached. Due to limited space, not all levels of the provided model can be shown here. However, to demonstrate the process of operationalization the aspect of *IT-Risk Management* is presented in detail:

As depicted in Figure 6 the aspect *IT-Risk Management* is a top level goal. It is decomposed to several sub aspects (see Figure 8):

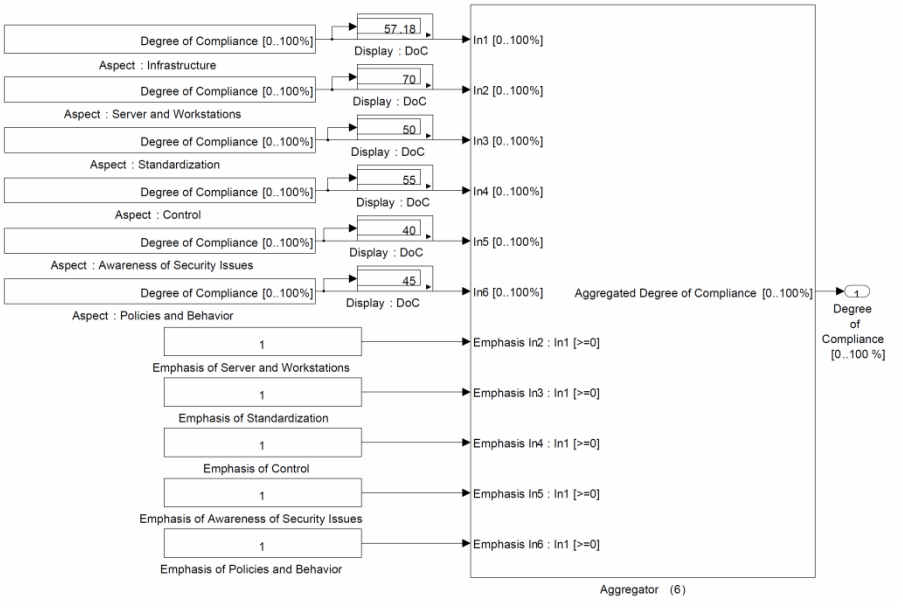


Fig. 8. Aspects of IT-Risk Management

- Aspect *Infrastructure*: issues regarding the functionality of the IT infrastructure
- Aspect *Server and Workstations*: issues regarding the security of end user computers and servers
- Aspect *Standardization*: use of standardized hard and software to allow the reduction of complexity
- Aspect *Control*: organizational aspects like responsibilities and hierarchies
- Aspect *Awareness of Security Issues*: degree to which members of an organization recognize / understand why certain issues are a threat to the organization's IT
- Aspect *Policies and Behavior*: the policies are procedures defining how to cope with reoccurring incidents while the behavior describes to which degree they are followed

Since the attainment of the aspects within the level of *IT-Risk Management* still cannot be directly measured further decomposition is required. For demonstrational purposes the decomposition of infrastructure related goals is subsequently traced (see Fig. 9):

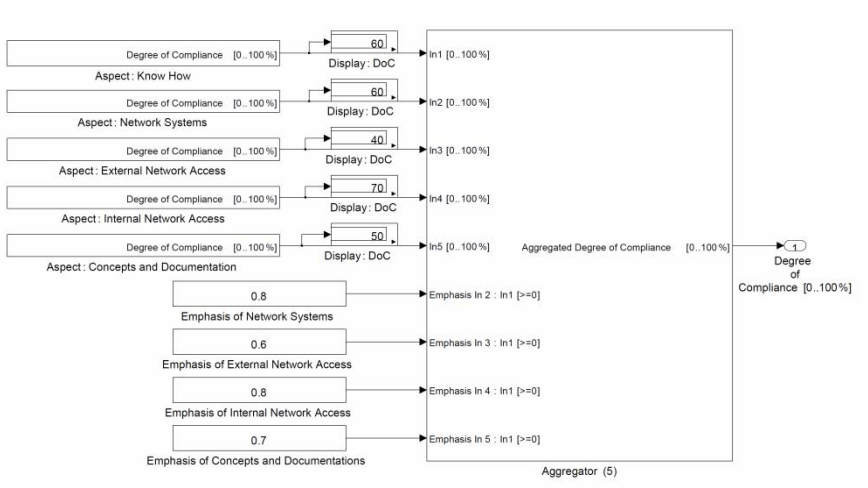


Fig. 9. Aspects of IT-Infrastructure, a sub goal within IT-Risk-Management

- Aspect *Know How*: issues regarding the professional knowhow / skill level of IT staff
- Aspect *Network Systems*: issues regarding the performance of the infrastructure and the ability to recover from severe incidents
- Aspect *External Network Access*: issues regarding remote access from unsecure networks
- Aspect *Internal Network Access*: issues regarding the local access to the organization's network
- Aspect *Concepts and Documentation*: issues regarding the existence and effectiveness of concepts und documentations for coping with risk related issues

Since our capital provider figured to measure know how the measurement level is reached for the aspect *Know How* (see Figure 10):

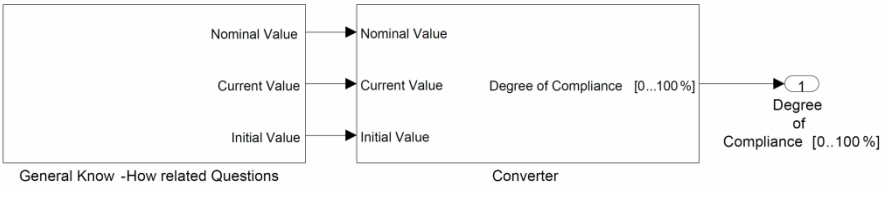


Fig. 10. Measurement within the aspect Know How

- The component *General Know-How related Questions* serves as a measurement procedure. It delivers values as specified in the metamodel: a *nominal value* expressing the target value, a *current value* being the result of the current measurement, and an *initial value* which is the worst value.

- The component *Converter* calculates a percentage degree from the results of the measurement procedure. It is a general component of the metamodel.

The inside of the component *General Know-How related Questions* is depicted in Figure 11:

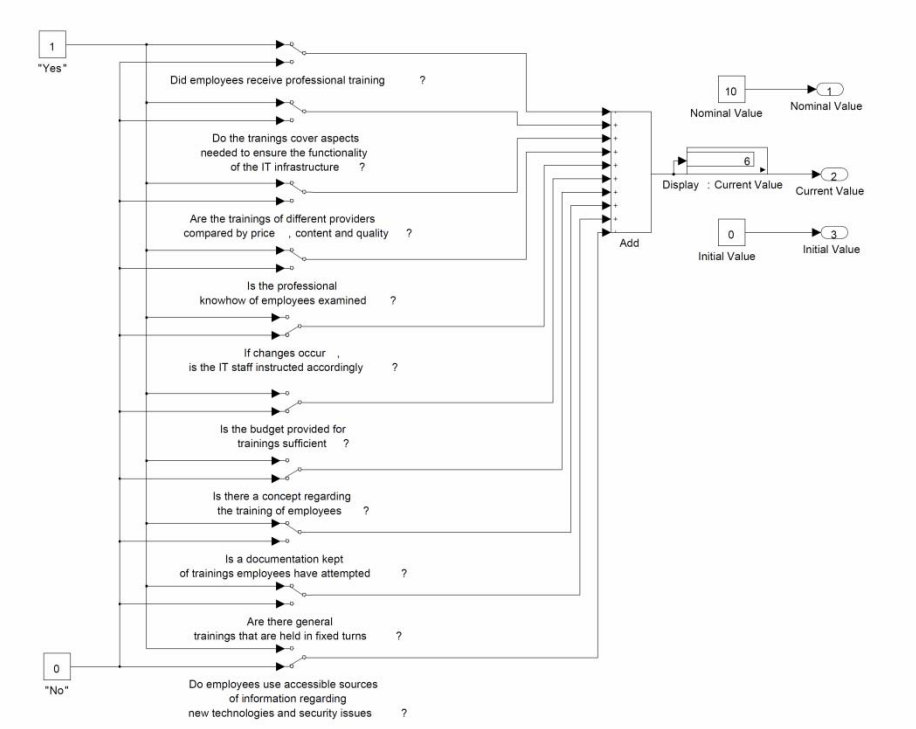


Fig. 11. Measurement procedure of general knowhow questions within the aspect K*nowHow*

There are ten questions. Each can be either "yes" or "no". Since all are treated as equally important, each "yes" counts one to the sum. Thus, if all questions are "yes" the *current value* would equal the *nominal value*. Of course, the worst situation is when all questions are "no" and consequently the *current value* would equal the *initial value*.

Instead of simple "yes/no" questions it is possible to integrate any measurement method. The only limitation is the requirement that for each measurement a *nominal*, *current*, and *initial* value is to be provided to the model.

We could do the validation, because the respective users were able to test model. This belongs mainly to two aspects. First, the testers were an integrated part of the model development whereby they could gain deeper knowledge about the model. Second, the simulation application allows to explore the model interactively, which helps to understand the concept. Thus, it is possible to address technical questions to the model and to test the aptitude in the intended application context. Our capital provider judges the model and metamodel to be valid for their intended applications.

4 Conclusion

The model presented in this paper is a contribution to the compliance measurement of IT governance goals. Additionally, an initial research work is pursued which assumes only the systematic splitting of the goals of IT governance to be possible. The measurability is accepted thereby as basic condition of the ascertainableness of the compliance. A metamodeling approach is chosen as a problem solution. An initial situation to the clarification of the problem understanding is described and enhanced by a requirement specification, which defines the solution goal. Thereupon, a solution idea was developed, which is examined in the context of the concept validation by intended users. This led to the design of a metamodel which serves as construction basis of specific models. The advantage of a metamodel is the clarity in its construction. The introduced metamodel is as simple as possible, in order to make the usage easily for users and to support an integration in the model construction process. The tests supported the examination not only in the context of the verification, but also in the understanding of the promoted model. Simulation parameters can be changed in real time and so the behavior model components can be explored interactively. That led in the validation step to the fact that the focus on the *learn-to-know* is an important part to understand the relation of the modeled components. Due to this, the users could examine technical questions of the IT-Governance within a short time using the proposed model. The validation results are positive, so that a company-wide announcement of the model is decided. The learnings of the validation example allow the conclusion for practical applications that abstract IT-governance objectives can be operationalized by applying the concept of decomposition. Since this process is arbitrary sufficient operationalization of goals depend on the individual judgment of users. For this reason no decisive rule can be provided that prescribes the process of IS-strategy operationalization. However, the metamodel presented in this paper is a general approach to support this process by providing application domain independent artifacts.

Further research work is pointed at the following aspects: Within the context of a deeper going system of objectives, organizations have to develop new ways of weighting partial goals. Zangemeister for example discusses a weight regulation systematic in the context of the *working-system-evaluation* [35]. Nevertheless, it remains unclear, which circles of acquaintances should be included into the goal education process. Furthermore, it has to be examined, how the influences of the IT-Governance stakeholder should be weighted. Moreover, approaches need to be developed that allow the inference of implementation sequences of objectives captured with model instances.

The metamodel is implemented in such a manner that a better integration into an operational area becomes possible. Additionally, the development of a domain specific language belongs directly to a software system. The metamodel itself is reflecting the domain of the language within its model. But within the referring of the tracking process of objectives for optimal compliance remains the question, in which way a decision support system for IT compliance has to be developed. The need and therefore the advantage is clear and is seen in a faster and more consistent decision making process. Due to this, further research activities are needed to define enterprise-wide decision support systems for IT Governance.

References

1. Antweiler, J.: Wirtschaftlichkeitsanalyse von Informations- und Kommunikationssystemen (IKS). Datakontext Verlag, Köln (1995)
2. Balci, O.: Verification, Validation and Certification of Modeling and Simulation Applications. In: Chick, S.E., Sanchez, P.J., Ferrin, D., Douglas, M. (eds.) Proceedings of the 2003 Winter Simulation Conference, pp. 150–158. Omnipress, Madison Wis (2003)
3. Balci, O., Sargent, R.G.: Some Examples of Simulation Model Validation Using Hypothesis Testing. In: Highland, H. (ed.) Proceeding of the Winter Simulation Conference, San Diego, pp. 621–629 (1982)
4. Banks, J., Carson, J., Nelson, B., Nicol, D.: Discrete-Event System Simulation. Pearson Prentice Hall, Upper Saddle River (2005)
5. Bhatt, G.: Managing Information Systems Competence for Competitive Advantage: An Empirical Analysis. In: Proceedings of the 24th International Conference on Information Systems, USA, pp. 134–142 (2003)
6. Biethahn, J., Mucksch, H., Ruf, W.: Ganzheitliches Informationsmanagement. Oldenbourg, München (1994)
7. Brynjolfsson, E.: The Contribution of Information Technology to Consumer Welfare. Information Systems Research 7, 281–300 (1996)
8. Carson, J.: Verification and Validation. In: Snowdon, J.L., Charnes, J.M. (eds.) Proceedings of the 2002 Winter Simulation Conference, San Diego, pp. 52–58 (2002)
9. Casolaro, L., Gobbi, G.: Information Technology and Productivity Changes in the Banking Industry. Economic Notes 36, 43–76 (2007)
10. Chan, Y.: Why Haven't We Mastered Alignment? The Importance of the Informal Organization Structure. MIS Quarterly Executive 1, 97–112 (2002)
11. Chan, Y., Huff, S., Barclay, D., Copeland, D.: Business Strategic Orientation, Information Systems Strategic Orientation, and Strategic Alignment. Information Systems Research 8, 125–150 (1997)
12. Chan, Y., Sabherwal, R., Thatcher, J.: Antecedents and Outcomes of Strategic IS Alignment: An Empirical Investigation. IEEE Transactions On Engineering Management 53, 27–47 (2006)
13. De Haes, S., van Grembergen, W.: IT Governance and Its Mechanisms. Information Systems Control Journal (2004)
14. Devaraj, S., Kohli, R.: Performance Impacts of Information Technology: Is Actual Usage the Missing Link? Management Science 49, 273–289 (2003)
15. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns. Addison-Wesley, Boston (1995)
16. Gass, S.: Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. Operations Research 31, 603–631 (1983)
17. Henderson, J., Venkatraman, N.: Strategic Alignment: A Framework for Strategic Information Technology Management. Center for Information Systems Research, Cambridge (1989)
18. IT Governance Institute, COBIT 4.0, Control Objectives, Management Guidelines, Maturity Models, released 2005 (2005), `http://www.isaca.org` (last access 4.3.2007)
19. Johannsen, W., Goeken, M., Just, D., Tami, F.: Referenzmodelle für IT-Governance. dpunkt-Verlag, Heidelberg (2007)
20. Kearns, G.: Strategic IT Alignment: A Model for Competitive Advantage. In: Applegate, L., Galliers, R., DeGross, J.I. (eds.) Proceedings of the 22. International Conference on Information Systems, Barcelona, pp. 1–12 (2001)

21. Kramer, U., Neculau, M.: Simulationstechnik. Hanser, München (1998)
22. Law, A., Kelton, W.: Simulation modeling and analysis. McGraw-Hill, Boston (2000)
23. Lee, S.: Modeling the Business Value of Information Technology. Information & Management 39, 191–210 (2001)
24. Lutter, M.: Deutscher Corporate Governance Kodex. In: Hommelhoff, P., Hopt, K., von Werder, A. (eds.) Handbuch Corporate Governance, pp. 737–748. Schmidt, Köln (2003)
25. Mukhopadhyay, T., Kekre, S., Kalathur, S.: Business Value of Information Technology: A Study of Electronic Data Interchange. MIS Quarterly 19, 137–156 (1995)
26. Nagel, K.: Nutzen der Informationsverarbeitung. Oldenbourg, München (1990)
27. Object Management Group, Meta Object Facility Core Specification, released 01.01.2006 (2006), http://www.omg.org/technology/documents/modeling_spec_catalog.htm#MOF (last access 15.08.2007)
28. Roepke, R., Agarwal, R., Ferratt, T.: Aligning the IT Human Rsources ith Business Vision: The Leadership Initiative at 3M. MIS Quarterly 24, 327–353 (2000)
29. Schmidt, R., Weiß, M.: Shareholder versus Stakeholder. In: Hommelhoff, P., Hopt, K., von Werder, A. (eds.) Handbuch Corporate Governance, Schmidt, Köln, pp. 107–127 (2003)
30. Sedgewick, R.: Algorithmen. Addison-Wesley, Bonn (1998)
31. Sommerville, I.: Software engineering. Addison-Wesley, Boston (2007)
32. Totok, A.: Entwicklung einer Business-Intelligence-Strategie. In: Chamoni, P., Gluchowski, P. (eds.) Analytische Informationssysteme. Springer, Berlin (2006)
33. Weill, P., Ross, J.: IT governance. Harvard Business School Press, Boston (2004)
34. Werder, A.: Ökonomische Grundfragen der Corporate Governance. In: Hommelhoff, P., Hopt, K., von Werder, A. (eds.) Handbuch Corporate Governance, Schmidt, Köln, pp. 3–27 (2003)
35. Zangemeister, C.: Erweiterte Wirtschaftlichkeitsanalyse (EWA). Wirtschaftsverlag NW, Bremerhaven (1993)

The Implementation of European Best Execution Obligations
- An Analysis for the German Market -

Peter Gomber, Gregor Pujol, and Adrian Wranik

Goethe University Frankfurt, E-Finance Lab e.V.,
Robert-Mayer-Str. 1, D-60054 Frankfurt/Main
{gomber,pujol,wranik}@wiwi.uni-frankfurt.de

Abstract. Since 01 November 2007 the provisions of the "Markets in Financial Instruments Directive" (MiFID) have to be applied by investment firms and regulated markets. This key date marks the (provisional) completion of a legislative process of the European Union which took several years. One major aspect of regulation concerns the best execution of client orders. Investment firms are obliged to make adequate provisions including processes and IT systems for order routing to achieve the best possible result ("best execution arrangements") and to disclose sufficient information of the most important measures to their clients ("best execution policies"). This paper presents the results of an analysis of 75 best execution policies from the 100 largest German financial institutions and the 15 largest online brokers in Germany with regard to the fulfillment of the statutory minimum requirements, identifiable "best practices" and the execution venues on which investment firms place significant reliance in meeting the best execution requirement.

Keywords: MiFID, best execution, best execution policy, order routing systems.

1 Introduction

From 01 November 2007 the provisions of the "Markets in Financial Instruments Directive" (MiFID) have to be applied by all companies which provide investment services as well as by all regulated markets throughout Europe. The central innovations of the MiFID are the classification of trading venues (regulated markets, multilateral trading platforms, systematic internalisers), the definition of "best execution" at European level, the introduction of far-reaching transparency regulations for OTC trading.

Investment firms are obliged to make adequate provisions including processes and IT systems for order routing for achieving the best possible result ("best execution arrangements") and to disclose sufficient information of the most important measures to their clients ("best execution policies"). Although "best execution" and the associated duty to inform and accounting obligations initially constitute a legal obligation in the relationship between client and investment firm, at the economic level this topic also decisively affects the interface between investment firms and execution venues.

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 126–144, 2009.

Because of the new regulations for “best execution”, the communication of the execution venues' performance capability, particularly in terms of price quality and execution costs, has become a major competitive factor for the execution venues with regard to the investment firms.

Best execution is discussed from different perspectives in academic literature, especially for US markets from a viewpoint of how best execution can be realised and measured. Macey and O’Hara analyzed legal and economic aspects of the duty of best execution and recommended that best execution for a particular trade is best achieved through competition between trading venues [1]. However, McCleskey suggested that best execution should be subject to regulation, as investors are not capable to evaluate execution quality due to limited access to appropriate information [2]. Furthermore, a number of papers examined costs as a key aspect involved with best execution for a single trading venue [3,4] as well as between different markets [5, 6]. These studies basically concluded that execution costs could be measured. Further research by Bacidore et al proposed six measures to quantify best execution [7].

With MiFID, the European Regulator addresses many of these viewpoints and particularly intends to harmonize the best execution requirements on a pan-European level. The new regime is characterized by a large diversity of influencing factors that investment firms have to consider in their order execution processes and systems and includes all financial instruments and all types of execution venues.

A survey among 200 investment firms in Germany in spring 2007 revealed that for German financial institutions MiFID is more of a regulatory obligation than a chance to leverage new competitive potentials [8, 9]. In the opinion of the institutions surveyed, competitive differentiation can be achieved best through the design of the best execution policies: 32% of the institutions which replied regarded this aspect as having a very high or fairly high competitive potential, and it was thus considered to have the best chances of all services connected with the introduction of MiFID.

Against this background the paper at hand for the first time documents the status quo of best execution policies *after* MiFID has entered into force in late 2007. As the European Directive had to be transformed into individual national legislations by all European member firms a consistent view can only be achieved for one member state. For this study, Germany was chosen as it represents the largest European economy. On the one hand the various “best execution policies” are analyzed as a whole and compared with each other. On the other hand this study checks how far the institutions' earlier assessment of the competitive potential of the topic “order execution” is also actually reflected in the best execution policies which were evaluated. In addition, it is examined whether firms selected a static or a dynamic approach as their preferred implementation option for the best execution requirement.

In chapter 2, the static and the dynamic approach as general implementation concepts for the best execution obligation are introduced. The composition of the sample and the methodology selected to analyze the best execution policies are explained in chapter 3. Chapter 4 presents the empirical results with regard to the legal requirements and chapter 5 describes further contents which were addressed by the investment firms in their best execution policies. The analysis of the information about the execution venues is provided in chapter 6. Chapter 7 summarizes the most important results of this analysis.

2 Implementation Options for Best Execution Processes

MiFID implies a number of challenges that affect the existing operational processes and systems in many financial institutions. The effort to implement these amendments highly depends on the business model and services provided by the individual firm. Fig. 1 provides an overview regarding the influence of MiFID on different areas of the securities business value chain [10]. Besides pre-trade and post-trade transparency requirements, order routing and best execution represent key aspects with a high level of impact. While order routing includes the decision of implementing a static or dynamic approach to deal with incoming client orders, best execution addresses the requirement to define a process that supports the paramount best execution target of MiFID [11].

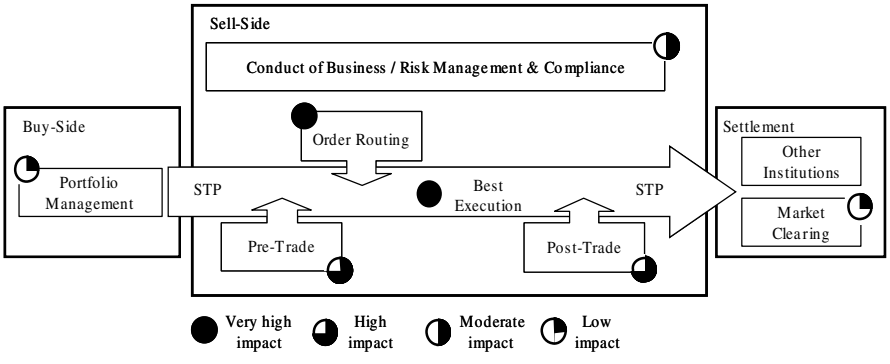


Fig. 1. MiFID impact on IT and business [10]

Although broadly discussed in the financial community, a clearly defined practice for the implementation of the best execution obligation has not been established yet. In general, two basic concepts have emerged. On the one hand investment firms may apply a static approach, i.e. the decision to route a client order is based on a pre-defined rule framework. On the other hand a dynamic approach represents a more sophisticated option considering real time market data for the order routing decision.

The rule framework is based on one or more order attributes and a best execution venue is pre-defined. MiFID leaves the definition of these rules to the investment firms. Depending on their product range and client structure different combinations of criteria such as client type, nature of the order or instrument class may be considered. However, the interpretation of the Committee of European Securities Regulators (CESR) suggests a minimum level of differentiation including the venue selection for different classes of instruments (security groups) and the distinction between retail and professional clients, if treated differently. Further differentiation of the execution policy is not mandatory and the decision is

left to the investment firm, for example to further distinguish by order type [12]. Thus, the minimum set up of such a rule framework takes into account two aspects: First, different classes of instruments (security groups) need to be defined, i.e. instruments with similar characteristics (stocks, bonds, derivatives, etc.) are grouped into one category. Such a category may also cover instruments of an entire market segment, a certain index or even according to geographical origin. Second, the different instrument classes are linked to client categories (retail/ professional), if a distinction is necessary due to different treatment of the orders. For example, all retail client orders in domestic shares (criteria) are executed in the most significant national stock exchange [10].

Fig. 2 illustrates such a static approach where investment firms have to select and list in their best execution policy those venues that on a consistent basis achieve the best possible result for the execution of client orders. MiFID does not prescribe explicitly a certain number of venues. The venue selection process follows a three-step process. First, all existing venues have to be identified, i.e. Regulated Markets, Multilateral Trading Facilities as well as Systematic Internalizes (existing venues). In the next step investment firms will have to decide which of these venues they wish to offer to their clients for transaction purposes (offered venues). It is likely that the number of offered venues is smaller than the number of existing venues due to costs of access. Finally, a subset has to be extracted from the population of offered venues. This sample then constitutes those venues that enable the firm to achieve the best possible result for client orders (best execution venues). Nevertheless, clients can use specific instructions to access venues in the sample of offered venues.

For the evaluation of the different execution venues MiFID prescribes to apply certain criteria, e.g. price, cost, speed, etc. and additionally requires a weighting of these criteria according to the attributes client type, financial instrument and nature of the order. As a result of this process an investment firm may define a rule framework as shown in Table 1. In this example exactly one execution venue is assigned to one securities group. For example, for the securities group stocks Venue A is considered to provide the best possible result on a consistent basis.

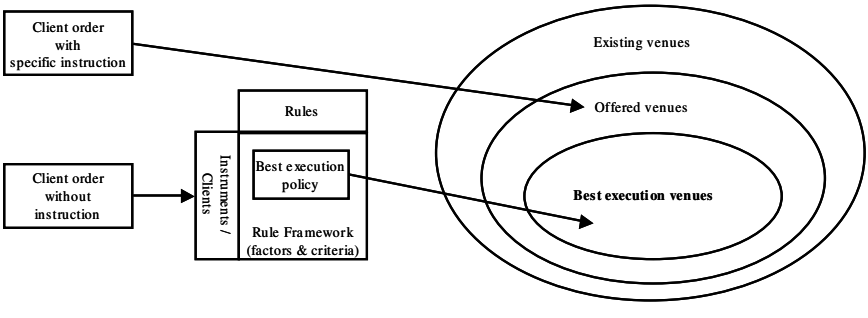


Fig. 2. Venue selection [10]

Table 1. Static approach – sample rule framework

Sample rule frame work				
Securities group	**Rank 1**	**Rank 2**	**Rank 3**	**…**
Stocks	Venue A	Venue B	Venue C	…
Bonds	Venue B	Venue A	Venue C	…
Derivatives	Venue C	Venue B	Venue A	…
...	…	…	…	…

Alternatively, the appropriate execution venue is selected by applying a dynamic approach which is an assumption that exceeds the current MiFID requirements. Instead of executing client orders according to the process defined in the execution policy each order is treated in an individual manner. Unlike the static approaches current market data feeds are attached to the rule framework and investment firms need to interface with providers to receive the relevant market and/or order book information. Such provisions enable a real-time evaluation and support a dynamic allocation to the execution venue offering the best conditions at the time of order entry. The real time comparison allows an optimized selection between venues increasing the quality of an order execution but it obviously leads to higher costs of implementation.

Fig. 3 shows an example of such a dynamic approach [13]. Smart Order Routing systems (SOR) access multiple liquidity pools, i.e. exchanges or alternative trading systems, to identify the best destination and apply proprietary algorithms to optimize order execution. They gather real-time data from the respective venues (A, B and C) concerning their order book situations, i.e. current quoted volumes and prices. Based on this information the routing engine slices incoming orders and decides where to route individual suborders in respect of the best prices available in that logical second. Then, the routing engine receives order confirmations or notes of unfilled orders. In the latter case, the routing process will be repeated or cancelled depending on a customer's instructions.

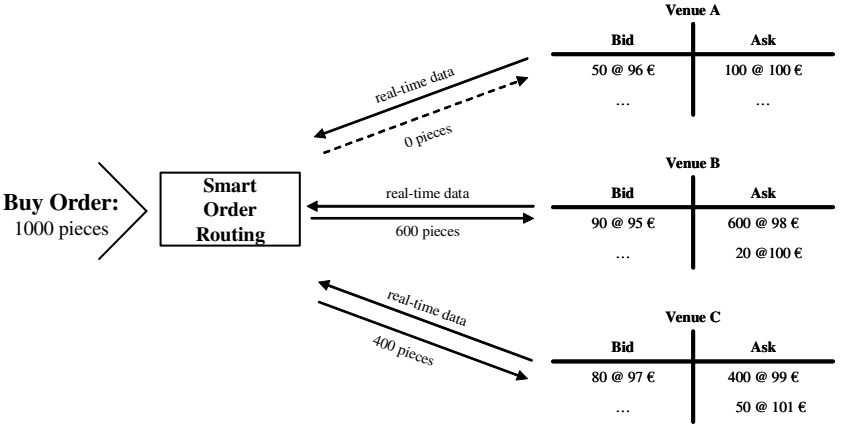


Fig. 3. Dynamic approach – real-time execution [13]

3 Sample and Procedure

The starting basis for this study is provided by the 100 largest German financial institutions in terms of their total assets in 2006 [14] and the 15 largest online brokers in Germany according to the number of security accounts [15]. The list of the 100 largest financial institutions was adjusted by removing the companies which do not provide any investment services (e.g. mortgage banks and home loan banks). After this correction, the original number of financial institutions with securities business reduced to 63. The final sample size for this group totals 60 investment firms because three best execution policies were not made available. The second group of online brokers includes a total of 15 companies leading to an overall total of 75 best execution policies (60 financial institutions and 15 online brokers) that were analyzed in the second quarter of 2008.

Various channels were selected for obtaining the policies. In addition to research over the Internet, the policies were collected by email or telephone contacts.[1]

A comprehensive list of criteria was developed for the analysis which was based on the legal requirements, but which also contains other aspects which resulted from the practical implementation by the investment firms. These include, for example, the explanations of the factors which were used to evaluate the execution venues. A second part of this analysis then focuses on the execution venues which were named. The execution venues were examined with regard to the assignment of execution venues to categories of financial instruments and with regard to the existence of a ranking of various execution venues in the policies. While in the text percentage values always refer to the total sample size (75 best execution policies) absolute figures are used to explain findings of a sub-sample (e.g. best execution policies applicable for private clients only.)

4 Minimum Legal Requirements for the Best Execution Policies

The best execution policies are a major part of the provisions which investment firms must make in order to ensure that they can regularly execute client orders in the best possible manner. The most important legal requirements of the MiFID implementation in Germany are specified in the German Securities Trading Act (WpHG)[2] and the Ordinance Specifying Rules of Conduct and Organisation Requirements for Investment Firms (WpDVerOV)[3].

In accordance with §33a (6) No. 1 WpHG an investment firm must inform *"its clients of its best execution policy before providing investment services for the first time and obtain the clients' acceptance of its policy."*

---

[1] Companies displayed differing degrees of willingness to cooperate. Some stated that they only provided this information - which is in principle public – to their existing clients.

[2] The WpHG in the version of 01 November 2007, amended by the law for implementing the Markets in Financial Instruments Directive (DIR 2004/39/EG, MiFID) and the implementing directive (DIR 2006/73/EG) of the Commission (law for implementing the financial markets directive) of 16 July 2007, Federal Law Gazette I 2007, 1330 of 19 July 2007.

[3] Ordinance Specifying Rules of Conduct and Organisation Requirements for Investment Firms (WpDVerOV) of 20 July 2007, Federal Law Gazette I 2007, p. 1432 of 23 July 2007.

Specific minimum requirements with regard to the scope and design of the contents are linked to this obligation. These are specified in §33a (5) and (6) WpHG in conjunction with §11 (4) WpDVerOV. On the one hand investment firms must show how they aim to achieve the best possible result for order execution in various categories of financial instruments. On the other hand the crucial factors for selecting a particular execution venue must be named, and finally a list containing at least those execution venues must be provided which can be considered for consistently achieving the best possible results. Fig. 4 shows the results for the minimum legal requirements which - irrespective of the business model concerned - apply for all investment firms. It displays the requirements and the levels of fulfillment. This is followed by a brief explanation of the individual results.

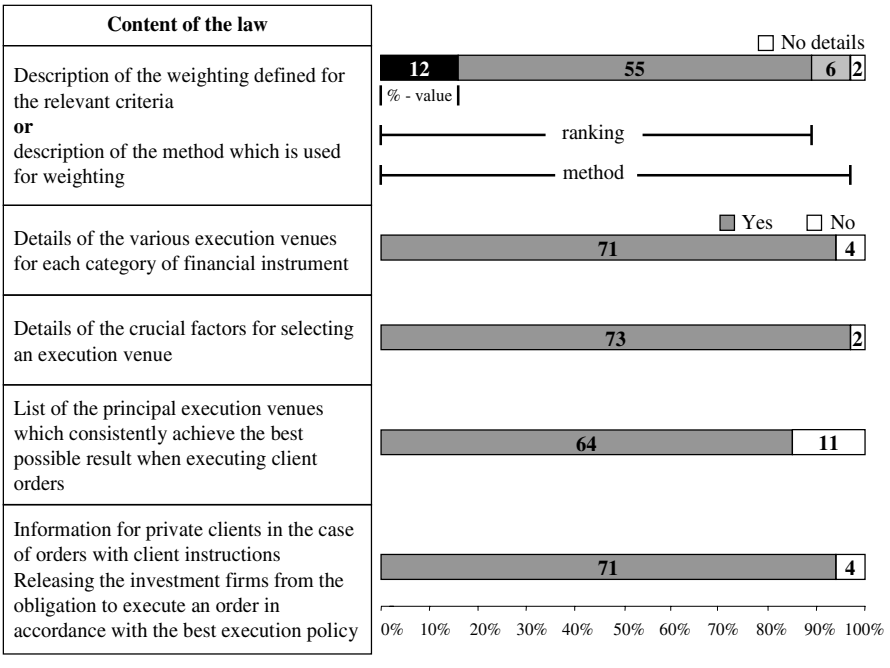


Fig. 4. Minimum legal requirements – irrespective of the particular business model

Description of the Weighting Implemented or Description of the Method

The basis for this requirement is provided by § 33a (6) No. 1 WpHG in conjunction with §11 (4) No. 1 WpDVerOV, according to which the policies must contain either a *"description of the weighting implemented for the relevant criteria to achieve the best possible result"* or "*a description of the method which is used for this weighting.*"

In the analysis a distinction is made between three examination criteria (percentage, ranking, method). Best execution policies which contain at least information on a particular procedure are acknowledged to have a method (e.g. "the bank assumes that their clients would like to achieve the best price" or "special importance is assigned to individual criteria"). That is the case with all investment firms (97.3%) except for two which provided no details on this.

The study examined whether the weighting defined was expressed as a percentage value or by a particular ranking. In 16.0% of the best execution policies concrete percentage values are specified for individual criteria (e.g. price: 80%; external costs: 20%), in 73.3% a ranking, e.g. price has priority over speed, is provided.

Details of Execution Venues for each Category of Financial Instrument
§33a (5) No. 1 WpHG (first half sentence) requires *"details of the various execution venues with regard to each category of financial instrument."* This provision specifies that a category must be assigned to the execution venue(s) when a best execution policy is created. The following variants of best execution policies exist: either precisely one execution venue is specified for one category or several execution venues are specified for one category or several execution venues are specified for multiple categories simultaneously. In 94.6% of the best execution policies an assignment of category to execution venue can be recognized.

Details of the Crucial Factors for Selecting an Execution Venue
In addition to linking the category and execution venue, in accordance with §33a (5) No. 1 WpHG (second half sentence), *"the crucial factors for selecting an execution venue"* must also be specified. In the analysis, information was taken into account which documents that the investment firms used various factors for the evaluation (e.g. "in particular the recognizable factors price and costs which arise through execution at an execution venue are used for the evaluation."). In nearly all policies (97.3%) comprehensible details are provided here.

List of the Principal Execution Venues
As a result of the analysis of various execution venues by the investment firms, a further obligation was derived from §33a (5) No. 2 WpHG in conjunction with §11 (4) No. 2 WpDVerOV according to which a *"list of the principal execution venues* […] *at which the investment firms can consistently achieve the best possible results when executing client orders"* must be contained in the policies. In 85.3% of the policies examined this list is provided either as a text list or as a table in the appendix.

Information for Private Clients about Execution in accordance with Instructions
Finally, in accordance with §33a (6) No. 2 WpHG in conjunction with §11 (4) No. 3 WpDVerOV investment firms must inform "*private clients expressly that when instructed by the client the investment firm will execute the order according to these client instructions and will therefore not be obliged to execute the order in accordance with its best execution policy to achieve the best possible result."* This information is clearly emphasized in most best execution policies (94.7%).

In addition, three further minimum legal requirements were examined which - depending on the particular business model - do not apply for all investment firms. Fig. 5 shows the results of implementing the regulatory specifications when client orders are forwarded to third parties, when financial portfolio management is offered, and in the case of OTC business.

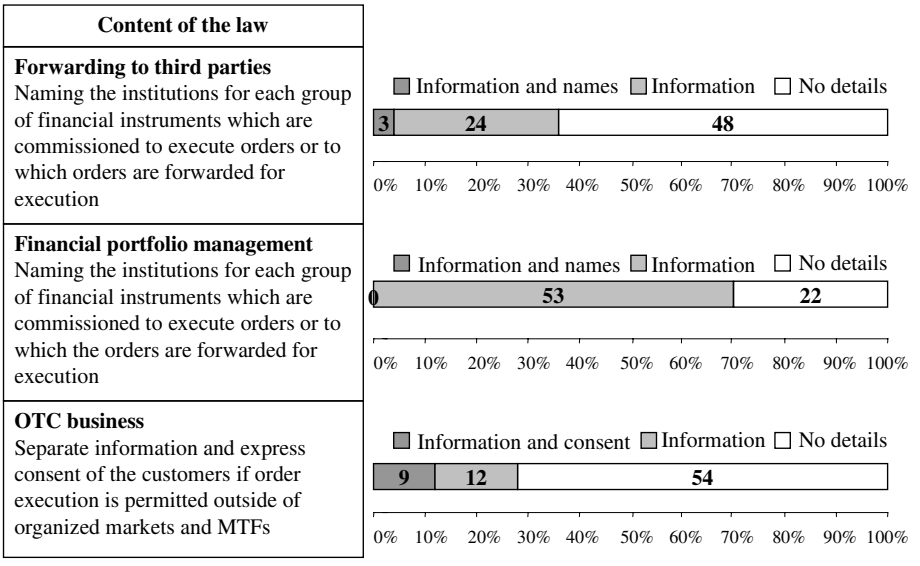


Fig. 5. Minimum legal requirements – depending on the particular business model

Special Feature when Forwarding Client Orders to Third Parties

An investment firm which forwards client orders to third parties must, in accordance with §33a (8) No.2 WpHG, "*with regard to each group of financial instruments, name all institutions which commission* [the investment firm] *to execute its decisions or to which it forwards the orders of its clients to be executed there."* In the analysis a check was first made to see whether information on forwarding to third parties existed. In the case of investment firms which commission third parties the study also examined whether it was merely noted that the order was to be forwarded (e.g. "forwards foreign securities which cannot be traded in Germany…") or, as prescribed by law, the institution commissioned to execute the order was also named.

The analysis in Figure 2 shows a very heterogeneous implementation of this ordinance. No clear statements can be made for around two thirds of the policies which were examined since the investment firms concerned provided no details regarding the forwarding of client orders to third parties. Information that orders were forwarded was found in around one third of the firms' best execution policies. In 4% of the best execution policies this information is also supplemented by the actual name of the institution to which client orders are forwarded.

Special Feature of Financial Portfolio Management

The same statutory specifications of §33a (8) No. 2 WpHG apply for investment firms which offer financial portfolio management as for forwarding orders to third parties. The analysis logic is therefore identical. 29.3% of the policies contain no details of financial portfolio management. In the remaining 70.7% of the investment firms which provide information on financial portfolio management the institution which was commissioned to execute an order was not named in any policy. This result can in part be explained by the fact that in many cases reference is made to the regulations for the asset management contracts involved which are not the subject of this study.

Special Feature of OTC Business
Finally, §33a (5) WpHG regulates how the clients are to be informed in the case of OTC business. *"If the best execution policies* […] *also permit order execution outside of regulated markets and multilateral trading facilities, the investment firm must inform its clients separately of this circumstance and obtain their express consent, either generally or for each business transaction, before the client order is executed at an execution venue."*

The policies were checked to see whether they contained explicit information on OTC execution. In addition, a check was made to see whether the information relating to the obligation to obtain consent for business in general or for each individual transaction could be recognized. 72.0% of the investment firms provide no information on OTC business. In 16.0% of the policies the option of OTC business is pointed out. In 12.0% both information on and a reference to obtaining prior express general consent or consent for each business transaction was provided.

5 Special Aspects of the Best Execution Policies

In addition to the minimum legal requirements, many investment firms also provide clients with further information. This information reveals further details on appropriate measures and processes which an investment firm must take internally but is not obliged to publish to clients. The results show that a "best practice" for many such measures and the communication thereof has become established which a large number of investment firms have followed in implementing their policies. In the following details relating to the scope of the best execution policies and the relevant criteria for achieving the best possible result and their weighting will be examined.

Scope
The policies examined display differences with regard to the scope for particular client groups. A definition of categories is contained in §31a (1-3) WpHG. According to this the clients are *"all natural or legal persons for whom the investment firms provide or initiate investment-services or ancillary investment services."*

Professional clients are clients *"of whom investment firms can assume that they have sufficient experience, knowledge and expertise to enable them to take their investment decisions and to assess the associated risks appropriately."* This definition is complemented in §31a (2) WpHG by a sample list of investment firms concerned and companies which are not subject to approval or supervision because they satisfy at least two of the three criteria total assets, sales revenues and capital resources.[4] Private clients are all clients who *"are not professional clients."*

Fig. 6 shows the distribution to the categories clients, private clients and professional clients. 17 best execution policies contain only details for private clients. 38 policies apply expressly both for private clients and for professional clients. No distinction between client categories can be recognized in the 20 remaining best execution policies. To classify the scope these policies use the term "Clients". For this analysis all the policies which speak in general terms of clients as a target group apply both for private clients and professional clients.

[4] Professional clients are firms which exceed two of the three criteria below: Total assets >20,000,000 Euro; sales revenues >40,000,000 Euro; capital resources >2,000,000 Euro.

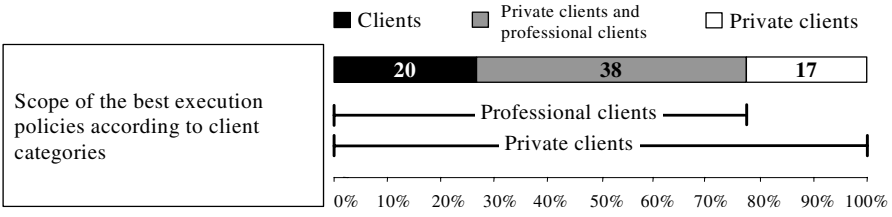


Fig. 6. Scope according to client category

Relevant Criteria and Recognizability of the Weighting

The relevant criteria for achieving the best possible result are derived directly from the wording of §33a (2) WpHG, in accordance with which "*in particular the prices of the financial instruments, the costs involved in order execution, the speed, the probability of execution and the processing of the order, as well as the scope and type of order"* must be taken into account as criteria. A check was made to see whether the relevant criteria had been weighted in the best execution policies (Table 2):

No weighting can be recognized in 10.7% (absolute: 8) of the policies. In two cases here no details at all are provided, in the other six cases the criteria have not been prioritized. Weighting of the relevant criteria can be recognized in 89.3% of the best execution policies (absolute: 67). In 12 best execution policies specific percentage values are named for individual criteria; in 55 policies a ranking, e.g. price has priority over speed, can be observed. In the group mentioned last one policy stands out on account of the execution venue being based on individual orders and the usage of real-time market data for selecting the execution venues providing the best result. In this case the weighting of the individual criteria is variable and depends decisively on the order attributes, e.g. the type of financial instrument, quantity, etc.

Table 2. Recognizability of the weighting of the relevant criteria

Number of policies evaluated			75
No details	2	No recognizable weighting	8
No ranking of criteria recognizable	6		
Ranking of the criteria recognizable **and** dynamic order execution (real time)	1	Recognizable weighting	67
Ranking of the criteria recognizable	54		
Percentage weighting of the criteria	12		

Concerning the discussion of the different implementation options for best execution processes and systems table 2 shows an important result: The majority of investment firms (except one) prefer a static implementation approach for the best execution requirement and thus do not utilize the competitive potential stated in recent studies. Instead of applying dynamic order routing they prefer a static rule based best execution approach.

The 67 best execution policies with a recognizable weighting were analyzed again in detail, with the focus on the selected ranking of the various criteria.[5] In this context the requirement of §33a (3) WpHG must be mentioned, according to which during order execution for private clients "*the best execution policies [must] contain* provisions *to ensure that the best possible result is based on the total consideration.*"

Table 3 contains the results with regard to the ranking of the relevant criteria for 67 best execution policies.[6] First it must be stated that, for private clients, 65 best execution policies are - as required by law - largely based on the total consideration, i.e. price plus the costs involved in order execution (external costs). At the same time, the criterion "price" is always ranked first in these cases. Two policies deviate from this principle. In one case an investment firm ascribes the greatest importance solely to the price. In another case the total costs, i.e. in-house charges and external costs, occupy Rank 1, without taking the price into account. This results in a total frequency of 66 nominations for the criterion "price." The second important criterion "external costs" is nominated a total of 65 times, 55 times in first place (always together in first place with price) and 10 times in second place.

Table 3. Ranking of the relevant criteria

Criterion	Frequency	Rank 1	Rank 2	Rank 3	Rank 4
Price*	66	66	0	0	0
All costs	1	1	0	0	0
External costs*	65	55	10	0	0
In house charges	34	0	34	0	0
Speed	56	2	18	35	1
Execution probability	55	2	18	35	0
Processing probability	7	1	3	3	0
Processing certainty	48	1	5	42	0
Complete execution	14	0	14	0	0
Liquidity	1	0	0	1	0

* Based on total consideration.

A second finding concerns the selection of the criteria. In addition to the statutory specifications (price, external costs, speed, execution probability, processing probability), slightly modified decision parameters are also used by the investment firms and explained in the policies concerned. These include the details relating to costs, in-house charges, processing certainty, complete execution and liquidity.

[5] 42 investment firms describe a process when several execution venues are considered simultaneously for executing a client order to achieve the best possible result.

[6] The first column contains both the legally prescribed criteria and decision parameters which are also used by the investment firms, e.g. processing certainty, liquidity. The second column shows the total number of times a criterion is named. The priorities of the criteria can be seen in the following columns.

Weighting of the Relevant Criteria Taking into Account the Attributes

The weighting itself was analyzed according to the specifications of §33a (2) WpHG "*taking into account the attributes of the client, of the client order, of the financial instrument and of the execution venue."* The results below show whether taking these attributes into account can be recognized in the policies. It must be borne in mind here that by suitable internal measures and processes, the investment firms are obliged to take each particular attribute into account, but communication with their clients is not mandatory. Fig. 7 shows to what degree it can be recognized that the individual attributes have been taken into consideration in the policies.

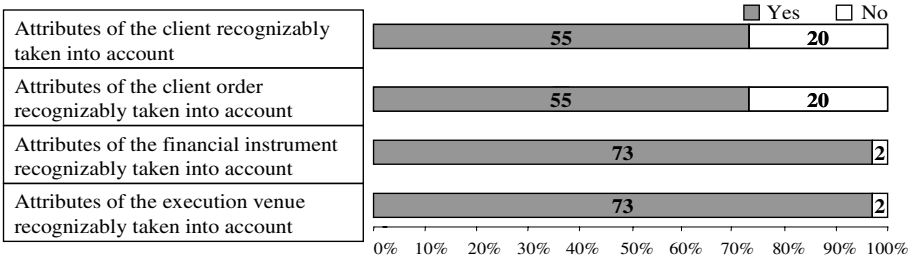


Fig. 7. Presentation / recognizability of the attributes in the best execution policy

In virtually all best execution policies (97.3%) it is possible to recognize that both the attributes of the financial instrument and the attributes of the execution venue have been taken into account. This unambiguous result can be explained by the minimum legal requirements: the attribute of the financial instrument can be recognized because details of various categories (e.g. shares, bonds) must be provided in the best execution policies. The fact that the attributes of the execution venue have been taken into account is also easy to recognize from the naming of various execution venues (list) for one or more categories of financial instruments.

Analyzing the policies makes it slightly more difficult to recognize the attributes of the client or of the client order. Details of this are given in 73.3% of the policies (absolute: 55). The results for these two attributes are presented in Fig. 8.[7]

The client attribute is frequently simplified and reduced to the classification between private clients and professional clients. However, the specific variant of the client group concerned is decisive for the particular client category. If, for example, the policies of the two groups apply equally, this means that they are treated equally and it can therefore not be claimed that the attributes of the client are taken into account. If, however, the scope of the policy is restricted to private clients, a distinction is made and the client's attributes are therefore taken into account.

A number of variants can be recognized for the client's attribute with regard to the scope, the different obligation to provide instructions and the weighting of the criteria: clients are most frequently distinguished from each other via the scope of the policies. Approx. one third of the policies (absolute: 17) apply exclusively for private clients.[8]

[7] These results are based on 55 policies which contain details about these two attributes.

[8] Taking this client attribute into account is consistent with the 17 nominations which are presented in Figure 3: Scope according to client category.

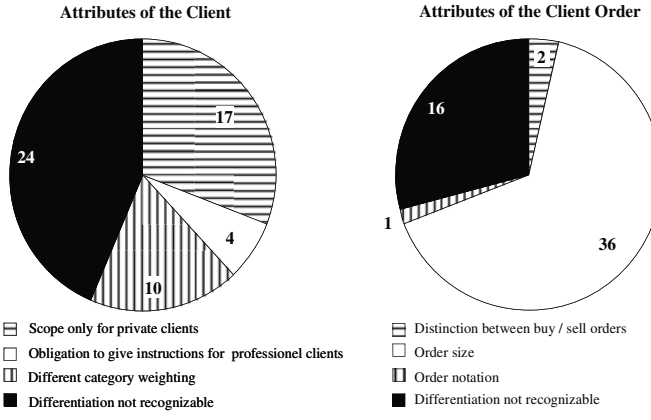


Fig. 8. Attributes of the client and attributes of the client order

Four policies do also apply for professional clients, but only when instructions are provided. Finally, 10 investment firms differentiate their clients by weighting the relevant criteria differently. Thus, for example, only the two criteria price and costs are taken into account for private clients. Additional criteria such as speed and execution probability are taken into account for professional clients. In 24 policies it is mentioned that client attributes are taken into account, but there is no description of an actual occurrence. For the attribute of the client order, variants relating to the distinction between buying and selling, order size and order notation can be observed. 36 policies specify various guaranteed order sizes for various categories of financial instruments. In two cases the attribute of the client order is taken into account on the basis of the distinction between buy / sell order or as in one policy in relation to the order notation. Finally, 16 investment firms state that they take the attribute of the client order into account, but give no further details of any particular occurrence.

6 Analysis of the Execution Venues Specified

Finally, details regarding the ranking of the execution venues were examined. Obviously, from a competitive point of view this ranking is highly relevant for execution venues. Therefore, it is of high interest how concrete and detailed policies list and rank the different execution venues.

Table 4 shows the distribution of the nominations for various securities groups, classified according to whether they can be traded domestically or abroad. It is noticeable that the investment firms primarily prefer abstract and summarizing descriptions to document their choice of an execution venue (e.g. domestic execution venue, foreign exchange) instead of naming specific execution venues. In addition to the abstract details relating to the execution venues, the table also includes special cases, such as forwarding, fixed-price business and also obligations to provide information. A concrete execution venue is named only in every fourth policy (24%).

Table 4. Execution Venues for different securities groups

Tradable domestically	**Shares**	**Bonds**	**Certified derivatives**	**Investment shares**	**ETF**	**Other securities**	**Other financial instruments**
Domestic execution venue	43	43	42	5	4	10	4
Domestic floor trading system	4	5	5	1	4	0	0
Domestic exchange	6	10	11	6	9	5	6
Domestic home exchange	0	2	2	0	0	0	1
Others (e.g. forwarding)	1	4	5	1	0	1	1
Fixed-price business	1	3	1	0	0	3	0
Instructions	0	1	1	46	2	34	35
No details	2	1	3	13	48	19	9
Not possible	0	0	0	2	1	0	1
Specific execution venue	**18**	**6**	**5**	**1**	**7**	**3**	**18***
Total	75	75	75	75	75	75	75

Tradable abroad	**Shares**	**Bonds**	**Certified derivatives**	**Investment shares**	**ETF**	**Other securities**	**Other financial instruments**
Foreign execution venue	2	5	3	1	1	2	12
Foreign exchange	27	17	17	0	0	10	1
Others (e.g. forwarding)	4	6	8	2	2	2	2
Fixed-price business	2	1	0	0	0	1	2
Instructions	35	36	34	46	46	34	34
No details	4	10	13	24	24	26	23
Not possible	0	0	0	2	2	0	1
Specific execution venue	**1**	**0**	**0**	**0**	**0**	**0**	**0**
Total	75	75	75	75	75	75	75

* 16 nominations here for Eurex.

For the categories of financial instruments which can be traded domestically, the abstract formulation “domestic execution venue” is chosen most frequently for the first three securities groups (shares, bonds, certified derivatives). For investment shares, other securities and other financial instruments orders are in most cases accepted only with express instructions. Investment firms most frequently specify - with 18 nominations each - a specific execution venue for the securities groups shares and other financial instruments; in the last group Eurex was named 16 times. The term “foreign exchange” is frequently used for financial instruments which can be traded abroad; with one exception, no specific details are provided here. It is noticeable in all the securities groups that order execution is normally only possible with instructions, or no details about the execution venues are provided at all.

Because of the few specific details about the execution venues and their ranking, a more detailed analysis was only performed for the securities group shares. In Table 5 18 best execution policies are listed which named at least one specific execution venue for the securities group and documented one recognizable ranking. Specifically, the segments[9] DAX 30, other DAX (MDAX, TECDAX, SDAX), EUROSTOXX 50, DJ STOXX 40, NASDAQ 100 and other domestic shares were analyzed.

Table 5. Segmenting of the securities group shares

Segment	**DAX 30, other DAX (MDAX, TECDAX, SDAX)**			**EUROSTOXX 50, DJ STOXX 40, NASDAQ 100**			**Other domestic shares**		
Execution venue	Freq.	Rank 1	Rank 2	Freq.	Rank 1	Rank 2	Freq.	Rank 1	Rank 2
Xetra-Best	4	4	0	3	3	0	1	1	0
Xetra	14	8	5	10	5	4	11	6	4
Berlin	2	0	1	2	1	0	2	0	1
Düsseldorf	2	0	1	2	1	0	2	0	1
Frankfurt	4	0	3	4	1	2	4	2	2
Hamburg	2	1	0	2	2	0	2	2	0
Hannover	1	0	0	1	1	0	1	1	0
Munich	2	0	2	2	0	2	2	0	2
Stuttgart	2	0	1	2	1	0	2	0	1
Tradegate	1	1	0	1	1	0	1	1	0
OTC	1	1	0	1	1	0	1	1	0
Domestic floor trading system	3	0	3	4	1	3	4	2	2
Domestic home exchange	2	0	2	2	0	2	1	0	2
Fixed-price business	3	3	0	3	3	0	3	3	0
No details[10]	3	0	3	6	1	5	6	0	6
Total		18	21**		22*/**	18		19*	21**

* The execution venues Hamburg / Hanover were counted twice in Rank 1.

** The execution venues Frankfurt/ Stuttgart/ Düsseldorf/ Berlin were counted four times in Rank 1 or 2.

[9] A segment is divided into three columns: the frequency (freq.) specifies how often a execution venue was named in total. Rank 1 and Rank 2 specify the number of times the execution venue concerned was nominated for this rank. The total of Rank 1 and Rank 2 does not necessarily match the frequency because nominations from Rank 3 and above were taken into account, but for reasons of clarity have not been included in this table. Thus, for example, Xetra for the DAX 30 values is listed eight times in Rank 1 and five times in Rank 2; one further nomination was registered, but with regard to the prioritization this is in one of the lower ranks.

[10] Some best execution policies do not contain specific details for every segment: for example, in the segment „EUROSTOXX 50“ the one nomination for Rank 1 means that one investment firm provided no details about order execution in this segment, but did document a ranking for the other segments (e.g. DAX 30). As regards the 5 nominations for Rank 2: the best execution policies only envisage one specific execution venue being ranked first, but no information is provided about alternative execution venues in the following ranks.

The execution venue mentioned most frequently in all segments is Xetra. In some policies precisely one specific execution venue is prioritized and occupies Rank 1; in these cases no specific nomination for Rank 2 exists - a "domestic floor trading system" or "domestic home exchange" was ranked second. This is why these abstract execution venues are also contained in the table. As execution venues of equal rank are also named twice or four times in the policies, this indifference with regard to selecting a venue results in a value greater than 18 in the bottom line. In three best execution policies order execution as fixed-price business is placed in Rank 1.[11]

It is noticeable that the regional stock exchanges are rarely - or more frequently in the case of foreign or other foreign shares which are listed domestically - placed in Rank 1, otherwise Xetra or Xetra-Best is dominant. Finally, in one policy the Hamburg stock exchange is also listed as the execution venue for other foreign shares which are listed abroad. New trading platforms like Chi-X were not mentioned at all in the policies examined in this study.

7 Summary and Outlook

On the basis of the analysis of the 75 best execution policies of the 100 largest German financial institutions and of the 15 largest online brokers in Germany, this study comes to conclusion that the minimum legal requirements have recognizably been implemented in nearly all policies. However, significant heterogeneity can be recognized between the policies of various investment firms: some best execution policies are extremely comprehensive and describe the procedure selected in great detail, while others are limited to minimum details and are not very meaningful for clients. The details about the execution venues offered are mainly formulated in an abstract manner (e.g. "domestic execution venue according to the evaluation result").

Only in 24 %, i.e. in approximately every fourth policy, the execution venues are named specifically or is their ranking provided. In the existing studies [8] from the time before the MiFID was applicable, the best execution principles are most frequently named as a key differentiator or competitive factor. However, in summarizing it must be noted that the use of the policies as a competitive instrument cannot at present be recognized in a large majority of German financial institutions. It is striking that only one policy applies a dynamic order routing strategy, i.e. the routing of the client order based on real-time market data.

The brief period between the implementation of the European regulations and their inclusion in national law has certainly resulted in a pragmatic implementation in most investment firms. When this analysis is repeated in the future, it will be possible to see whether this is the main reason for the - current - fairly uncompetitive implementation of the topic, and in how far the financial institutions will, on the one hand, recognizably include in their best execution policies adjustments and further data analyses to provide more specific statements on the various execution venues, for example, or on their ranking, and on the other hand also recognizably incorporate the latest developments on the capital markets, e.g. as a result of new MTFs emerging.

[11] In these cases investment firms offer order execution primarily as fixed-price business. However, if fixed-price business does not come about, the order is directed to a concrete execution venue (e.g. Xetra) which occupies Rank 2.

Naturally it must be noted that these changes will possibly also be reflected in the internal measures taken by the institutions, but they will not be listed explicitly in the best execution policies. Nevertheless, from the client's viewpoint it is desirable that these analyses and their concrete results should be communicated to the client in a more transparent form than the best execution policies and that the MiFID can function not just as a regulatory necessity but also as a driver of competition between investment firms and also between the execution venues.

Ende et al assess the economic relevance of Smart Order Routing engines based on a four week data set of EURO STOXX 50 securities consisting of 8 million executed trades with an overall value of € 262 billions [13]. The analysis shows that on a gross basis, i.e. neglecting execution costs, there is a relevant and statistically significant extent of suboptimal order executions where a different execution venue provides a better executable limit enabling for total savings of € 9.5 million. These findings suggest that dynamic order executions are superior to static approaches that have been implemented by the vast majority of investment firms.

An analysis of the actual results of the different policies based on real order executions and a case-study based comparison of the internal order execution arrangements, processes and systems represent relevant future research topics in this field. On the one hand it is intended to examine the execution results provided by the application of static and/ or dynamic best execution policies, on the other hand a focus may be directed towards the influence of new market entrants and their impact on incumbent trading venues. Moreover, the MiFID perception from a client perspective and potential implications in terms of changes in reputation and performance of investment firms can also add supplementary insights to this topic.

References

1. Macey, J., O'Hara, M.: The Law and Economics of Best Execution. Journal of Financial Intermediation 6(3), 188–223 (1996)
2. McCleskey, S.: Achieving Market Integration - Best Execution, Fragmentation and the Free Flow of Capital. Butterworth-Heinemann, Elsevier (2004)
3. Roll, R.: A simple implicit measure of the bid-ask spread in an efficient market. Journal of Finance 39(4), 1127–1139 (1984)
4. Stoll, H.R.: Inferring the components of the bid-ask spread: theory and empirical tests. Journal of Finance 44(1), 115–134 (1989)
5. Huang, R.D., Stoll, H.R.: Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics 41(3), 313–357 (1996)
6. De Jong, F., Nijman, T., Roell, A.: A comparison of the cost of trading French shares on the Paris Bourse and on Seaq International. European Economic Revue 39(7), 1277–1301 (1993)
7. Bacidore, J., Ross, K., Sofianos, G.: Quantifying best execution at the NYSE, Working paper, New York Stock Exchange (1999)
8. Gomber, P., Chlistalla, M., Gsell, M., Pujol, G., Steenbergen, J.: Umsetzung der MiFID in Deutschland - Empirische Studien zu Status Quo und Entwicklung der MiFID-Readiness der deutschen Finanzindustrie, Books on Demand (2007)

9. Gomber, P., Chlistalla, M.: Implementing MiFID by European execution venues – Between threat and opportunity. Journal of Trading 3(2), 18–28 (Spring, 2008)
10. Gomber, P., Pujol, G.: Best execution in electronic banking and brokerage: an analysis of business and technical requirements. International Journal of Electronic Banking 1(1), 1–15 (2008)
11. European Commission: Commission Directive 2006/73/EC of 10 August 2006 implementing Directive 2004/39/EC of the European Parliament and of the Council as regards organisational requirements and operating conditions for investment firms and defined terms for the purposes of that Directive (Text with EEA relevance), Official Journal of the European Union, L 241, 2.9.2006, pp. 26–58 (2006)
12. CESR: Best Execution under MiFID, Questions & Answers, Ref CEST/07-320 (2007)
13. Ende, B., Gomber, P., Lutat, M.: Smart Order Routing technology in the new European equity trading landscape, Working paper, Chair of e-Finance, University of Frankfurt (2008)
14. Karsch, W.: Top 100 der deutschen Kreditwirtschaft: Auf Wachstumskurs, Die Bank; 2007, Heft 8, S. 34-37 (2007)
15. Kundisch, D., Holtmann, C.: Competition of Retail Trading Venues – Onlinebrokerage and Security Markets in Germany. In: Schlottmann, F., Seese, D., Weinhardt, C. (Hrsg.) Handbook on Information Technology in Finance, pp. 171–192. Springer, Berlin (2008)

The EU Insurance Mediation Directive – Bureaucracy or Opportunity?

Kathrin Braunwarth[1], Hans Ulrich Buhl[1], Marcus Kaiser[1], Alexander Krammer[1], Maximilian Röglinger[1], and Alexander Wehrmann[2]

[1] University of Augsburg, Department of Information Systems Engineering & Financial Management, Universitätsstraße 16, 86159 Augsburg
{kathrin.braunwarth,hans-ulrich.buhl, marcus.kaiser,alexander.krammer, maximilian.roeglinger}@wiwi.uni-augsburg.de
[2] Senacor Technologies AG, Vordere Cramergasse 11, 90478 Nuremberg
alexander.wehrmann@senacor.com

Abstract. The insurance mediation directive of the EU contributes to the recent regulation of financial services markets in order to improve customer protection. Many financial services providers as well as insurance intermediaries fear expensive documentation overhead. In this paper, we argue that the documentation requirements offer a variety of chances. As neither the directive nor national law explicitly specify which customer data has to be collected, we analyse the propositions of respective associations. Moreover, we reason that overall data quality, i.e. completeness, correctness, currency, and consistency of customer data, will probably be influenced in a positive way. With this knowledge in mind, we finally present a set of scenarios from the field of customer relationship management such as advisory process and campaign management that undeniably benefit from a better documentation of customer data.

Keywords: Regulation, insurance mediation, data quality, customer relationship management.

1 Introduction

In recent years, the EU commission began regulating the financial services markets, in order to improve consumer protection. Both the "insurance mediation directive" (IMD, [1]) and the "markets in financial instruments directive" (MiFID, [2]) are designed to support this objective.

While the EU member nations had until November 2007 to convert the MiFID directive into national law, the deadline for the IMD implementation expired in January 2005. The German transposition took effect in May 2007 and affects approx. 500,000 consultants [3]. In the meantime, most of the German financial services providers (FSPs) have introduced procedures that prepare them to fulfill the documentation requirements, but they did not adjust their business processes and IT Systems to capitalize on the opportunities presented by the IMD.

This article explains why it is short-sighted to solely focus on fulfilling the minimum IMD documentation requirements because of the inherent benefits for the FSP

D. Kundisch et al. (Eds.): FinanceCom 2008, LNBIP 23, pp. 145–161, 2009.

(e. g. better quality of financial advisory [4]). Our premise is that – besides others – the IMD comes along with a higher customer data quality (DQ) and therefore provides the FSP with increased value. Both MiFID and IMD present clear business opportunities for financial FSPs, but due to the still outstanding implementation of MiFID and the limited space we further on focus on IMD.

The paper is organized as follows: section two presents the proposed German IMD legislation. Next, we analyze several interpretations of the IMD documentation requirements in order to identify exactly which data FSP are required to collect from the customer. Section three deals with DQ, presents metrics to quantify four dimensions of DQ and discusses the general impact that the IMD has on customer DQ. Subsequently, we illustrate the economic effects of the higher DQ by means of two case studies. Section four summarizes the results and provides further aspects of the topic.

2 Documentation Requirements

In order to build a fundamental understanding, this section presents the salient topics covered in the IMD. We focus on Germany because it constitutes the largest market for insurance services, and the transformation process, converting the directive into national law, will most likely be very similar for other European countries. Subsequently, we analyze which customer data has to be collected in order to comply with legal requirements.

2.1 The Insurance Mediation Directive and Its Transposition into German Law

The Directive 2002/92/EC of the European Parliament and the European Council of 9 December 2002 on insurance mediation was adopted in January 2003. In particular it aims at ensuring a higher level of consumer protection by prescribing professional reliability to be verified by a registration authority. Furthermore, it specifies information and documentation requirements for insurance intermediaries. That is, "insurance […] intermediaries should be registered with [a] competent authority" [1] ensuring that intermediaries "meet strict professional requirements in relation to their competence, good repute, professional indemnity cover and financial capacity" [1]. The IMD also specifies "the obligations which insurance intermediaries should have in providing information to customers" [1]. Additionally, "prior to the conclusion of any specific contract, the insurance intermediary shall at least specify, in particular on the basis of information provided by the customer, the demands and the needs of that customer as well as the underlying reasons for any advice given to the customer on a given insurance product".

As its name implies, the IMD applies to insurance intermediaries. It defines insurance mediation as "activities of introducing, proposing or carrying out other work preparatory to the conclusion of contracts of insurance, or of concluding such contracts, or of assisting in the administration and performance of such contracts, in particular in the event of a claim" [1]. Hence, an insurance intermediary is "any natural or legal person who, for remuneration, takes up or pursues insurance mediation" [1]. Moreover, a tied insurance intermediary is "any person who carries on the activity of insurance intermediation for and on behalf of one or more insurance undertakings"

[1]. As this distinction is only relevant for registration issues and does not concern documentation requirements, we refer to the notion of insurance intermediaries throughout the remainder of this paper. As the concentration process within the financial services industry leads to more and more banks offering insurance products and vice versa, the IMD applies not only to genuine insurance companies, but also to a wide range of other FSP.

Originally, the Member States should have had transposed the IMD into national law by January 2005. Among others, France, Great Britain, Italy, Spain, and Germany failed to meet the deadline [5]. In Germany, the advanced re-election in 2006 caused further delay. In June 2006, the Federal Ministry of Economics and Technology proposed a draft law named "Gesetz zur Neuregelung des Versicherungsvermittlerrechts" [6] that was adopted by the Bundestag in October 2006. The justification of this document shows the guideline of the IMD: "The Directive aims at consumer protection […]. Customer interests shall be protected by an obligation of registration and a standardization of information and documentation requirements of the intermediary." ([6], translated by the authors). As we intend to elaborate the chances resulting from the IMD in the field of customer relationship management, we particularly focus on the documentation requirements described above. This is rooted in that insurance intermediaries will benefit most from obtaining additional customer data of high quality.

In German law, the documentation requirements are specified in §42c of the "Gesetz über den Versicherungsvertrag" (VVG): „The insurance intermediary must interview the insuree according to the complexity of judging the insurance offered or the insuree and the insuree's individual situation. Taking into account the appropriate proportion between effort of advice and the premium to be discharged by the insuree, the insurance intermediary must give advice to the insuree and justify each council concerning certain insurance. The intermediary must document this with respect to the complexity of the insurance contract according to §42d." ([6], translated by the authors).

According to §42d VVG the documented information has to be provided to the customer "in a clear and comprehensible textual form" ([6], translated by the authors) before a contract is signed.

There have been many reactions on the legislative procedure of this law as well as on the results published so far. On the one hand, parts of the law have been commented by various associations. On the other hand, as neither the IMD nor national law explicitly specify which customer data has to be documented. Various institutions have interpreted the results so far. As a consequence, first drafts of advisory protocols and documentation recommendations have been proposed. Some of them are presented and compared below.

2.2 Proposed Documentation Templates

"The advisory documentation has to contain all information that is surveyed from and all recommendations including the reasons given to the customer." ([7], translated by the authors) To enable and ensure this requirement, different organizations are currently working on interpretations of the IMD in order to suggest documentation templates. The results are advisory protocols and documentation recommendations. Some of them are already integrated in software applications.

Among others, two comparably mature documentation templates are the one of the Arbeitskreis EU-Vermittlerrichtlinie and one called Beratungsprotokoll. We analyzed these with focus on the required data fields. As a more technical counterpart we also describe the de facto standard for personal information management vCard [8] and the XML-based Extensible Customer Information Language (xCIL, [9]). Details of the analysis can be seen in the table in the appendix.

It is reasonable to tell apart three groups of data claimed in the templates, because the documentation templates are structured similarly:

- *Basic customer data fields* store personal customer information such as name, date of birth or number of children. Data fields from the technical standards vCard and xCIL as well as the corresponding fields of the other documentation templates are clustered into this group.
- *Product-specific customer data* fields contain information that is especially required for a certain product category. These fields differ for each product category. Additionally, not every field is required for each product category, e.g. living space is required for a householders insurance but not for a life assurance. Both the documentation template of the Arbeitskreis EU-Vermittlerrichtlinie and the Beratungsprotokoll provide different sheets for each product category. Therefore, we clustered these data fields into the group product-specific customer data.
- *Process-specific data* describes the advisory process, such as date of advisory, topic of advisory or advisor's name. In the documentation templates developed with respect to the IMD, these fields can usually be found at the end.

The analysis illustrates several problems regarding the comparability of the surveyed data. First of all, fields that are semantically identical differ with regard to name, e.g. birthday vs. date of birth, and type of data, e.g. marital status as a multiple choice vs. text field. Secondly, each template specifies a set of unique data fields which cannot be found in one or more of the other templates. For instance, the customer's nationality is only asked for by the Beratungsprotokoll and xCIL.

With regard to the data groups, the templates differ quite significantly. As far as basic customer data are concerned, the documentation requirement template of the Arbeitskreis EU Vermittlerrichtlinie and the Beratungsprotokoll show many missing fields compared to the technical standards vCard and xCIL. In the field of product-specific data the difference between the templates is mainly a question of granularity. In return, the process-specific data fields are quite homogeneous.

It is remarkable that none of the proposed templates is based on established standards. The vCard standard for example is integrated in the majority of today's E-mail applications, cell phones or PDAs and covers major parts of the data fields for basic customer data. Therefore, the authors suggest considering established technical standards – e.g. the vCard – when setting up new documentation solutions.

After the comparison of data fields in different templates we will have a look at more general aspects of DQ in the next section.

3 Data Quality

This section investigates to what extent the documentation requirements specified in the IMD may improve the overall quality of customer data. This is necessary because

the quality of customer data is often quite poor [10]. Main reason for this is that customer data are often collected incompletely, at irregular intervals and that it is stored in several independent databases, which is a source of inconsistencies. Moreover it becomes "naturally" obsolete over time. For our investigation, we select an appropriate set of DQ dimensions. After a general definition, we analyze how each dimension must be interpreted in our context and how it may be influenced by the documentation requirements.

In the last years, a lot of research has been done in the field of DQ and many approaches concerning the identification and classification of DQ dimensions have been proposed [11]. Some of them generally apply to data, information, and knowledge management systems. Others focus especially on data warehouses, internet-based or cooperative information systems [12].

As a consequence, the selection of an appropriate set of DQ dimensions is not straightforward. In literature there are two different perspectives on the measurement of quality [13]: Quality of Design (QoD) and Quality of Conformance (QoC). QoD denotes the degree of correspondence between the users' expectations and the specification of the information system (IS) (e.g. specified by means of data schemes). In contrast, QoC represents the degree of correspondence between the specification and the existing realisation in the IS (e.g. data schemes vs. set of stored customer data). In the following we focus on QoC as the existing realisation in the IS (stored data values) are of high relevance regarding the documentation requirements imposed by the IMD.

Considering the definition above, QoC is mainly related to data values. According to [14], the DQ dimensions correctness, completeness, consistency and timeliness are most important in this context.

Completeness

Generally defined, completeness means that data attributes must have a value that semantically differs from NULL. In this context, NULL is not a defined value, but a mere wildcard for unknown or non-present values:

Let w_I be an attribute value stored in the IS. Then the metric for completeness on the level of attribute values $Q_{Compl.}(w)$ is defined as follows [15]:

$$Q_{Compl.}(w_I) := \begin{cases} 0 & \text{if } w_I = NULL \text{ or } w_I \text{ is semantically equal to } NULL \\ 1 & \text{else} \end{cases} \quad (1)$$

Considering an advisory process, this definition does not consider whether an attribute actually should have been documented according to a customer's specific advisory situation. For instance, a customer that solely had an advisory interview on indemnity insurances will probably not indicate data that is relevant for the conclusion of a householders' insurance. Nevertheless, the customer's data are complete with respect to the individual advisory situation. As the definition above treats each data attribute value in equal measure, on this level, we can not emphasize particular attributes, e.g. according to their importance within an individual advisory situation. However, this problem is solved on higher levels of data storage, namely the levels of tuples, relations/views and the whole database. We will describe this in the following:

Based on the level of attribute values, [15] define the metric on the level of tuples: T is a tuple with the values $T.A_1, \ldots, T.A_{|A|}$ for the attributes $A_1, \ldots, A_{|A|}$ and $g_i \in [0;1]$ is the relative importance of A_i regarding completeness. Thus we define the metric for completeness on the level of tupels based on the metric on the level of attribute values as a weighted arithmetic mean:

$$Q_{Compl.}(T) = \frac{\sum_{i=1}^{|A|} Q_{Compl.}(T.A_i) g_i}{\sum_{i=1}^{|A|} g_i} \tag{2}$$

This formula provides the possibility to emphasize particular attributes when quantifying DQ. For instance in our example, if we want to quantify the completeness of attributes which are relevant for indemnity insurances, we can (but need not!) put a higher weight on these attributes. Thus, the metric can be adapted to the data requirements imposed by specific product groups. The results of the metric on the level of tuples can then be aggregated to the next higher level, the one of relations/views.

Let R be a non empty relation or a view. Then the completeness of R bases on the arithmetic mean of the completeness of the tupels T_j in R (j = 1, 2, …, $|T|$) and is defined as (cf. [15]):

$$Q_{Compl.}(R) = \frac{\sum_{j=1}^{|T|} Q_{Compl.}(T_j)}{|T|} \tag{3}$$

Considering a relation consisting of several tuples, this formula seems intuitive. More interesting is the fact that the definition explicitly includes views: This means that we can define particular views e.g. for every product group. These views include only or put a higher weight on those attributes relevant for the corresponding product group and these attributes can also be distributed over several relations. Hence, by defining particular views, we can quantify the completeness according to the specific characteristics of the advisory data.

Finally, let D be a data set (e.g. a database) which can be represented as a disjoint decomposition of the relations or views R_k (k = 1, 2, …, $|R|$). I.e. the whole data set can be decomposed into pairwise non-overlapping relations R_k, so that each attribute in the data set is assigned to exactly one of the relations, or formally noted: $D=R_1 \cup R_2 \cup \ldots \cup R_{|R|}$ and $R_i \cap R_j = \varnothing \ \forall i \neq j$. (Note: in cases when a key attribute is part of several relations or views, it has to be weighted with a positive value only once. This avoids a multiple consideration within the metric for completeness and does not prohibit the applicability of the metric). Hence, we define the completeness of a data set D (based on the completeness of the relations R_k (k=1, 2, …, $|R|$)) as follows:

$$Q_{Compl.}(D,R) := \frac{\sum_{k=1}^{|R|} Q_{Compl.}(R_k) g_k}{\sum_{k=1}^{|R|} g_k} \tag{4}$$

The adoption of the IMD will improve the completeness of customer data with high probability because it is of high importance for both insurance intermediaries and customers. From the perspective of an intermediary, completeness of customer

data proves that the intermediary sedulously carried out his duties, a fact that protects him against claims for indemnities. Moreover, intermediaries are now able to document why fractions of customer data are missing, i.e. whether a customer was not able or did not want to provide the information required. In addition, more complete customer data provides also economic benefit, which we will illustrate later. From the perspective of a customer, complete data enables to identify suitable insurance cover and to achieve acceptance of claims for indemnity against intermediaries or insurance companies. Of course some customers are afraid of their personal data being recorded and abused. It is up to the intermediary to establish trust so that these concerns do not prevent the chances mentioned below for the FSP.

Correctness

In concordance with [16] – and their definition of accuracy – [17] define correctness as the closeness between a value w_I and a value w_R, considered as the correct representation of the real world phenomenon that w_I aims to represent. To quantify closeness, we first need an adequate distance measure d. Examples for distances measures normalized to the interval [0;1] are:

- $d_1(w_I, w_R) := \begin{cases} 0 & \text{if } w_I = w_R \\ 1 & \text{else} \end{cases}$, which is independent of the field of application,
- $d_2(w_I, w_R) := \left(\frac{|w_I - w_R|}{\max\{|w_I|, |w_R|\}} \right)^{\alpha}$ with $\alpha \in R^+$ for numeric, metrically scaled attributes and
- n-grams, edit (or Levenshtein) distance or Hamming distance for strings (all normalized to the interval [0; 1]).

Based on such distance functions, [17] define the metric on the level of attribute values as follows:

$$Q_{Corr.}(w_I, w_R) := 1 - d(w_I, w_R)$$

In analogy to the metric for completeness, the results can be aggregated to the levels of tuples, relations/views and the database.

This definition of correctness directly applies to our context because it concerns basic customer data, product-specific customer data as well as process-specific data in equal measure.

Analogously to completeness, the adoption of the IMD will probably increase the quality of customer data in terms of correctness. Intermediaries must document their recommendations and corresponding reasons according to the needs of the customer. If it turns out that recommendations and reasons cannot be justified properly, intermediaries lose the cover of their professional indemnity. For customers, it is crucial to provide information to the best of their conscience because otherwise insurance contracts loose validity in many cases.

Currency

Currency (which is also often named timeliness) refers to whether "the recorded value is not out of date […]. A stored value, or any data item, that has become outdated is in error in that it differs from the current (correct) value." [18] Though being closely related to correctness, both dimensions do not mean the same. Currency explicitly

addresses the decay of data with respect to the point in time when it has been acquired or revised.

[17] were the first to propose an approach for quantifying currency which is based on probability theory. That means, their metric on the level of attribute values returns a value which indicates the probability that the considered attribute value still corresponds to its real world counterpart. This probability depends on the distribution of the shelf life of an attribute value. For an exponentially distributed shelf life of a particular attribute value, they define the metric for currency as follows:

The parameter *decline*(*A*) is the decline rate indicating how many values of the attribute considered become out of date in average within one period of time. E. g., a value of *decline*(*A*)=0.2 has to be interpreted as follows: on average 20% of the attribute *A*'s values lose their validity within one period of time. The variable *age*(*w*, *A*) denotes the age of the attribute value *w,* which is computed by means of two factors: the instant when DQ is quantified and the instant of data acquisition. The metric on the level of an attribute value is – under the assumption of an exponentially distributed shelf life of the attribute – then defined as:

$$Q_{Time.}(w, A) := \exp(-decline(A) \cdot age(w, A)) \tag{5}$$

Again, the results can be aggregated to higher levels.

As mentioned above, attributes differ in the shelf life and therefore in the decline rate of their values. In cases, where the shelf lives are not exponentially distributed, other metrics have to be developed. [19] design a procedure consisting of six steps, which allows developing a metric for currency according to the specific characteristics of the considered attribute with respect to its values' shelf life.

Currency does not directly depend on the documentation requirements of the IMD. It rather depends on how frequently the intermediary contacts a customer. If this is quite rarely the case, specific fractions of customer data, e.g. marital status or net income, are not updated sufficiently and will become obsolete sooner or later. However, if the customer frequently contacts the intermediary, customer data are updated more often. Therefore, basic customer data are updated each time, product-specific data and process-specific data on occasion only. For instance, if a couple moves together, new insurances must be contracted and already existing insurances like a householders' insurance must be updated. Assuming that customers call on their intermediaries in regular intervals (et vice versa), currency will be improved by the documentation requirements of the IMD.

Consistency

Consistency requires that two or more values of data attributes are free of internal contradictions. Distinguishing only between either *consistent* or *not consistent,* [20] provide a logical definition of consistency: Let *w* be an attribute value within the information system and $\Re$ a set of consistency rules with $|\Re|$ as the number of set elements that shall be applied to *w*. Each consistency rule $r_s \in \Re$ ($s = 1, 2, \ldots, |\Re|$) returns the value 0, if *w* fulfils the consistency rule, and 1 otherwise:

$$r_s(w) := \begin{cases} 0 & \text{if } w \text{ fulfils the consistency rule } r_s \\ 1 & \text{else} \end{cases} \tag{6}$$

Using (6), the metric for consistency is defined as follows:

$$Q_{Cons.}(w,\Re) := \prod_{s=1}^{|\Re|} \left(1 - r_s(w)\right) \tag{7}$$

The value of the metric is 1, if the attribute value fulfils all consistency rules defined in $\Re$ (i. e. $r_s(w) = 0 \; \forall \; r_s(w) \in \Re$). Otherwise the result is 0, if at least one of the rules specified in $\Re$ is violated. (i. e. $\exists r_s \in \Re : r_s(w) = 1$). Such consistency rules can be deduced from business rules or domain-specific functions, e. g. rules that check the value range of an attribute (e. g. $00600 \leq$ *US zip code*, *US zip code* ≤ 99950, *US zip code* $\in \{0, 1, \ldots, 9\}^5$ or *marital status* $\in$ {"single", "married", "divorced", "widowed"}).

Based on particular assumption concerning the consistency rule set $\Re$, the results can be aggregated to higher levels, too. In contrast to the other dimensions, [20] propose not to weigh particular attributes on the level of tuples.

In contrast to the dimensions presented so far, it is impossible to make any a priori assumptions of whether consistency will be influenced positively or negatively by the IMD. If – in order to provide a correct documentation – the insurance intermediary uses an application system which verifies whether the consistency rules are fulfilled or not, the quality of customer data will be higher in this regard. However, this is not directly influenced by the IMD. The status quo can be characterized as follows: Despite considerable standardization efforts, many FSPs still dispose of a historically grown, heterogeneous landscape of product-centric application systems, i.e. often each application system is especially geared to exactly one financial or insurance product. Besides FSPs, individual intermediaries often employ a mixture of paper-based and application system-based documentation techniques. Beyond, product-specific data are maintained in non-integrated application systems, a fact that complicates the realization of an integrated advisory process. As in many cases no integrated advisory process is implemented, process-specific data are scattered to several application systems. The fact that customers contact the FSP over different channels which are not always integrated as well still worsens these problems. Hence, it is a complex task to implement such consistency rules.

Concluding, it can be stated that the documentation requirements of the IMD will presumably have a positive impact on the DQ dimensions investigated above. That is, the additional work caused by recurrent documentation will lead to a higher level of overall DQ. This, in turn, offers several chances, some of which are illustrated in the next section.

4 Case studies

Up to now, we identified which customer data should be documented according to the interpretations of the IMD. Additionally, we analyzed how the IMD influences the quality of customer data. In this section, we exemplarily sketch two chances that arise from the IMD.

4.1 Advisory Process: Economic Effects of a Higher Completeness of Customer Data

In this subsection we present the results of a project which was conducted at a major German FSP (for details see [15]). The project started with the premise that the IMD comes along with a higher completeness of customer data. The goal was now to analyze whether more complete customer data provides higher economic benefits. Therefore, 75 comparable branches of the FSP were selected: To quantify completeness, the metric outlined above was applied to data which are acquired during an individualized advisory process. The measure for economic success was the goal realization level (GRL), which was computed as the ratio between the realized sales of a branch and the target sales (the latter being equal for all branches). Figure 1 depicts the results of this analysis for a subset of the data.

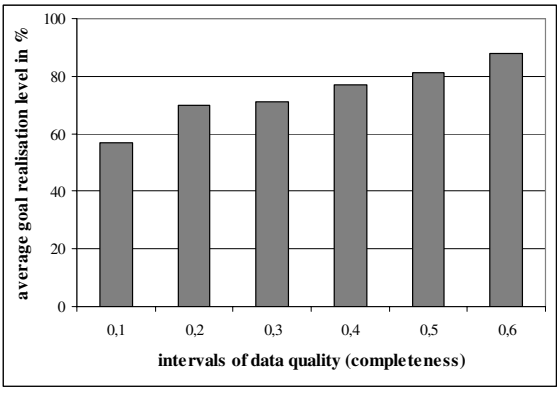


Fig. 1. Average goal realization level depending on the metric for completeness [15]

For instance, all branches with completeness of customer data between 0 and 0.1 achieved in average a GRL of about 58%. We see, that the completeness of customer data seems to be positively correlated with the economic success. Assuming the IMD causing a higher completeness of customer data, it is not only bureaucracy, but will also provide economic advantages. [15] also discuss which subsets of customer data are likely to provide economic benefit and which ones are bureaucracy.

4.2 Campaign Management: Economic Effects of a Higher Currency of Customer Data

Campaign management is defined as "the planning, realization, control, and monitoring of marketing activities aimed at known recipients, who are either current or prospective customers" [21]. As a sub-activity to customer relationship management, its success rate is heavily influenced by DQ [10]. As [19] illustrate, DQ problems arise in the activities of customer selection and customer contacting. The goal of customer selection is to identify target group(s) and to select the corresponding customers by using available data. However, the data stored in the IS are often not up-to-date anymore.

Taking a campaign from a German mobile services provider (MSP), [19] illustrate the economic results of developing a metric for currency according to their procedure. As the campaign addresses students, their metric shall indicate the probability, whether the professional status "student" stored in the customer data base is still up-to-date. Hence, they compute this probability for each of the customers with the professional status "student" based on the age of the attribute value. It turns out that many of these customers are very likely not to study anymore and therefore cannot accept the offer.

Finally, they compare the economic effects of two customer selection methods:

1. Select the top 30% according to sales volume out of all customers with the professional status "student"
2. Select the top 30% according to their individual expected additional profit out of all customers with the professional status "student"

It turned out, that the additional profit was 1.7 times higher when using the second selection method compared to the first one (309,200 € to 178,200 €). This is due to the fact that many of the customers with the professional status "student" were indeed not studying anymore. The second selection procedure - based on a metric for currency, indicating the probability whether a customer is in fact still studying - takes into account this possibility.

5 Conclusion

In this paper, we illustrated that the documentation requirements specified by the IMD do not only cause additional work, but that they also offer a variety of chances. In order to clearly define the notion of customer data, we analyzed advisory protocols and documentation recommendations with respect to data attributes recommended for the documentation of customer-specific, product-specific, and process-specific advisory issues. Furthermore, we identified that the documentation requirements have the potential of improving overall DQ, i.e. completeness, correctness, currency, and consistency of customer data. Finally, we exemplarily illustrated some chances offered by such solutions with scenarios from the field of customer relationship management. We illustrated the IMD being not – as understood by most of FSP at present – just a costly obligation but rather a chance to justify investments to advance DQ. As an adoption of the regulation is mandatory anyway, the investment should be very well considered in order to be able to support strategic decision as well as to improve the day-to-day business. In this way the expenses for the mandatory documentation requirements have not to be seen as sunk but can also generate return in form of revenue resulting from better DQ.

References

1. European Parliament and the Council, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0092:EN:NOT
2. European Parliament and the Council, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:145:0001:0044:EN:PDF

3. Maschmeyer, C.: Vermittlerrichtlinie: Unabhängige als Gewinner. bank und markt 36, 16–18 (2007)
4. Kaiser, M.: Financial Services Advisory. Individualisation and the Role of Customer Data. Dissertation at the University of Augsburg (2008)
5. Freshfields Bruckhaus Deringer, http://www.freshfields.com/practice/disputeresolution/publications/pdfs/13289.pdf
6. Deutscher Bundestag, http://217.160.60.235/BGBL/bgbl1f/bgbl106s3232.pdf
7. Warth, W.P., Beenken, M.: Bankassurance nach EU-Recht: Abschied vom Produktvertrieb. bank und markt 55, 36–38 (2005)
8. InternetMailConsortium, http://www.imc.org/pdi/
9. Organization for the Advancement of Structured Information Standards, http://www.oasis-open.org/committees/ciq/Downloads/xCIL/Versions/xCILv1_0/
10. Ryals, L., Payne, A.: Customer relationship management in financial services: towards information-enabled relationship marketing. Journal of Strategic Marketing 9, 3–27 (2001)
11. Wang, R.Y., Storey, V.C., Firth, C.P.: A Framework for analysis of data quality research. IEEE Transactions on Knowledge and Data Engineering 7, 623–640 (1995)
12. Helfert, M., Heinrich, B.: Analyzing Data Quality Investments in CRM - A model-based approach. In: Proceedings of the 8th International Conference on Information Quality, pp. 80–95. MIT Press, Cambridge (2003)
13. Juran, J.M.: How to think about Quality (1998)
14. Redman, T.C.: Data Quality for the Information Age. Artech House, Boston (1996)
15. Heinrich, B., Kaiser, M., Klier, M.: Does the EU Insurance Mediation Directive help to improve Data Quality? - A metric-based analysis. In: Proceedings of the 16th European Conference on Information Systems (ECIS) (2008)
16. Batini, C., Scannapieco, M.: Data Quality. Concepts, Methodologies and Techniques (Data-Centric Systems and Applications). Springer, Berlin (2006)
17. Heinrich, B., Kaiser, M., Klier, M.: How to measure data quality? – a metric based approach. In: Proceedings of the 28th International Conference on Information Systems (ICIS) (2007)
18. Ballou, D.P., Pazer, H.L.: Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems. Management Science 31, 150–162 (1985)
19. Heinrich, B., Kaiser, M., Klier, M.: A procedure to develop metrics for currency and its application in CRM. In appraisal for: Journal of Data and Information Quality (2008)
20. Heinrich, B., Kaiser, M., Klier, M.: Metrics for measuring data quality – Foundations for an economic data quality management. In: 2nd International Conference on Software and Data Technologies (ICSOFT) (2007)
21. Geib, M., Reichold, A., Kolbe, L., Brenner, W.: Architecture for Customer Relationship Management Approaches in Financial Services. IEEE Computer Society Press, Washington (2005)

Appendix

	Documentation Template "Arbeitskreis EU-Vermittlerrichtlinie"	Beratungsprotokoll	vCard[1]	xCIL
Basic customer data				
Customer identifier	---	---	Unique Identifier (UID)	Customer identifier
Salutation	Salutation	Salutation	Full Name (FN) Name (N) with type values Family Name, Given Name, Additional Names, Honorific Prefixes, Honorific Suffixes Nickname (NICKNAME)	[can be deduced]
Title	Title	---		Name and address details
Name	Name	Name		
First Name	First Name	First Name		
Address details	Address	Address	Address (ADR) and Label (LABEL)	
Birth details	Date of birth	Date of birth	Birthday (BDAY)	Birth details
Age details	[can be deduced]	[can be deduced]	[can be deduced]	Age details
Gender	[can be deduced]	[can be deduced]	[can be deduced]	Gender
Marital Status	Marital Status (multiple choice)	Marital Status (multiple choice)	---	Marital Status
Nationality	---	Nationality (German or other)	---	Nationality
Photo	---	---	Photo (PHOTO)	---
Organisation details	---	---	Organisation (ORG)	Organisation details

[1] Labeling of the particular data field is written in brackets and capitals.

Occupation	Occupation (multiple choice)	Occupation	Job Title (TITLE) Occupation/Role/Business category (ROLE)	Occupation
Qualification details	---	Qualification (multiple choice)	---	Qualification details
Passport details	---	---	---	Passport details
Religion	---	---	---	Religion
Ethnicity	---	---	---	Ethnicity
Telephone details	Telephone (private, official, mobile)	---	Telephone (TEL) with type values home, work, msg (for voice messaging support), pref (means preferred use), voice, fax, cell, video, pager, bbs (for bulletin board system), modem, car, isdn, pcs (indicates personal communication services)	Telephone details
Cellular Phone details		---		Cellular Phone details
Facsimile details	Facsimile (private, official)	---		Facsimile details
Pager		---		Pager
E-mail details	E-mail (private, official)	---	E-mail (EMAIL); types (TYPE): internet, pref, x400 E-Mail Programme (MAILER)	E-mail details
URL	Internet address	---	URL	---
Availability	Availability	---	[can be deduced]	---
Legal representative	Legal representative	---	---	---
Time Zone	---	---	Time Zone (TZ)	---
Global Position	---	---	Global Position (GEO);	---
Living situation	Living situation (multiple choice)	---	---	---
Account details	---	---	---	Account details

Identification card details	---	---	---	Identification card details
Person Identification Number details	---	---	---	Person Identification Number details
Vehicle Information details	---	---	---	Vehicle Information details
Tax number details	---	---	---	Tax number details
Spouse details	Own fields for spouse	Own fields for spouse	---	Spouse details
Children details	Name, First Name, Birthday, living (not) with me, articled (for 4 children), further field family planning	Name, First Name, Birthday, Note (for 2 children)	---	Children details
Note	---	Additions	Note (NOTE)	---
Category	---	---	Category (CATEGORIES)	---
Generated with Product	---	---	Product generated this vCard (PRODID)	---
Sort String	---	---	Sort String (SORT-STRING)	---
Sound	---	---	Sound (SOUND)	---
Version	---	---	Version of vCard specification (VERSION)	---
Access classification	---	---	Access classification (CLASS)	---
Public Key	---	---	Public Key (KEY, ENCODING)	---
Product-specific customer data: [old age provision]				
Income details	Income in the last 3 years Average monthly income after tax Monthly disposable income	Finances (Net income, liquid assets, funds, liabilities)	---	Income details

Existing contract	Existing contracts	Existing contract (cancelled or not)	---	---
Insurer		Insurer	---	---
Identifier		Insurance number	---	---
Pension needed	Monthly Pension needed	Pension needed	---	---
Expected pension	Expected benefits from... Until age... About... One-time sum... Monthly sum... Comment...	Expected provision	---	---
Pension shortfall	Monthly Pension shortfall in	Pension shortfall	---	---
Required Sum	---	Required Sum	---	---
Demand analysis	---	Demand analysis	---	---
Type of insurance	---	Type of insurance	---	---
Additions	---	Additions	---	---
Monthly pension	---	Monthly pension	---	---
Sum in case of survival	---	Sum in case of survival	---	---
Expiration of the contract	---	Expiration of the contract	---	---
Increasing benefits	---	Increasing benefits	---	---
Contribution per year	---	Contribution per year	---	---
Additions	---	additions	---	---
Advisor's recommendation	---	Recommendation (take out/ leave/ change/cancel) Market Basis	---	---

		Reason Changes		
Customer's decision	---	Decision (take out/ leave/ change/cancel) Motives Changes	---	---
Process-specific data				
Copy of advisory protocol given to the customer?	---	Copy of advisory protocol given to the customer?	---	---
Topic of Advisory	[can be deduced]	As multiple choice	---	---
Customer, Advisor and other Attendees	Customer, Advisor and other Attendees	Customer, Partner, Mediator, other Attendees	---	---
Date, time and Place of advisory	Date and Place	Date, Time, Length and Place of advisory	---	---
Customer's signature	Customer's signature	Customer's signature	---	---
Advisor's signature	Advisor's signature	Mediator's signature	---	---

Author Index