Daniel J. Veit Dennis Kundisch
Tim Weitzel Christof Weinhardt
Fethi A. Rabhi Federico Rajola (Eds.)

Enterprise Applications and Services in the Finance Industry

LNBIP 4

3rd International Workshop, FinanceCom 2007
Montreal, Canada, December 2007
Revised Papers

Springer

Lecture Notes
in Business Information Processing 4

Series Editors

Daniel J. Veit Dennis Kundisch
Tim Weitzel Christof Weinhardt
Fethi A. Rabhi Federico Rajola (Eds.)

Enterprise Applications and Services in the Finance Industry

3rd International Workshop, FinanceCom 2007
Montreal, Canada, December 8, 2007
Revised Papers

Volume Editors

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

Dennis Kundisch
University of Freiburg, 79085 Freiburg, Germany
E-mail: dennis.kundisch@vwl.uni-freiburg.de

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

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

Fethi A. Rabhi
The University of New South Wales, Sydney, NSW 2052, Australia
E-mail: f.rabhi@unsw.edu.au

Federico Rajola
Catholic University of Milan, 20123 Milan, Italy
E-mail: federico.rajola@unicatt.it

Library of Congress Control Number: 2008924059

ACM Computing Classification (1998): J.1, H.4, C.3

ISSN 1865-1348
ISBN-10 3-540-78549-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-78549-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12236473 06/3180 5 4 3 2 1 0

Preface

The rapid development and adoption of technology-based information, communication and product delivery systems have facilitated globalization of the financial services industry and removed the barriers between its vertical segments. New business and market models require the adaptation of existing IT infrastructures as well as the development of innovative solutions for both experimental and operational systems.

In this context, several aspects are creating a highly innovative and at the same time competitive setup in the financial industry. In the workshop series on enterprise applications and services in the finance industry (FinanceCom), we try to capture this development. Here, service-oriented computing concepts and architectures play a major role. When applying these novel concepts in the banking and finance sector, the business value of such architectures has to be assessed quantitatively. Also, the heterogeneous and evolving service-oriented modeling methodologies have to be evaluated. Pulling these novel modeling methodologies up to a larger picture, the technology-driven transformation of the financial industry is emphasized, where a highly fragmented and heterogeneous situation is currently developing towards banking value networks.

In this field, financial business standardization and the introduction of standardized service modules play a core role. Applying standardized services certainly allows one to utilize economies of scale in a better way and hence fosters outsourcing and offshoring tendencies in the sector of banking information systems. While addressing this aspect, of course novel technologies that are being developed in other fields also entering the banking and finance domain. Web services and grid computing are two important examples. However, these standards can only be utilized when being applied homogeneously in financial systems. As an example, international standards such as the "markets and financial instruments directive (MiFID)" of the European Union are of key interest.

All of these aspects lead to a deeper integration of information technology into banking and finance. Enablers for a more dynamic and cross-institutional financial field are also innovative design patterns for electronic markets and e-finance business models. While focusing on such specific fields, large rules which set the legal frameworks, such as Basel II or the usage of IFRS (IAS), have a tremendous impact on the domain. Finally, decision support systems in banking as well as implementation exercises and case studies build the foundations of innovative concepts.

Within the Third International Workshop on Enterprise, Applications and Services in the Finance Industry (FinanceCom 2007) many of these aspects were addressed in several contributions. The workshop was hosted with the International Conference on Information Systems (ICIS 2007) in Montreal.

In the first contribution, Peter Gomber and coauthors investigate the issue of volume-weighted average price (VWAP) executions in electronic trading. They consider this concept as a benchmark to measure execution quality. The authors propose a conceptual market model for non-intermediated crossing sessions in a fully electronic environment, which provides anonymity and lifts the constraints of full-day VWAP by introducing the possibility of crossing investors' orders at forward intraday VWAPs.

In the second contribution, Conte and Burghardt elaborate on the MiFID. They evaluate the concepts in MiFID concerning the best execution of client orders and show the potential of the directive in that respect. Extending the directive contents, they present an approach that allows for a dynamic best execution.

Kundisch et al. present a concept about the transfer of portfolio selection theory to customer portfolio management in the case of an e-tailer. In their work, they focus on the assessment of the alternative of investing in existing customers instead of acquiring new ones. From a first evaluation of the application of the portfolio selection theory to this problem, they derive a model for the alignment of customer relationship management activities.

Schaper addresses trends in European cross-border securities settlements. He elaborates on the fact that the settlement phase – often underestimated in trading literature – is still highly fragmented and heterogeneous within the European landscape of financial institutions. An overview about the market, regulation, recent approaches as well as the European code of conduct and TARGET2-Securities is given.

In the following contribution Schulte and others present a framework in which they focus on the general requirements of banks on IT architectures and the service-oriented architecture paradigm. They present the results from an empirical study which quantifies the qualification of SOA for the German banking industry. They identify the heterogeneity of expectations within the banking field.

Marabelli and Rajola draw attention to the interrelation between IT capabilities and organizational change in the banking industry. They investigate how far the Italian financial industry is moving towards compliance with the MiFID directive. In their work, they combine qualitative and quantitative methodologies to investigate the research questions.

Lutz identifies cash tokes for the security assertion markup language (SAML)-based federations. In his work he outlines an approach for digital cash within such setups.

Ureche-Rangau and Carugati in their contribution look at the foreign delisting and domestic stock value. Based on the withdrawal of delisting of their stocks by foreign companies from the Tokyo Stock Exchange, they measure the impact of the delisting decision on the domestic stock price. While highlighting significant clusters, they provide in their content analysis a valuable addition to traditional event study and useful guidance in understanding the delisting phenomena.

In their contribution, Kohlmann and coauthors present instruments for an integrated business network redesign in the financial industry. They explicitly elaborate on factors such as globalization, increased competition and declining customer loyalty as a structural transformation. Within this setup, they identify redesign in the example of the investment process and thus present the fundamentals: a reference network, a reference process and a reference service map for investments.

Berger and Martin explain the adoption of value metrics in retail banks' customer management. In their work, they elaborate on a competitive customer management system which requires the use of adequate metrics for performance controlling and management. They show that the competitive pressure is identified as a dominating driver. Ease of use is also of high relevance and a negative link from firms' profitability to the adoption of the customer lifetime value is revealed.

Sinclair and others provide substantial insight into information risk in financial institutions. In their field study, they establish a research roadmap which allows further investigation of this domain. They discuss lifecycle management and entitlement review processes in the context of large financial institutions and describe both: the results of a field study research in retail as well as in investment banks.

Finally, Weber elaborates on his FinanceCom 2007 keynote paper on technology for trading: what works and what fails. He outlines the most important steps in trading technology of the last 20 years of research and gives a comprehensive and insightful overview on the steps in technology development as well as the adoption of these by the market participants.

Many people were involved in the preparation of this workshop. First of all I would like to thank Jörn Altmann, Dieter Bartmann, Mike Briers, Carl Chen, Giovanni Damiani, Jochen Dzienziol, Lars Friedrich, Heiko Gewald, Peter Gomber, Michael Grebe, Bernd Heinrich, Terry Hendershott, Franz Hollich, Gabriel Jakobson, Steffen Krotsch, Markus Lammers, Dirk Neumann, Mihir Parikh, Hans Gert Penzel, Omer Rana, Marcus-Julian Rumpf, Stefan Sackmann, Oliver Schein, Gerhard Schwabe, Robert Schwartz, Andrew Schwarz, Matthias Tomann, Pascal van Eck, Heinz-Theo Wagner, Bruce Weber and Henning Weltzien for preparing more than 70 reviews within a tight time schedule. Based on these reviews, 11 full contributions were selected for publication (as revised versions) in this volume, which is equivalent to an acceptance rate of 31%. Concerning the workshop at ICIS I would like to thank France Belanger for his support in hosting the event as well as Lisa Rucker for her outstanding management of the event. For the substantial support in preparing the material and running the pre-workshop processes I would like to thank Nils Parasie.

Also, I would like to thank Ralf Gerstner and Christine Guenther from Springer for their excellent support in producing this proceedings volume. Last but not least I would like to thank the Organizing Committee Dennis Kundisch, Tim Weitzel, Christof Weinhardt, Fethi Rabhi and Federico Rajola for the fruitful cooperation.

Finally, I would like to wish all the best for the follow-up FinanceCom Workshops in the coming years.

January 2008 Daniel J. Veit

Organization

FinanceCom 2007 was organized by the Dieter Schwarz Chair of Business Administration and Information Systems, Business School, University of Mannheim, Germany and was technically co-sponsored by the IEEE Computing Society. The workshop was carried out at the International Conference of Information Systems 2007 (ICIS) on December 8, 2007, in Montreal, Canada.

Organizing Committee

Chair	Daniel J. Veit (University of Mannheim, Germany)
Program Chairs	Dennis Kundisch (University of Freiburg, Germany) Tim Weitzel (University of Bamberg, Germany)
Sponsoring Chair	Christof Weinhardt (University of Karlsruhe (TH), Germany)
Technical Sponsoring Chair	Fethi Rabhi (University of New South Wales, Australia)
Publication Chair	Federico Rajola (Catholic University of Milan, Italy)

Program Committee

Jörn Altmann, Intl. University of Bruchsal, Germany
Dieter Bartmann, University of Regensburg, Germany
Mike Briers, SIRCA, Australia
Carl R. Chen, University of Dayton, USA
Giovanni Damiani, Banca Popolare di Milano, Italy
Jochen Dzienziol, finalix, Switzerland
Lars Friedrich, SEB AG, Germany
Heiko Gewald, IBM Global Business Services, Germany
Peter Gomber, Frankfurt University, Germany
Michael Grebe, Boston Consulting Group, Germany
Bernd Heinrich, University of Augsburg, Germany
Terry Hendershott, University of California at Berkeley, USA
Franz Hollich, SAP Deutschland AG and Co. KG, Germany
Gabriel Jakobson, Altusys Corp., USA
Steffen Krotsch, Dresdner Bank, Germany

Markus Lammers, Commerzbank, Germany
Dirk Neumann, University of Karlsruhe, Germany
Mihir Parikh, University of Central Florida, USA
Hans Gert Penzel, European Central Bank, Germany
Omer Rana, Cardiff University, UK
Marcus-Julian Rumpf, Direktberatung AG, Germany
Stefan Sackmann, University of Freiburg, Germany
Oliver Schein, McKinsey & Company, Inc., Germany
Gerhard Schwabe, University of Zurich, Switzerland
Robert A. Schwartz, Baruch College, New York, USA
Andrew Schwarz, Louisiana State University, USA
Matthias Tomann, Senacor Technologies AG, Germany
Pascal van Eck, University of Twente, The Netherlands
Heinz-Theo Wagner, Sell GmbH, Germany
Bruce W. Weber, London Business School, UK
Henning Weltzien, x-markets GmbH, Germany

Table of Contents

Trading Models and Concepts

Banks, IT-Architectures and Settlement

Business Networks and Payment Systems

Information Risk and Customer Management

Keynote Paper

Flexible VWAP Executions in Electronic Trading

Peter Gomber, Marco Lutat, and Adrian Wranik

Johann Wolfgang Goethe-University Frankfurt, Chair of e-Finance, E-Finance Lab,
Robert-Mayer-Strasse 1,
60054 Frankfurt, Germany
{gomber,lutat,wranik}@wiwi.uni-frankfurt.de

Abstract. For the execution of large equity orders, institutional investors often use the Volume Weighted Average Price (VWAP) as a benchmark to measure execution quality. To achieve this, they have the possibility to either cross their orders in a non-intermediated electronic system or to submit a VWAP agency order to a broker that executes the orders manually. Though more expensive in explicit costs, in particular due to higher flexibility, agency VWAP is still more attractive to investors than VWAP crossings. This work proposes a new electronic crossing model addressing and solving the flexibility restrictions present in today's VWAP crossing.

Keywords: Electronic trading, Crossing, VWAP.

1 Introduction

The size of institutional investors' orders can range up to several percent of the average daily volume. Trading such orders on today's electronic trading venues is subject to explicit as well as implicit transaction costs of trading, e.g. opportunity costs and market impact [1] [2]. The market impact results from the information carried by the orders and from the premium paid for liquidity provision.

In order to manage and minimize these costs, benchmarks are applied to measure execution performance and to be able to compare execution venues and the execution services of brokers. As of today, the most established benchmark is the Volume Weighted Average Price (VWAP) as it is easy to measure, easy to communicate and as it is provided from most information vendors (Reuters, Bloomberg, etc.)[1]. To achieve this benchmark, institutional investors (i) execute large orders without any broker intermediation within an electronic crossing system that imports the full day VWAP from a trading venue (reference market) or (ii) these orders are handed over to a broker that executes the order successively and manually in the markets applying data on market volume distributions.

As shown in [3], the VWAP benchmark can be calculated in various ways with various nuances related to the included transactions, e.g. by excluding own trades, by excluding over-the-counter transactions which have been reported with delay or by excluding transactions on other execution venues. If the transaction volume is concentrated on one dominant market, the excluded few trades would have a low influence on the full day VWAP. However, the influence can rise if the calculation

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 1–14, 2008.

period is shortened. Mathematical effects can also occur, when trades from an intra-day crossing session are included in the calculation influencing the VWAP for overlapping later sessions.

The existing trading model of VWAP crossing (see (i)) separates price and volume negotiations from each other, such that the size of the order has no market impact. The process starts by submitting an order into a closed order book before the reference market opens. This enables investors to search counterparties and seek size on the opposite side of the market. Once there is quantity available at both sides, the orders are crossed within a session before the reference market opens and trades occur. After the reference market closes, the VWAP gets imported and the previous trades are enriched by setting this price. The strategy of trying to cross big orders in the first place before trading them in chunks over time has been shown to be cost effective [4]. However, the long period between execution and price determination is subject to the risk of price movement, as the market can move into an unfavorable direction. Based on transaction costs in trading, the long period of one whole trading day leads to high opportunity risk.

The other option (see (ii)) is an agency VWAP order, where the investor hands the order to a broker for execution bound to VWAP as benchmark. The broker splits the order into smaller chunks and executes them over time or even across multiple execution venues trying to achieve the VWAP benchmark. Although the agency VWAP is subject to transaction costs and broker trading behavior is not without critique [5], it offers flexibility to the investor: First, the investor can specify a time period other than the whole trading day, for example the VWAP for several hours. Additionally, the order can be canceled during execution, for example if important news regarding the traded instrument occurs. Finally, the investor has the chance that his order is executed at a price better than the VWAP.

Given that the flexibility advantages of the agency VWAP trading is attractive to institutional investors (despite high explicit costs, i.e. brokerage fees), a new crossing model has to include at least some comparable value proposition. In this paper, a new electronic VWAP crossing model is developed and presented that offers full flexibility relating to the time period of VWAP executions based on an order book market model.

The remainder of this paper is organized as follows: Section 2 will give an overview of the existent academic literature in the context of VWAP trading and presents two examples of VWAP trading solutions in today's financial industry. In section 3 we will present a forward VWAP trading model offering investors a high grade of flexibility in trading. Finally, section 4 will conclude.

2 Related Works

In the academic literature, work on VWAP trading or crossing in general can be found, whereby most of the literature focuses on the investor's perspective, i.e. optimizing VWAP strategies or reducing transaction costs by the use of crossing networks. The design of mechanisms for VWAP trading from a market operator's perspective, i.e. the derivation of market models to satisfy investors' needs, can not be found in literature yet.

An extensive overview on the topic of VWAP trading from the investor's perspective is given e.g. by [3]. The meaning of transaction costs in general and the problematic of implementing an - in theory - alpha generating portfolio investment strategy is presented for hedge funds in [6].

By means of theoretical modeling a "static optimal execution strategy" of a VWAP trade is derived and proved appropriate with empirical data in [7]. Here, the optimal execution strategy is calculated by an iteration of a single variable optimization. It is shown that optimal execution times lag behind expected market trading volume distribution as price volatility has a positive correlation with market trading volume. Essentially, execution error for a basket trade can be reduced by spreading out execution times according to the correlation of price movement.

Optimal trading strategies with the purpose to minimize expected costs when trading large block orders over a fixed time period are also derived in [8]. In this paper the optimal sequence of trades as a function of market conditions is obtained. An extension to the portfolio case where price impact across stocks can have an effect on a portfolio's total trading cost is provided.

Models for two important aspects of modern financial markets, namely VWAP trading and limit order books are developed in [9] and extended to a study of competitive algorithms in these models and a relation to earlier algorithms for trading.

The costs of trading in crossing networks is highlighted in [10] by a comparison of effective trading costs and the costs of non-trading (when an order or a part of it could not be executed in a crossing network). By means of one institutional investor's data set they provide evidence that the low effective trading costs for crossing networks are offset by the costs of non-trading.

Against the background of best execution in institutional investor trading a discussion on the quality of benchmarks like VWAP or High-Low-Open-Close (HLOC) can be found in [11].

Two trading mechanisms for large-block trades are analyzed in [12], namely the "downstairs" markets, such as NYSE floor, and the "upstairs" markets where counterparties to a trade are actively searched and prices are determined by negotiation in order to avoid adverse pricing by insufficient liquidity in the "downstairs" market. Those negotiations often rely on benchmarks like VWAP.

As stated in [5], the evaluation of a trader's performance by portfolio managers is mainly based on a comparison of the price per share that the trader has reached and the VWAP during a whole trading day. It is empirically shown that this fact gives a trader some incentives which do not comply with the portfolio manager's objectives and thus fortifies the principal agent problem.

As already indicated, related research addresses various topics regarding VWAP and institutional investors' trading needs, but does not deal with a suitable VWAP crossing model itself. Therefore, the goal of this work is to extend existing research by presenting a new flexible model for forward VWAP crossing in a fully electronic environment, utilizing the benefits of crossing [13, 4] and addressing the risk of price changes inherent to full-day crossing. Forward VWAP crossing relates to the fact that investors have a facility to submit their orders before the trading session - from which the VWAP is imported - starts. Consequently, the execution price is first unknown to the investor at the time of order submission and second not yet determined when the actual crossing of orders takes place. Thus, the model is different from "off-hours"

VWAP crossing where investors have the opportunity to cross their orders after a regular trading session (off-hours) in which a daily VWAP or a VWAP for a certain period of time has been determined.

2.1 VWAP Landscape in Today's Financial Industry

Some exchanges and agency only brokers (like Instinet) have already recognized the investors' desire for VWAP crossing or crossing related to benchmarks other than VWAP (e.g. midpoint) and offer respective trading models to their clients. In general, public information on VWAP crossing in the industry is spare and hard to find. This subsection is to provide the reader with two examples of the industry's solutions on crossing sessions:

New York Stock Exchange (NYSE) Crossing Sessions
The NYSE provides its clients with four different crossing sessions [14]. Crossing Session I enables member firms to submit their single stock orders into the SuperDot system for 45 minutes starting at 4.15 p.m. Orders will be executed at the end of the session at the NYSE closing price determined during the regular trading session which ends at 4 p.m.

Crossing Session II is restricted to the trading of baskets of at least 15 securities and a minimum value of $1 million. This session is open to NYSE member firms for 2 ½ hours starting at 4 p.m.

In Crossing Session III, NYSE offers its members a facility to execute "guaranteed price" trades for their customers whereby prices may lie outside the price range of the regular NYSE trading session on that day. Again, this session is open for order submission for 2 ½ hours starting at 4 p.m.

Crossing Session IV may be used by members for full-day VWAP orders or trades designated as VWAP for a specific period of time. The crossing session for VWAP trades follows at the end of the regular NYSE trading session and lasts for 2 ½ hours.

As one can see from the descriptions above, all crossing sessions provided by NYSE can be regarded as off-hours trading facilities rather than forward VWAP crossing as associated trade prices are already known before an order is submitted to the system.

Instinet Crossing
Instinet provides its customers with three different forms of crossing with each of them related to a different crossing benchmark and crossing time [15].

First, Instinet customers are offered the opportunity of crossing their orders at the (future) full-day VWAP twice a trading day, where crossing takes place in two separate sessions at 8.30 a.m. and 9.15 a.m. Orders may be submitted from 8.00 a.m. through the beginning of those sessions [16].

Alternatively, investors may cross their orders at seven different times during a trading day ranging from 9.50 a.m. to 3.50 p.m. Those crossings will take place at the security's market mid-spread prevailing in the reference market at the respective point in time.

One last option of crossing investors' orders is presented by Instinet's Last Daily Cross, where orders are matched at the primary exchange closing price at 6.30 p.m. Orders specified for Last Daily Cross may be submitted from 3.30 p.m. through 6.30

p.m. [16]. In any of the three cases described above, orders that do not find a match in a crossing session are returned to the client unexecuted.

The existing approaches are either providing ex post VWAPs where investors have the high risk that – if they don't get an execution – they have no other alternative to achieve the benchmark or are providing full day VWAPs only where the flexibility is significantly lower than in the alternative of an (expansive) broker execution.

3 Flexible VWAP Trading – The Market Design

Based on the discussion in the previous section, this section presents a new model for crossing orders based on the VWAP. As transactions resulting within a crossing system are excluded from the price finding mechanisms implemented in the respective reference market the VWAP is not influenced by them.

3.1 General Market Model Characteristics

This subsection provides an overview on the key market model parameters and refers to some details that will be further described in subsections 3.2 to 3.5: The model is investor order driven, that is, neither external market makers nor the provider of the platform provide liquidity or act as counterparty. Such a model is in-line and shares the advantages of other crossing platforms as well as order driven trading venues.

The overall trading period is bound to the opening times of the reference market with the first crossing at the opening and the last crossing before the closing of the reference market.

The key idea of the proposed model is that crossings are triggered based on the start times (and end times) of the VWAP periods submitted as order parameters from investors (see subsection 3.2). This provides full flexibility concerning the time windows in which the VWAP is determined and thereby differs from existing VWAP crossings. In the model, the submitted start times and end times for the VWAP calculations by the investors are shown in an open order book (see figure 1). Thereby, other institutional investors can react to the order submissions and liquidity can concentrate at specific time windows (see subsection 3.3), i.e. investors will try to join time periods already present rather than specify new ones.

Besides the desired start and end times, the order book is closed in a way that neither volume nor market side information is shown in order to prevent market impact. Trading is anonymous, so investors do not know each other before the trade.

A generic restriction for the order size is not necessary; however, a size per period restriction can be included, e.g. based on a fraction of average daily volume.

Trades are finally and fully confirmed at time of completion of all necessary parameters, including the trade related price. This means that after a successful cross, the investors get a preliminary execution confirmation including the crossed size (see subsection 3.4) without price information. The final trade confirmation (including the execution price) is sent after the VWAP has been calculated, that is, after the specified period has passed. At this point, the trade price is available and will be reflected in the trade confirmation. In order to protect investors and as additional incentive, safeguards against extreme events and reference market movements are presented (see subsection 3.5).

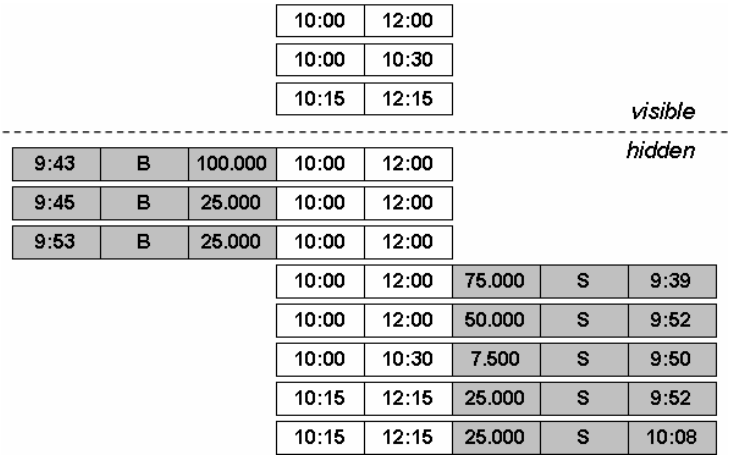


Fig. 1. Shows an example of the visible and hidden parameters of the order book. Only the time intervals (start and end times) are shown to the investors (visible), while quantity, entry time and market side included in the order book are hidden.

3.2 Order Types and Parameters

The key new parameters that are not yet implemented in other crossing systems constitute the start times (t_1) and end times (t_2) for the desired window for VWAP calculation. Based on this, there are several possible types of orders designed for the system which is developed in this paper. The first one is a 'plain vanilla' order named strict. This type carries the minimum parameters needed, namely buy/sell indication, quantity, start time and end time. At the time of order arrival within the system, the order entry time is attached to maintain time priority for the matching mechanism. The start and end time could be alternatively replaced by start time plus period information.

The strict order can be modified by parameters like "all or none" or "minimum execution quantity", further constraining the order. A strict order participates only in crossing auctions with exactly matching parameters (start and end times), and unexecuted quantity is deleted as in state of the art full day crossing systems.

An additional order type, "start fix", allows for partial period matches. Such orders can be executed against orders with the same start time but different end time. In order to enable the investor to achieve the VWAP, the quantity of these orders available for crossing against different periods is adjusted according to the historical volume pattern. Details are depicted in the matching description.

Further order types are "carry forward" orders where unexecuted quantity is forwarded to the next crossing session. Two subtypes are possible, either "end fix" orders which can only be carried to sessions with a later start time but the same end time and "period fix" carried into sessions of the same length. In the first case, quantity is again adjusted to the historic volume pattern, while in the second, the quantity remains unchanged. This "carry forward" order type can be substituted by the investor by canceling the remaining quantity (e.g. a strict order) and submitting a

new one for the later period. As an incentive to use these orders, the time priority is kept. The benefit for the system is an enhanced liquidity, as the trading interest remains within the system. While the carry forward orders are waiting for the start of the next auction, the unexecuted quantity can be canceled by an investor. The period fix version addresses investors who are not bound to the period of VWAP calculation and to timing of the order but lay the focus on the average price. The carry forward orders can optionally have a maximum forward parameter, e.g. the latest start time. Parameters like 'all or none' are also applicable.

3.3 Trading Phases

The phases are order entry phase, crossing auction, VWAP calculation period, price determination and confirmation (see figure 2). There is no fixed order entry phase; however, the first order submitted into the system sets an event driven time window between this first order entry time (t_0) and the specified VWAP calculation start time (t_1). The order book shows the submitted start and end times to enable other traders to join these VWAP calculation windows. All orders willing to interact with the first one or with the same start time (t_1) must be entered before this start time in order to participate within the crossing auction at t_1. Additionally, orders can be canceled before the start of the crossing auction.

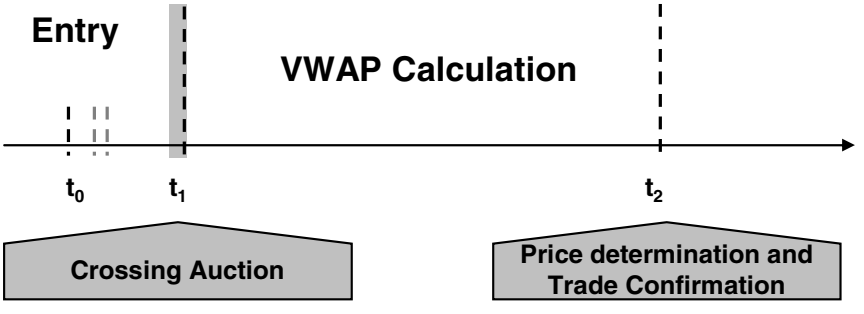


Fig. 2. Trading phases

The order entry time is important for the time priority as one factor for the matching mechanism. Immediately (e.g. 1 min.) before the start time t_1, a crossing auction occurs, matching orders based on the matching rules described below. Orders can not be withdrawn as soon as the auction starts. After the crossing, execution confirmations with the executed quantity are sent to investors. The execution confirmations have no counterparty information to prevent a black board effect where investors could submit only a small quantity to the crossing system and negotiate their real quantity bilaterally with counterparties disseminated through the confirmations. These investors already have shown their acceptance of the time period and would be potential counterparties for bigger trades, which could lead to lower liquidity in the actual system.

At the time the auction ends, the calculation period for the corresponding VWAP begins. The VWAP for the specified period will represent the price for the trades

already crossed in the auction. The executed orders can not be withdrawn from the system by the investors. Additionally, after the crossing auctions end, at the beginning of the VWAP calculation, unexecuted quantity is handled (see flow described in figure 3). Remaining quantity for strict orders is deleted. Remaining quantity or unexecuted orders from carry forward types (either end fix or period fix) are moved to the next suitable crossing session, for which the entry phase is running. If there is no further entry phase, that is, there are no further orders in the system or the orders do not match strict time constraints, the remaining quantity is also deleted.

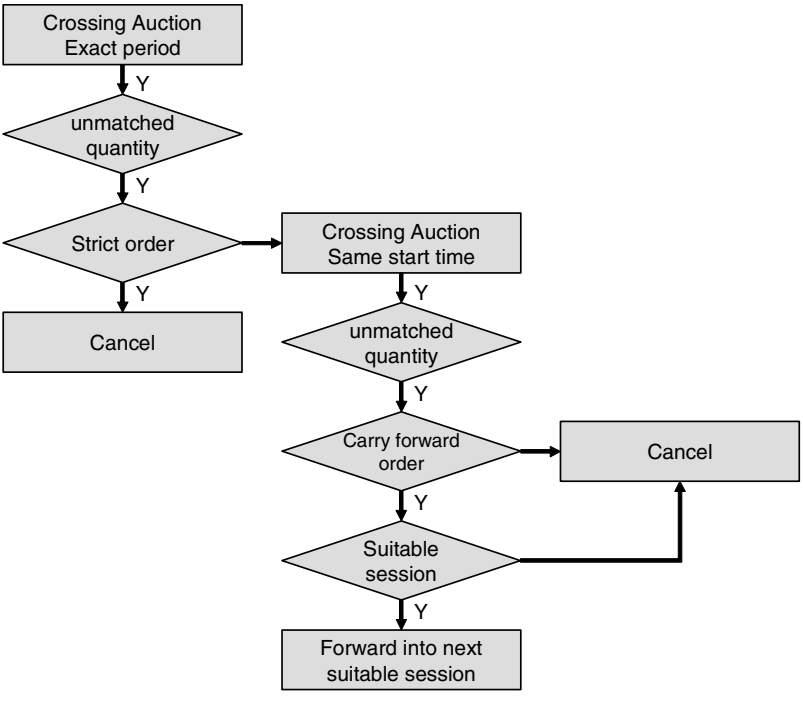


Fig. 3. Processing of orders after a crossing auction

At t_2, the calculation of the VWAP ends and the price of the transaction is determined. This price allows for enrichment of the trade data completing the transaction.

After t_2, the trade confirmations are prepared and sent to the investors, including all data required for post trade processing. This includes data previously included in the execution confirmations as well as price and counterparty information. Together with confirmation of the trades, reports required by regulation are prepared and sent and the transactions are disseminated to the market.

Since investors are free to specify the time periods, these trading phases are present for every single period. This includes overlapping ones, which are relevant within the matching mechanism.

3.4 Matching Rules

Corresponding orders on opposite sides of the market are matched based on time period, entry time and calculation start time priority. Since the VWAP is imported from a reference market and set as the transaction price for all executed orders, there exists no price priority by design.

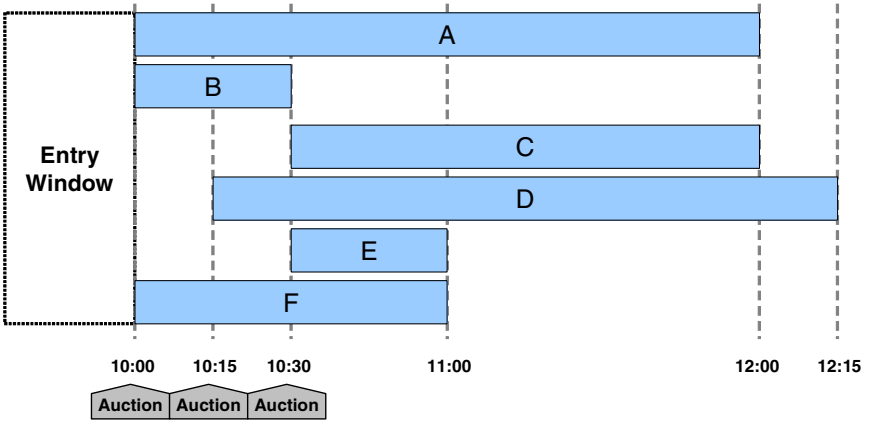


Fig. 4. Various partially or fully overlapping time periods specified within submitted orders. In order to provide examples for the matching rules, the depicted points in time and periods are used below. Additionally, for the examples it is assumed that a crossing auction at 10:00 takes place.

At t_1, orders are matched within a crossing auction. First, all orders with exactly matching time periods - i.e. t_1 (start time) and t_2 (end time, see figure 2) are identical - are matched. In the example in figure 4, this would affect orders marked as A, with specified intervals between 10:00 and 12:00 and the crossing would occur at 10:00 (the corresponding time t_1). Additional auctions would occur at the same time (10:00) for orders depicted as B in figure 4 (for the period of 10:00 – 10:30) as well as for orders depicted as F. Thus, the crossing auctions are triggered by the start time parameter of orders. To this point, the crossing is comparable to existing full day VWAP crossing systems except for the flexible time interval. Orders are matched based on time priority of order entry to reward users which have revealed their preferred time period to the market.

Unexecuted orders as well as remaining quantities from partially executed orders marked for strict matching (order type) are deleted at the end of the auction. Whenever there exists unexecuted quantity enabled for partial period matches (start fix) and which is residual from the first auctions at 10:00, additional auctions including different time periods but same start times are triggered. As an example, orders or remaining quantity of orders marked A and B (F is not present for this example) would be included in an additional crossing auction. In order to enable investors to achieve an overall price as close to the VWAP as possible, remaining quantity of orders marked A in the example is not included in full within this additional auction. Instead, only a fraction of the unexecuted quantity reflecting the average volume traded in the shorter period (here B) in proportion of the average volume traded within the full period can be crossed.

Suppose the volume pattern between 10:00 – 12:00 is linear with time. Here, only one quarter (25%) of the remaining (not matched against other 10:00 – 12:00 orders) quantity of a 10:00 – 12:00 order would be included to match a 10:00 – 10:30 order. By this mechanism, the investor is left 75% of the remaining quantity to either work the order through agency execution or to submit an order with the remaining quantity for another period (10:30 – 12:00). By these mechanisms, the submitters of the A marked orders have shares left allowing to submit orders between 10:30 and 12:00 such that the VWAP can be achieved. The investor is free to submit a carry forward end fix order into the crossing system within the entry time which would automatically perform forwarding and would keep time priority or to cancel the unexecuted quantity and submit additional orders into the crossing system for 10:30 – 12:00 (marked C in the example). To further detail the example above, suppose an investor has submitted an order for 200,000 shares with the interval 10:00 – 12:00 (A in the example). Suppose within the crossing auction for exact time intervals at 10:00, 100,000 shares are executed, leaving another 100,000 as remaining quantity. If the order is suitable for partial match and again assuming a linear volume pattern on the reference market, 25,000 (a quarter of the remaining quantity) shares are included within the additional auction against orders marked as B (10:00 – 10:30). Assuming that these 25,000 shares are matched, the remaining 75,000 shares are either forwarded into a 10:30 – 12:00 interval (carry forward end fix order) or canceled such that the investor can resubmit them on his own (giving up the time priority) or perform an agency execution.

If there are several overlapping intervals with the same start time and unexecuted quantity (e.g. an additional interval 10:00 – 11:00, marked F) would exist, the adjusted remaining quantity traverses the available auctions sequentially, starting with the longest interval in order to maximize matched quantity. Because quantity is adjusted based on historical volume, the longer the period the more shares can be included. Based on the previous example (where 100,000 shares have been already matched), but including the additional interval F, 50,000 shares would be included within the auction for 10:00 – 11:00 (F). If they would be completely matched, no shares are left for the auction between 10:00 and 10:30 because of the historical volume pattern, as the investors need to execute the remaining 50,000 shares between 11:00 and 12:00 in order to match the VWAP. If, say, only 25,000 shares would be matched between 10:00 and 11:00, then 12,500 shares can participate in the auction for the 10:00 – 10:30 interval.

The quantities can be calculated by subtracting the quantity needed for uncovered periods first (100,000 remaining – 50,000 for 11:00 – 12:00), then subtracting matched quantity after every auction (50,000 – 25,000 matched between 10:00 – 11:00) and calculating the fraction based on the historical volume pattern afterwards (25,000 left for 10:00 – 11:00 multiplied by 50% as the fraction of 10:00 – 10:30 in comparison to 10:00 – 11:00 based on the assumed linear volume pattern).

Afterwards, remaining, unmatched quantity is handled as described above. In the example, carry forward end fix orders would be carried to C (auction at 10:30), carry forward period fix orders would be carried to D.

Last possibility would be matching of intra periods, e.g. matching A and E orders against each other based on adjusting of quantity. However, E is involved regularly within the crossing auction at 10:30, where it can be matched against start fix and carry forward end fix orders (which itself can result from unexecuted quantity of A orders) and thus, matching A orders with E orders would reduce liquidity for the

10:30 crossing auction, where it has the possibility being executed in full against various orders. Thus, intra period matching is not available.

The pseudo code for order handling at t_a (marginally before t_1) as well as the used data structures and functions (Table 1) can be seen below.

Table 1. Data structures, parameters, functions and procedures

O_i, EO_j , *UEO,*	Sets of orders, EO for executed and UEO for unexecuted orders. A partial execution can be treated as a combination of an executed and an unexecuted order. The operations $O_i \setminus EO_j$ or UEO $\setminus EO_j$ will care of partial executions.
α_i , ω_i	Parameters containing the start (t_1) and the end times (t_2) of the VWAP calculation of the given order o_i
$P_i \epsilon \{0,1\}$	Selection parameter, where 0 = strict and 1 = start fix order
match{S} $S \rightarrow EO$	Function matching orders in the set *S*, returning executed orders *EO*. The outcome depends on the matching rule e.g. based on the submission time priority. Partial executions lead to an executed and an unexecuted order, $o_i \rightarrow o_{ie}$, o_{iu} where volume(o_i) = volume(o_{ie}) + volume(o_{iu}).
select{S, α, ω, P}	Returns a subset of the given order set *S*, where the parameters *α, ω, P* of the included orders match the specified variables. Unused parameters are specified by a wildcard *
adjust{S, a, w}	Returns a set of adjusted orders with rescaled volume
endtimes{S, α, w}	Returns an ordered set (latest first) of all end times from orders with start time *α* and optional with end times before *w* ($t < w$) in the given set *S*.
carryforward{S, α}	Adjusts carry forward orders with originally earlier start times for auction at time *α* in set *S*, setting new start time and rewriting the volume (carry forward end fix)

```
EO_t := ∅, UEO := Ω
(At the beginning, all orders are unexecuted)

At t_a (marginally before t_1)

      carryforward{UEO, t_1}
            (Forward unexecuted 'carry forward' orders if
            existent from earlier auctions to the current
            auction)

      for t in endtimes{UEO, t_1, *}:
            (Select all end times of orders with start
            time t_1 , e.g. 10:00)

            O_1     := select{UEO, t_1, t, *}
                  (Select orders for auction with exactly
                  corresponding time periods t_1 - t, e.g.
                  10:00 – 12:00)
```

```
      EO_t    := match{O_1}

      UEO     := UEO \ EO_t
      (New unexecuted order set, reduced by the set
      of executed orders)

end for

for t in endtimes{UEO, t_1, *}:
      (Select all remaining available endtimes)

      O_2     := select{UEO, t_1, t, 1}
              (Select previously unexecuted or
              partially executed start fix orders,
              e.g. 10:00 - 12:00)

      for z in endtimes{UEO, t_1, t}:
              (Select all endtimes before t, e.g.
              11:30 if t is 12:00)

              O_3     := adjust{O_2, t_1, z}
                      (Orders originally designated for
                      the longer period, e.g. 10:00 -
                      12:00 need a volume adjustment)

              O_4     := select{UEO, t_1, z, 1}
                      (Select start fix orders for the
                      shorter period, e.g. 10:00 -
                      11:30)

              EO_z    := match{O_3 U O_4}
                      (Match both the adjusted orders
                      and the orders for a shorter VWAP
                      period)

              UEO := UEO \ EO_z
              O_2     := O_2 \ EO_z
                      (Remove executed orders from all
                      unexecuted and from our subset O_2.
                      The remainder of O_2 will be
                      further matched against shorter
                      VWAP period orders, e.g. 10:00 -
                      11:00)

      end for
end for
```

3.5 Safeguarding

The risk for the individual trader/investor concerning the forward crossing mechanisms results from the fact that an order is crossed before a respective price is determined. Because of this setup, investors cannot cancel their orders during unusual market conditions, e.g. periods of high volatility caused e.g. by ad-hoc news for a security. Agency VWAP orders that are executed by brokers have such a cancel option, but by design only for the remaining, unexecuted quantity.

A crossing design has the potential to offer higher investor protection. In a full day VWAP cross, for example, late information disseminated to the market, which will lead e.g. to a higher opening price the next trading day, leaving one side of the executed investor with an unfavorable price. To overcome this risk, ITG offers safeguards for their after hour crossing sessions by monitoring late news and the possibility to cancel a crossing session.

Although the probability of unfavorable news during a crossing session is higher for longer rather than shorter sessions, the impact of positive or negative news on a short-term VWAP might be even greater, leaving an investor with potential price risk and no possibility to react. For this reason, it is necessary to secure also short VWAP sessions with safeguard mechanisms to protect investors' interests. As a consequence, a crossing session in a particular security is canceled and all orders returned to the investors unexecuted, if price jumps in the reference market exceed a certain limit upwards or downwards and thus the volatility in that security becomes too high. That limit can be set as a percentage deviation from the reference market's last price at the beginning of the crossing session and is not disclosed to the investors to avoid manipulation. For the same purpose the percentage should not be set too low as an investor might "cancel" an order submitted to a crossing session by inducing price movements in the reference market and thus stop the entire session otherwise.

4 Conclusions

In order to have their orders executed at the VWAP, investors can submit a VWAP agency order to a broker or make use of a fully electronic crossing facility. Existing VWAP crossing facilities are characterized by low costs, but also inflexibility and price risk as the full-day VWAP is applied for crossing sessions. The crossing systems of two providers, NYSE and Instinet, have been described. Based on the analysis above, a model with flexible VWAP sessions concerning start time, end time and time period appears necessary and in the investors' interest.

We proposed a conceptual market model for non-intermediated crossing sessions in a fully electronic environment, which provides anonymity and lifts the constraints of full-day VWAP by introducing the possibility of crossing investors' orders at forward intraday VWAPs. This model is based on an order book of start and end times for the VWAP periods that are specified by the users of the system, i.e. institutional investors. Thereby it offers – like brokerage agency orders – the opportunity to execute at flexible VWAPs. Investors may specify their orders as "carry forward" to make unexecuted quantity from one crossing session participate in the next session and thus increase execution probability. Besides, orders can be

combined with restrictions like “all or none” or “minimum execution quantity”. Corresponding orders will be matched based on time period, entry time and calculation start time priority. For investor protection the trading model features safeguards which will cancel a session in case of heavy price movements induced by unusual circumstances.

The next step in the project will be a systematic discussion of the model with potential users and – if positively evaluated – a development of a prototype of the proposed model that can be further analyzed e.g. by means of experiments.

Acknowledgments. The authors gratefully acknowledge the support of the E-Finance Lab, Frankfurt for this work.

References

1. Schwartz, R.A., Francioni, R.: Equity Markets in Action: The Fundamentals of Liquidity, Market Structure & Trading. John Wiley & Sons, Hoboken, New Jersey (2004)
2. Massimb, M.N., Phelps, B.D.: Electronic Trading, Market Structure and Liquidity. Financial Analysts Journal 50(1), 39–50 (1994)
3. Madhavan, A.: VWAP Strategies. Transaction Performance Spring, 32–38 (2002)
4. Naes, R., Skjeltorp, J.A.: Equity trading by institutional investors: Evidence on order submission strategies. Journal of Banking and Finance 27, 1779–1817 (2003)
5. Edelen, R.M., Kadlec, G.B.: Agency Costs of Institutional Trading (November 25, 2006)
6. Madhavan, A.: Implementation of Hedge Fund Strategies. Hedge Fund Strategies Fall, 74–80 (2002)
7. Konishi, H.: Optimal slice of a VWAP trade. Journal of Financial Markets 5(2), 197–221 (2002)
8. Bertsimas, D., Lo, A.W.: Optimal control of execution costs. Journal of Financial Markets 1, 1–50 (1998)
9. Kakade, S.M., Kearns, M., Mansour, Y., Ortiz, L.E.: Competitive algorithms for VWAP and limit order trading. In: EC 2004: Proceedings of the 5th ACM conference on Electronic commerce, pp. 189–198. ACM Press, New York (2004)
10. Naes, R., Odegaard, B.A.: Equity Trading by Institutional Investors: To Cross or Not to Cross. SSRN eLibrary (2005)
11. Schwartz, R.A., Wood, R.A.: Best Execution - A candid analysis. The Journal of Portfolio Management Summer, 37–48 (2003)
12. Madhavan, A., Cheng, M.: In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets. The Review of Financial Studies 10(1), 175–203 (1997)
13. Almgren, R., Lorenz, J.: Bayesian Adaptive Trading with a Daily Cycle. Journal of Trading 1(4), 38–46 (2006)
14. New York Stock Exchange: NYSE Crossing Sessions – Extending the Market’s Reach, http://www.nyse.com/pdfs/crossingsessions.pdf (accessed July 14, 2007)
15. Instinet: http://www.instinet.com/pdf/downloads/Instinet_Crossing.pdf (accessed July 14, 2007)
16. Instinet: http://pages.stern.nyu.edu/~jhasbrou/Teaching/2004%20Spring%20T&M/Readings/Instinet%20xing-overview_1203.pdf (accessed July 15, 2007)

A Process Model for Best Execution

Tobias Conte[1] and Matthias Burghardt[2]

[1] Research Center for Information Technologies (FZI),
Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, Germany
conte@fzi.de
[2] Institute of Information Systems and Management (IISM),
University of Karlsruhe, Englerstr. 14, 76131 Karlsruhe, Germany
burghardt@iism.uni-karlsruhe.de

Abstract. The MiFID is regarded as the core of EU financial market regulation. One of its quintessential features is the best execution of client orders. Best execution is virtually impossible to define; it is rather the definition of a suitable process for best execution compliance. To identify an appropriate approach, we analyze MiFID's requirements and develop a best execution process model with emphasis on minimal requirements to ensure MiFID compliance without real-time information. We also present an approach that extends the process model towards dynamic best execution.

Keywords: MiFID, best execution, process model, smart order routing.

1 Introduction

The European Union's "Markets in Financial Instruments Directive" (MiFID) is considered as the core of financial market legislation in recent years. It replaces the "Investment Services Directive" (ISD) and will be effective on November 1st, 2007. MiFID requires that investment firms and market operators adapt to significant changes in the securities trading process. The most significant changes are caused by new requirements for best execution and market transparency. The legislation therefore adapts to recent developments in information and communication technologies that link market places with investment firms worldwide and enables customers to choose from a multitude of execution venues.

Although the MiFID itself was adopted on April 30, 2004, many investment firms haven't completely understood the requirements yet. Instead of taking the chance to differentiate themselves from their competitors with innovative processes and systems, investment firms regard the required changes as a regulatory burden or just another compliance issue.

The central MiFID requirement to provide best execution is a strategic opportunity for investment firms to gain market share and attract retail order flow. The big question is what is understood by "best execution". An explicit definition of best execution can't be found in the MiFID, it is rather a high-level approach that shall give the investment firms a high flexibility for implementation. Furthermore, there is an abundance of legal documents at different levels of granularity making it even

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 15–31, 2008.

more difficult for investment firms to capture and implement the required changes. In this paper we suggest a process model that fulfils these requirements and presents a customizable solution for investment firms.

The contribution of this paper is twofold: First, we analyze the MiFID's best execution requirements and discuss the criteria and factors relevant for investment firms and market operators. Second, we present a process model for best execution that is suited to fulfill the requirements identified in the first part of the paper. Moreover, we identify systematic misplacements of criteria in the legal text which require further interpretation.

The paper is organized as follows: In section 2, we will give a short overview of the MiFID and identify the relevant documents and existing literature. Section 3 is concerned with the analysis of MiFID requirement and the design of the best execution process model. In section 4 we discuss the impacts of the proposed model. In particular, we highlight the differences between the static order routing model and a dynamic best execution model. Section 5 summarizes the findings.

2 The Markets in Financial Instruments Directive

2.1 Overview

The cross-border integration of the European financial markets is one of the most important objectives of the European Union but it has been delayed due to insufficient or inconsequent legislation. The Markets in Financial Instruments Directive [1] is the core element of the European Union's Financial Services Action Plan and includes comprehensive rules targeted at financial services as well as financial markets. These rules shall facilitate the integration of the European financial markets as well as foster investor protection.

MiFID is targeted at investment firms and regulated markets (Article 1)[1]. MiFID also classifies trading venues into three explicit categories: regulated markets, multilateral trading facilities (MTF), and systematic internalizers. A fourth category is implicitly defined in the directive, covering all trading venues which do not belong to one of the three first categories, e.g. for bilateral OTC transactions. These four categories of execution venues differ in several aspects, and the choice of the execution venue influences the execution quality to a high extent. We will discuss these differences in section 3.2 in detail.

Two global goals shall be achieved by the MiFID: First, the development of an efficient and integrated financial market shall be facilitated by the introduction of comprehensive rules for the initiation and execution of transactions in financial instruments. Second, investor protection shall be fostered by the introduction of rules governing the practices of investment firms.

2.2 MiFID's Best Execution Concept

The best execution regulation can be regarded as the core element of legislation intended to achieve investor protection as one of the two global goals of the directive.

[1] References to articles refer to Articles in the Directive [1] unless indicated otherwise.

Article 21 concretizes the obligation for investment firms to "take all reasonable steps to obtain, when executing orders, the best possible result for their clients". In accordance with this obligation, investment firms are required to establish and implement an order execution policy which includes different trading venues and takes into account factors affecting the choice of the execution venue. The order execution policy is intended to select a trading venue that enables firms "to obtain on a consistent basis the best possible result for the execution of client orders".

It becomes obvious that MiFID's best execution is intended to be a static concept, i.e. current market conditions do not have to be considered on an order-by-order basis. The analysis of past performance of trading venues suffices to make statements on the best result on a consistent basis.

2.3 Related Literature

Since the implementation of the MiFID followed a Lamfalussy process [2], there are different levels of legal documents: On the first level, there is the original MiFID text with the framework principles and guidelines defining the scope of the required implementation measures. On level 2, there are two documents: the implementing regulation [3] and the implementing directive [4] which both further specify the MiFID framework [5]. On a third level, consultation papers of the Committee of European Securities Regulators (CESR) include guidelines and valuable comments on the implementation measures. All three types of documents have to be taken into consideration when firms develop MiFID-compliant business processes.

For the U.S. securities market where the term "best execution" has existed for decades, there is a stream of market microstructure literature discussing best execution requirements. Macey and O'Hara [6] examine several market practices that make best execution unwieldy and unworkable as a mandated duty. Hughes [7] discusses the fiduciary duty from a regulatory point of view. Schwartz and Wood [8] discuss best execution from different views and underscore the virtual impossibility of quantifying best execution.

Regarding MiFID adoption in Europe, there are several studies about MiFID readiness of European financial institutions. Fuller [9] conducted one of the first surveys on MiFID readiness of London based investment firms while Gomber and Gsell [10] did the same for German financial institutions. They conclude that the majority of the surveyed investment firms are still insufficiently prepared for the regulatory requirements MiFID imposes.

3 Definition of a Best Execution Process Model

The process model introduced in this section is designed to support investment firms in implementing an appropriate execution policy. In particular we examine the minimal degree of differentiation which is consistent with the requirements regulated by law. Further emphasis is put upon the identification of an approach to select appropriate execution venues consistent with regulatory decrees. The wording of the MiFID suggests a three-step approach.

Initially, investment firms have to decide which weight is being assigned to the relevant factors listed in Article 21(1). To this end, customer needs have to be analyzed with respect to their type and the nature of the order placed (Step 1, cp. section 3.1). Secondly, existing and appropriate trading venues are to be evaluated. A trading venue turns into an execution venue as soon as it is included into a firm's execution policy (Step 2, cp. section 3.2). The third step comprises a procedure which matches the weighted aspects with the selected execution venues. Utilizing the outputs of the first two steps, this execution strategy finally determines which execution venue is able to obtain the best possible result (cp. section 3.3).

3.1 Determination of Relevant Factors and Their Weight

Best Execution Factors. Condition precedent to formulate a process model is to describe the factors referred to in Article 21(1): price, costs, speed, likelihood of execution and settlement, size, nature or any other consideration relevant to the execution of the order, hereafter *best execution factors*.

Since liquidity and transaction costs are considered to be major indicators of market quality, we define the factors in reference to Harris' liquidity dimensions [11]. Furthermore, we concentrate on a disjunctive differentiation of the factors.

Prices of financial instruments consist of bid and ask. By considering the spread (*breadth*) as a benchmark, the factor *price* reflects one major part of implicit transaction costs.

The factor *costs* comprises explicit costs which can be divided into external and internal costs. External direct costs are composed of taxes and fees which accrue for accessing venues and completing transactions there (e.g. commissions). Internal costs have to be considered whenever a firm's execution policy allows for more than one execution venue [12]. In this case investment firms have to incorporate their own commissions into the execution policy.

Speed denotes the time period between placing an executable order and the actual execution by the venue. If an order is not immediately executable, speed refers to the time elapsed between the moment the order is theoretically executable and the actual execution. Therefore, the factor *speed* is related to the liquidity dimension *immediacy*. Furthermore, opportunity costs are being implicitly considered. Speed deficits increase the risk of indicated conditions being already obsolete when an order is actually executed.

Likelihood of execution and settlement addresses the risk of an order not being executed at the indicated price. The more liquid a venue, the more likely an order is executed. Thus, this factor is mainly affected by the depth of the order books.

The factor *size* is confusing since it seems to describe a feature of the customer's order instead of the venue's quality. Therefore, size must be interpreted as the ability of a venue to execute orders without affecting the price ("market impact").The market impact is often responsible for the bulk of transaction costs if orders are large.

The factor *nature of an order* has to be interpreted as well. Its wording is equally misleading since it circumscribes characteristics of the order. In contrast to *size*, there is no reasonable interpretation. The nature of an order is part of the order itself. Evaluating a venue by means of an order's nature is not possible so that this factor takes an exceptional position as an exclusion criterion.

Any other consideration relevant to the execution of the order allows investment firms to define further criteria if they have an impact on best execution (e.g. security features like market supervision or safety measures).

Best Execution Criteria. In the next step, investment firms have to weigh the aforementioned aspects according to their relevance. We examine the degree of detail the best execution criteria must hold below.

Characteristics of the clients are being explicitly defined; the differentiation between retail clients and professional clients complies with MiFID's requirements.

Characteristics of the client orders include any characteristics of an order that determine its execution. Article 21(6) explicitly mentions size and type of order. Type of order refers to the status as a limit order, market order, or other specific type of order. For size evaluation purposes we suggest to introduce intervals whose size each investment firm has to individually settle according to their service and customer portfolio. Some evidence can be found in Article 26 of the implementing regulation [3] where the retail order size is limited to 7.500€. According a firm's customer topology, two intervals I_1[0;7.500€) and I_2[7.500€;∞) might sufficiently reflect customer needs.

The *characteristics of financial instruments* are probably the most important. In order to handle complexity, investment firms should define product groups within classes of instruments that aggregate financial instruments with similar or identical features.

The *characteristics of the execution venues* appear to be a foreign object compared to the other criteria. This criterion has been put at an unfortunate and misleading position, since it actually *is* the task of an execution policy to identify the best possible venues using the weighted factors relevant to best execution. The weighting determines the execution venue; consequentially a venue simply cannot be a best execution criterion. Thus, the legislators' intention must be interpreted. We suppose that the characteristics of an execution venue circumscribe the matching of the weighted factors (step 1) and the evaluated venues (step 2).

Weighting of the Best Execution Factors. Fig. 1 gives an overview of the best execution factors and weighting criteria that have to be included in the execution policy.

The characteristics of the execution venues will be considered in the third step of the process model. The nature of an order as a best execution factor is an exclusion criterion and must be reflected whenever concrete client orders are being executed (cp. section 3.4).

In fact, the best execution factors have a different importance, e.g. in respect to different order sizes. As an example, let's consider the factor *size* which reflects the market impact. The execution of client orders being larger than the size allocated to an indicated quote at a specific venue accounts for a deterioration of the achieved average execution price. Thus, the market impact can account for the largest proportion of transaction costs when it comes to the execution of large-volume orders [13].

Each investment firm is obliged to keep appropriate weightings for all hypothetical shapes of client orders. To this end, possible client orders should be classified into

order groups using the best execution criteria. Subsequently, a specific weighting vector $\vec{w}_h$ for each order group h can be configured dependent on the applicable customer needs.

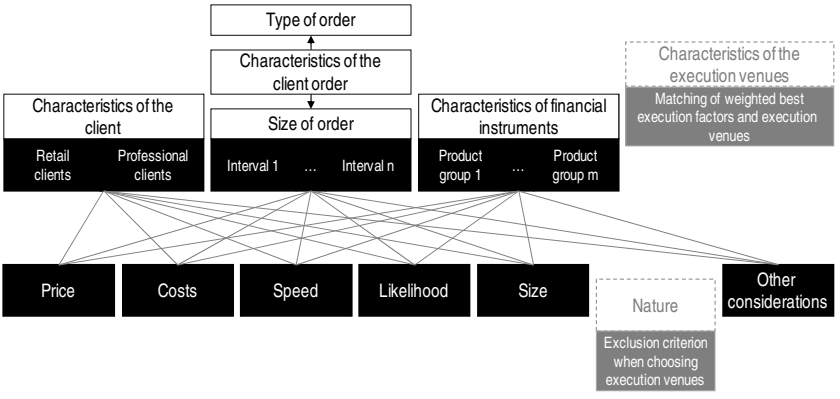


Fig. 1. Relevant best execution factors and criteria

In the course of this procedure some combinations are likely to get identical or very similar weighting vectors. Inter-product group consolidation of these combinations into order clusters would in fact make the weighting of the execution factors more compact. However, the allocation of execution venues would turn much more opaque since venues have to be tested for their best execution potential regarding individual product groups[2]. For the sake of clarity, we refrain from further clustering of order groups although clustering within product groups would not cause any issues.

Table 1. Example of weighting best execution factors

	Characteristics of the client	
Best execution factor	Retail client (A)	Professional client (B)
Price (w_1)	0.4	0.3
Costs (w_2)	0.4	0.3
Speed (w_3)	0.1	0.2
Likelihood (w_4)	0	0.1
Size (w_5)	0	0
Other considerations (w_6)	0.1	0.1

To visualize a specific weighting vector, we have to hold all criteria constant except for one. The following matrix shows an exemplary weighting, considering the product group *limit orders* in *liquid domestic shares* up to a *transaction volume of 7.500 €*. We differentiate retail and professional clients (Table 1). This example does not claim to be accurate. We do not propose a universally valid methodology to

[2] It is unrealistic to assign the same performance to a trading venue for different product groups.

specify the weights since *the* correct weighting shall not exist anyway since every investment firm has its own peculiarities.

For both column (A) and column (B), a separate order group h with weighting vector $\vec{w}_h = (w_h^1, w_h^2, \ldots, w_h^n)$ should be created.

3.2 Selection of Execution Venues

The choice of execution venues is established by Article 21(3) and "shall at least include those venues that enable the investment firm to obtain on a consistent basis the best possible result for the execution of client orders" [1]. It is a crucial determinant of the best execution factors and thus the quality of the execution.

An example may clarify the importance of execution venue choice to retail investors[3]: Retail investors' orders can be routed to electronic exchanges (e.g. Deutsche Börse Xetra), to regional exchanges with intermediaries or specialists providing additional services such as liquidity provision (e.g. Stuttgart Stock Exchange), or to an over-the-counter execution venue. The factor *price* which includes the implicit transaction costs such as the spread may differ across the venues: While the spread in electronic trading systems can potentially be very high, intermediated exchanges guarantee a certain maximum spread to lower implicit transaction costs. At the Stuttgart Stock Exchange, intermediaries offer prices with a spread of zero for the DAX 30 stocks. The factor *cost* varies for the different venues, with the OTC execution venue having the lowest explicit transaction costs whereas the regulated markets such as Xetra and Stuttgart are more expensive in terms of explicit transaction costs. The *speed* of a market order execution also differs between the venues: Xetra may provide the fastest order execution whereas the execution at the Stuttgart Stock Exchange may take several seconds because of the intermediaries' additional services. The *likelihood* of order execution may be very different from one venue to the other depending on whether a limit order book, a market maker system or a hybrid market model is used. Xetra uses a limit order book combined with designated sponsors to guarantee a certain liquidity whereas the Stuttgart Stock Exchange uses a limit order book with a best-price principle that guarantees that orders will be executed as least as good as in the reference markets.

Therefore, investment firms do not have to include each and every venue, they have to consider venues that are *generally able* to provide best execution over a certain time spread, not relative to each individual transaction. The second statutory standard is the linkage of execution venues to classes of financial instruments. If we assume that classes of financial instruments are considered to be a generic term for features of financial instruments, it is in fact possible to structure financial instruments into classes and subsequently refine these classes into product groups. In terms of the actual selection of venues we suggest a twofold approach. First, investment firms should cover venues that are basically worth to be considered. Second, firms should apply a comprehensible evaluation system which separates high-potential venues from negligible venues.

[3] Although the example execution venues are taken from Germany, the example could be extended to include international trading venues as well.

Evaluation of Venues. According to Article 21(3), venues have to be evaluated in respect to the best execution factors. Consideration within the same product group is essential whereas the size of the order should not be reflected in the assessment since it is already accounted for in the best execution factor *size*. Evaluation of trading venues requires data about their execution quality. Availability as well as comparability of such data matter. MiFID's pre-trade transparency provisions ensure the disclosure of data referring to completed transactions. Disclosure in a standardized format is not prescribed [4] so that investment firms have to condense this data themselves or buy it from third party vendors. After identifying sound data[4], service offerings and performance indicators have to be linked to customer needs. To handle this as transparent and verifiable as possible, we introduce a simple rating system.

Rating System. MiFID does not only require investment firms to list all high-potential trading venues in their execution policies, it also requires them to justify the choice of venues. From our point of view, only mathematical approaches are capable of meeting this requirement by formally motivating the firms' decision. The concept of the rating is to sort the pre-selected trading venues by their performance indicators for each best execution factor[5]. Generally there are two kinds of measurement applicable for the rating: ordinal or cardinal measurement.

Ordinal Measurement. Ordinal scales sort venues without quantitatively measuring differences in quality. Let's assume that (with reference to the product group at hand) there are m venues available. Using the quality data, we can sort the venues for each best execution factor in ascending order, assigning the highest score $f(m,r)$ to the venue with the highest rank r. Thus, a specific score $f(m,r)$ refers to the best execution potential of a trading venue regarding a specific product group and a specific best execution factor (see Table 2). Dependent on m and r, we assign a score to each trading venue for each best execution factor referring to a specific product group:

$$f(m,r) = \frac{m - r + 1}{m} \tag{1}$$

The formula is being normalized to the interval $I(0;1]$. It is based on the assumption that a ranking is continued at rank $(r + l)$ after equally ranked venues at rank r. This is consistent with the nature of ordinal rankings and ensures comparability between aspects with equally ranked venues and aspects with unique rankings. If venues exhibit very similar but not identical quality data, it is often reasonable to assign the same rank since minimal variations might not justify different ranks. Importantly, the applied mechanism must be discretionary and documented.

Table 2 shows an example for $m = 4$ trading venues. Let $a_1, \ldots, a_6$ denote the best execution factors in the following sequence: Price, costs, speed, likelihood of execution and settlement, size, and other consideration.

[4] We will not elaborate on suitable data quality. Possible key performance indicators may be historical data on spread, reaction time, and the structure of rates and charges.

[5] If there is more than one performance indicator for a best execution factor, these indicators must again be (internally) weighted in order to determine their relative importance.

Table 2. Ordinal rating of trading venues

Factor	a_1	a_2	a_3	a_4	a_5	a_6
Trading venue α with respect to product group						
Rank r	1	2	2	2	3	1
Score $f(m,r)$	1.0	0.75	0.75	0.75	0.5	1.0
Trading venue β with respect to product group						
Rank r	4	4	1	3	3	1
Score $f(m,r)$	0.25	0.25	1.0	0.5	0.5	1.0
Trading venue γ with respect to product group						
Rank r	3	1	2	4	1	1
Score $f(m,r)$	0.5	1.0	0.75	0.25	1.0	1.0
Trading venue δ with respect to product group						
Rank r	1	3	2	1	2	4
Score $f(m,r)$	1.0	0.5	0.75	1.0	0.75	0.25

Ordinal scales only allow for qualitative comparisons such as "q positions better or worse" or "equal". A trading venue that yields a much better quality than all subsequent venues in respect to a specific factor would be discriminated against. A further disadvantage may arise from a later addition or removal of venues since such actions require readjustment to the new conditions.

Cardinal Measurement. Use of cardinal measurement allows for exact quantification of quality differences. Thus, the actual quality of trading venues can be defined more accurately. Cardinal measurement requires an adequate score interval which is valid for all best execution factors. A feasible solution might be a normalized score interval $S(0;1]$ with fixed increments. Based on this interval, actual quality features have to be mapped to an appropriate score which introduces a high level of complexity into the rating process. On the other hand, a cardinal scale minimizes ongoing correction efforts since the addition or removal of execution venues will only trigger a change if this venue outperforms other venues in one or several factors. Table 3 maps the example from above to a cardinal system, using $S(0;1]$ as the score interval and increments of 0.05 (just considering trading venue α). Concretely, the scores depend on the mapping that has previously been undertaken.

Table 3. Cardinal rating of trading venues

Factor	a_1	a_2	a_3	a_4	a_5	a_6
Trading venue α with respect to product group						
Score	0.85	0.7	0.75	0.7	0.45	0.95

Choice of Execution Venue. After rating the trading venues, either all of the venues or the most attractive ones can be adopted according to the execution policy. The trading venues that become part of the execution policy are called *execution venues.*

3.3 Matching of Weighted Best Execution Factors and Execution Venues

In the third step of the process model we consolidate the weighted best execution factors and order groups (step 1) and the evaluation, rating, and selection of venues

(step 2). Thus, an order group specific value is assigned to each execution venue so that investment firms can predefine a static sequence of venues for every order group. We denote the best possible execution venue as "default execution venue".

First, for every order group investment firms have to determine the associated product group since the weighting vectors $\vec{w}_h$ are being set in respect to order groups while the execution venues are rated according to product groups. Therewith, execution venues and their factor- and product-wise rating for every order group can be predefined for each order group by means of a firm-specific "best execution value" (BEV). The BEV determines the sequence execution venues are being accessed. Defining more than one execution venue per order group ensures MiFID conformity and capacity to act if the default execution venue must be overridden due to partial customer instructions or features that are not supported (see Fig. 2).

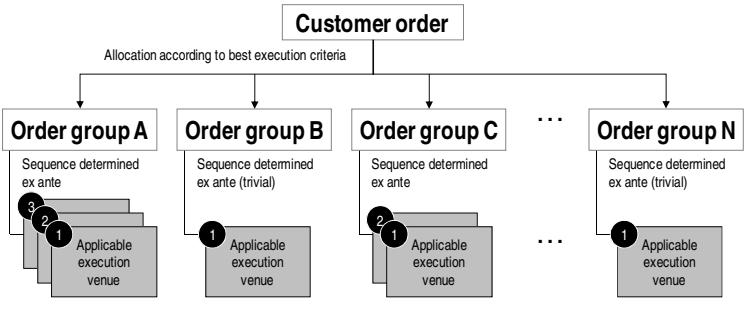


Fig. 2. Ex-ante allocation of execution venues

Let *h* denote the order group at hand. Let $\vec{w}_h = (w_h^1, \ldots, w_h^n)$ denote the weighting vector assigned to *h* where w_h^i is the weighting coefficient for best execution factor *i*. The indices shall be interpreted as in Table 1. $x_{j,k}^i$ shall denominate the rating for venue *j* by means of best execution factor *i* and product group *k*. There are *j*=1 … *m* execution venues that are allocated to *k*. Calculation of best execution value V_j^h for venue *j* results from following equation:

$$V_j^h = \sum_{i=1}^{n} w_h^i x_{j,k}^i \tag{2}$$

The default execution venue j_{max}^h for *h* is the venue having scored best:

$$j_{max}^h = j_j \text{ where } \max_j \left\{ V_j^h \right\} = \max_j \left\{ \sum_{i=1}^{n} w_h^i x_{j,k}^i \right\} \tag{3}$$

The following example clarifies the above-defined approach. Let's assume that an investment firm has conducted step 1 and step 2 for a retail customer order group *A* and applies the ordinal measurement to rate venues. The calculation of the BEV is made according to equation (3).

$$V_\alpha^A = 0.4 \cdot 1.0 + 0.4 \cdot 0.75 + 0.1 \cdot 0.75 + 0 \cdot 0.75 + 0 \cdot 0.5 + 0.1 \cdot 1.0 = 0.875 \tag{4}$$

Analogously, we get $V_\beta^A = 0.4$, $V_\gamma^A = 0.775$ and $V_\delta^A = 0.7$ Execution venue α is the default execution venue since $V_\alpha^A > V_\gamma^A > V_\delta^A > V_\beta^A$.

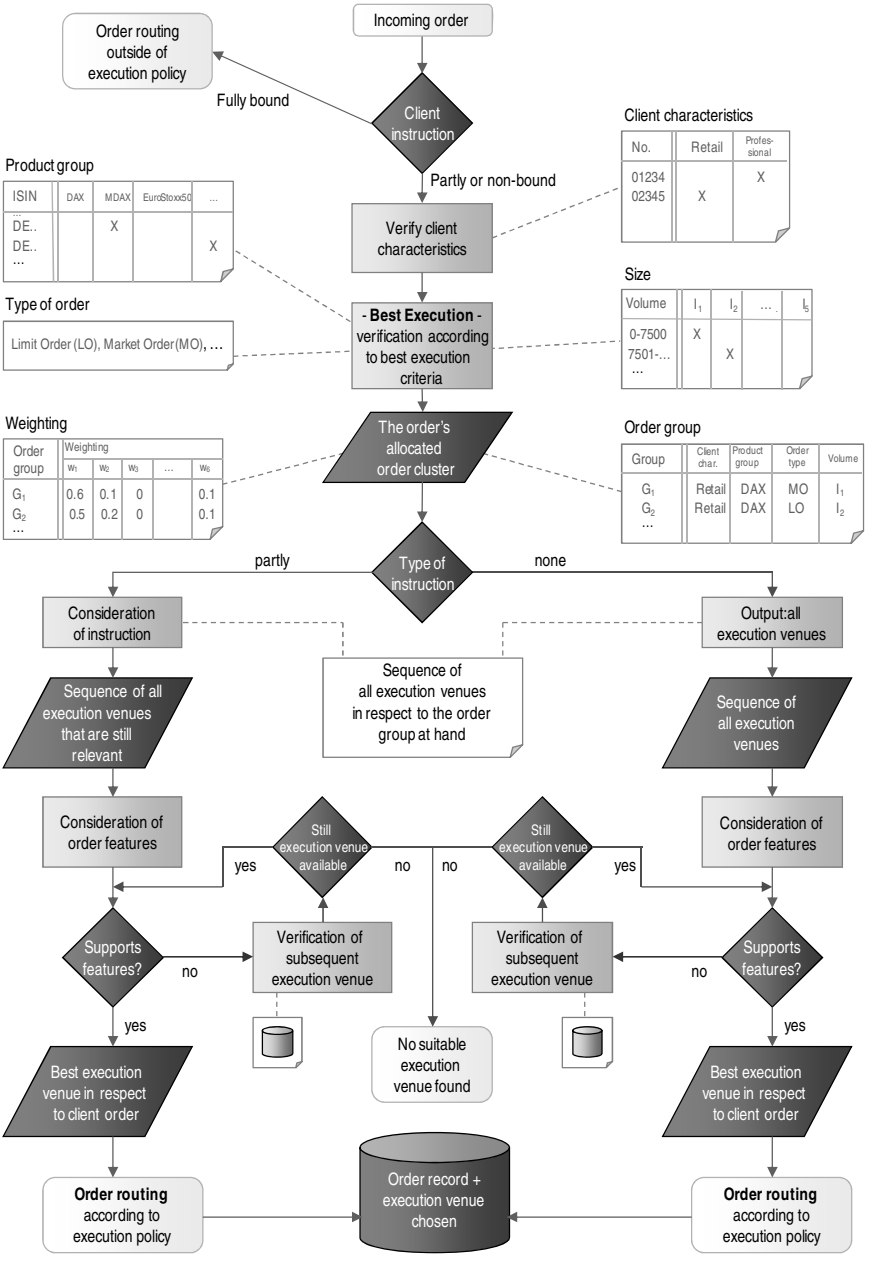


Fig. 3. Process model for executing client orders

3.4 Example for the Execution of a Specific Client Order

After the customer has agreed to the execution policy an investment firm is obliged to handle his or her orders according to that execution policy. First, there should be a

differentiation between orders fully bound to instructions and orders partly or non-bound to instructions. Orders containing explicit instructions can be executed outside of best execution at the venue the customer desires. Second, the order has to be checked for its client characteristics (e.g. via customer ID) to identify orders which best execution does not apply for. Retail and professional client orders have to be processed according to the execution policy. Along with the type of instrument and the order size, the order can be classified into one of the predetermined order groups.

By identifying the applicable order group *h* on the basis of a systematic procession of the weighting criteria, the ex-ante determined sequence of execution venues is known. Subsequently, any part of the client instruction that impairs standard execution can be considered. E.g. explicit exclusion of OTC venues actually disqualifies non-regulated marketplaces which are allocated to the order group at hand. Eventually, before finally executing the order, both features of partly bound and non-bound orders ought to be compared to the service offerings of the applicable execution venues. If the default execution venue does not support the desired features, a renewed verification of the subsequent execution venues is carried out in descendent order until either a suitable execution venue has been identified or all venues have been checked without a match. These subroutines equal an evaluation of the best execution factor "nature" that has earlier been identified as exclusion criterion (see section 3.1 and Fig. 1). Records of all orders that have been processed through the execution policy have to be saved for verification purposes. Fig. 3 exemplarily illustrates the client order handling.

4 Discussion and Outlook

4.1 Discussion of a Minimal Approach to Comply with MiFID's Best Execution

With MiFID's best execution, the European Commission de jure abolishes any concentration rules[6] by explicitly stating that regulated markets and OTC-markets must be considered in a firm's execution policy if capable of providing best execution. However, we might expose this new equality to be of theoretical nature, since de facto continuously liquid market places with high trading volumes profit from static best execution [14]. Temporarily available best conditions offered by alternative marketplaces do not suffice to attract order flow, which will continue to accrue towards established trading venues. These are mostly regulated markets.

From a client point of view, best execution will noticeably increase investor protection. Due to the disclosure of investment firms' execution policies investors will be able to compare investment firms and their interpretations of best execution and choose the service provider whose execution policy is aligned best with own investment objectives. This increase in transparency will diminish information asymmetries existing between investment firms and clients by default.

However, additional value for clients and the capital market as a whole could be created by implementing best execution on an order-by-order basis incorporating current market data. Thus, best execution would turn from an organizational obligation to

[6] MiFID's predecessor ISD did not prohibit national transpositions that oblige investment firms to route client orders to regulated markets.

provide a suitable process to an advanced procedure that significantly converges to *true best execution*. Nevertheless, we assume that the industry will initially focus on implementations with low differentiation and static venue allocation. Therefore, the best execution process model presented in section 3 is supposed to suit a large proportion of investment firms.

4.2 Towards Best Execution Order-by-Order as Optional Strategy

As outlined in the previous section, we suppose that investment firms implement the minimal requirements of best execution. There are basically two reasons for this development: On the one hand, implementation time is very tight so that firms turn their attention to on-time MiFID-compliance rather than to sophisticated order-by-order best execution engines. On the other hand, implementing best execution order-by-order, i.e. enabling a dynamic routing decision for each individual client order, will involve vast investment costs [14]. Yet these investment needs might be higher than the marginal utility generated by such new features. However, after the introduction of MiFID and a first consolidation phase it is in fact possible that investment firms enhance the standards of their execution policies towards dynamic best execution.

The main idea of best execution order-by-order is a process that guarantees best execution for each individual client order by considering current market conditions. The fundamental difference to the process model introduced in section 3 is to be found in step 3: the matching of weighted best execution factors and selected execution venues. Order-by-order routing provides real-time-based routing without ex-ante sequenced venues while the consolidation into order groups is still a valid concept. Dynamic routing decisions are supported by innovative concepts such as smart order routing (SOR). SOR algorithms seek optimal conditions at the time an order is placed to ensure the best possible execution. The definition of "best possible" results from the weighting of the client order at hand.

Fundamental prerequisite for real-time scanning of market conditions is a direct connection to execution venues via real-time protocols for the electronic communication of trade-related messages such as the FIX (Financial Information eXchange) protocol [15]. Market data disclosure is facilitated by MiFID's pre-trade transparency rule. The connection of trading venues using real-time protocols creates a virtual network of liquidity pools which enables investment firms to make centralized routing decisions in a decentralized environment.

However, not all best execution factors are being affected by real-time decisions so that the routing decision still requires a rating of execution venues. The factors price, size and likelihood of execution can be acquired using real-time data. Explicit costs, speed and other considerations like security mechanisms are real-time-invariant, therefore require ex-ante evaluation. This multi-attributive approach complicates the routing algorithm.

In the course of routing an order, real-time data meets real-time-invariant data[7]. This heterogeneous information has to be converted into a standardized format to ensure consistent routing decisions. Fig. 4 recapitulates the requirements and influencing factors of dynamic best execution.

[7] Thus, a static approach as presented in chapter 3 will not be obsolete once an investment firm implements order-by-order best execution since the real-time-invariant aspects still need to be evaluated ex-ante.

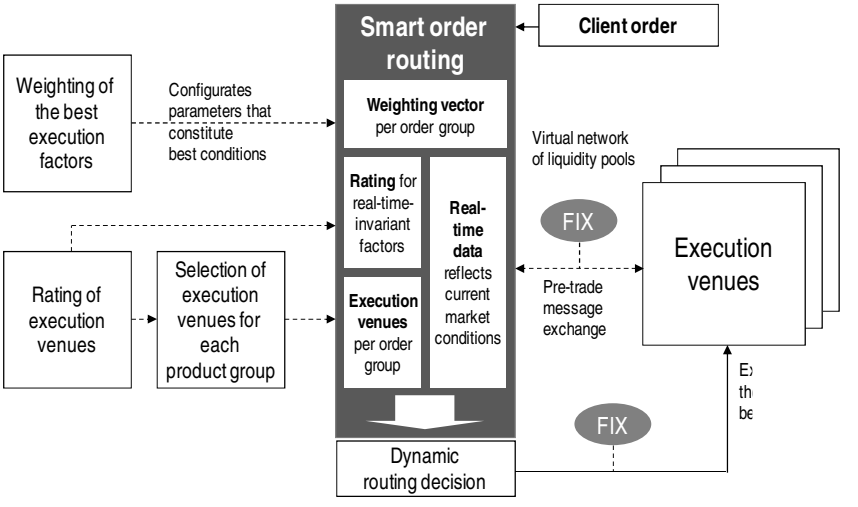


Fig. 4. Best Execution Order-by-Order (SOR)

The bottom line is that MIFID-compliant SOR algorithms must provide more functionality than "ordinary" SOR algorithms fulfilling the requirements of the U.S. National Market System. However, some investment firms perceive implementing order-by-order algorithms as an opportunity to differentiate themselves from competitors opting for static best execution. Although changing over to order-by-order mechanisms might be quite cost-intensive, it could turn out to be an investment for the future. It may very well be the case that third party best execution providers sell the service "best execution order-by-order" to smaller investment firms that do not have the capacity or infrastructure to handle dynamic best execution themselves. Such outsourcing suppliers in turn might be able to offer dynamic best execution comparatively moderate priced exploiting economies of scale due to their high trading volumes.

4.3 MiFID's Best Execution vs. SEC's Reg NMS

The best execution regime imposed by U.S. Securities and Exchange Commission (SEC) is a twofold approach, comprising both static and dynamic elements. On the one hand, brokers are subject to a fiduciary obligation to provide best execution which is similar to the concept pursued by the MiFID. The big difference is its legislation: U.S. best execution is common law only. Principles like multiple factors affecting best execution or the static nature of the process have been highlighted, yet overall obligations for investment firms remain vague due to inconsistent and unofficial statements.

On the other hand, SEC implemented a second (regulatory) principle to ensure best execution on top of the common law fiduciary obligation. This *National Market System* (NMS) is a centralized system which provides a *National Best Bid and Offer* (NBBO) that is displayed on a pre-trade basis for exchange-traded securities. One of the basic features of NMS is the *trade through rule* which obliges trading venues to

route client orders to the venue quoting the NBBO. Thus, there is only one factor that constitutes best execution in NMS – the price.

With Regulation NMS (Reg NMS) [16], the SEC updates the aforesaid trade through rule, henceforth called *order protection rule* (Rule 611). Venues are being categorized in *slow markets* and *fast markets* according to their mechanism of price determination. Rule 611 continues to protect fast markets against trade-throughs while quotations with manual interference (provided by slow markets) are no longer protected. I.e. client orders may be executed at a fast market even though the NBBO is set by a slow market. Thus, *speed* becomes an additional factor for the choice of venue.

Although MiFID's best execution and Reg NMS may have similar objectives, they are based on quite different concepts. While MiFID addresses investment firms, Reg NMS externalizes the duty to provide best execution to trading venues. An order protection or trade through rule is a concept that is not known in European legislation - multilateral trading venues are not involved in best execution provisions. By establishing a virtual national market, Reg NMS is based on centralization whereas MIFID's best execution is built on individual and decentralized execution policies provided by each investment firm. These execution policies incorporate multiple parameters that impact the routing decision. MiFID explicitly rejects a clearly defined benchmark as a safe harbor. By contrast, Rule 611 solely takes account of the factors price and speed, the former being benchmarked by the NBBO. However, neglecting other factors such as price effect robustness or explicit costs is a major simplification which might not be sufficient to effectively increase investor protection. Lastly, MiFID opts for a process which is fitted to static best execution, whereas Reg NMS requires dynamic best execution order-by-order.

If dynamic best execution takes hold of the MiFID, European and U.S. best execution will in fact converge, still being fundamentally different with regards to addressees and the multi-attributive routing decision putting state-of-the-art SOR-algorithms to the test.

5 Conclusions

Best execution is often referred to as MiFID's central element and the catalyst to reach the global goals of the regulation. Macey and O'Hara [6] raised the question whether best execution is a definable goal or a shapeless concept akin to market efficiency. The obligation to execute orders most favorably for the client lies between these extremes.

MiFID's best execution can surely not be defined explicitly. It has a different meaning dependent on different client needs. *The* best execution does not exist; characteristics are much too heterogeneous to be satisfied by a one-size-fits-all definition. Therefore, the framework delivered by the EU does not contain an explicit definition of best execution, yet it requires the definition of a process that is capable of providing a verifiable and repeatable best execution. Our process model introduced in section 3 represents such a procedure.

Step 1 of the process model deals with weighting the best execution factors. According to the weighting criteria, client orders are classified into order groups. Differentiation

between types of financial instruments is done via product groups. Step 2 uses a score-based ranking to mathematically justify a firm's selection of execution venues. Step 3 merges the first two steps such that a product group specific best execution value is assigned to each execution venue. This allocation can be done ex ante before executing any specific client orders.

In chapter 4 we discussed advantages and disadvantages of the proposed static best execution process model and gave an idea of the development of dynamic models that incorporate real-time data. We find that upgrading the best execution policy to support dynamic best execution order-by-order holds differentiation potential which raises new questions on the MiFID implementation measures since they are obviously designed to support static best execution models.

Hence, best execution will remain a hot topic after the implementation deadline in November 2007. Further research should be directed to develop real-time based best execution methods and innovative business models for financial service providers offering best execution services.

References

1. European Commission: Directive 2004/39/EC of the European Parliament and of the Council of April 21, 2004 on markets in financial instruments (2004)
2. Davies, R., Dufour, A., Scott-Quinn, B.: The MiFID: Competition in a new European equity market regulatory structure. In: Ferrarini, G., Wymeersch, E. (eds.) Investor Protection in Europe. Corporate law making, the MiFID and beyond, pp. 163–197. Oxford University Press, Oxford (2006)
3. European Commission: Commission Regulation No. 1287/2006 of (August 10, 2006)
4. European Commission: Commission Directive 2006/73/EC of (August 10, 2006)
5. Ferrarini, G.: Contract Standards and the Markets in Financial Instruments Directive (MiFID): An Assessment of the Lamfalussy Regulatory Architecture. European Review of Contract Law 1(1), 19–43 (2005)
6. Macey, J., O'Hara, M.: The Law and Economics of Best Execution. Journal of Financial Intermediation 6, 188–223 (1997)
7. Hughes, P.: What Is Best Execution? A Primer For Satisfying the Investment Adviser's Fiduciary Duty Under U.S. Securities Regulation. In: Mössle, K. (ed.) Best Execution. Executing Transactions in Securities Markets on behalf of Investors. A collection of essays, pp. 75–82. European Asset Management Association (2002)
8. Schwartz, R.A., Wood, R.: Best Execution: A candid analysis. The Journal of Portfolio Management 29(4), 3–48 (2003)
9. Fuller, R.: MiFID Readiness Survey. MiFID Joint Working Group (JWG) document (2005)
10. Gomber, P., Gsell, M., Reininger, C.: MiFID-Readiness - Die Umsetzung der MiFID "Markets in Financial Instruments Directive" in der deutschen Finanzindustrie. In: Oberweis, A., Weinhardt, C., Gimpel, H., Koschmider, A., Pankratius, V., Schnizler, B. (eds.) eOrganisation: Service-, Prozess-, Market-Engineering. Internationale Tagung Wirtschaftsinformatik, vol. 8, pp. 741–758. Universitätsverlag Karlsruhe (2007)
11. Harris, L.: Trading and Exchanges: Market Microstructure for Practitioners. Oxford University Press, Oxford (2003)
12. CESR: Ref.: CESR/07-050b. Best execution under MiFID. Public Consultation paper (2007)

13. Schwartz, R.A., Francioni, R., Weber, B.: The Equity Trader Course. John Wiley and Sons, Chichester (2006)
14. Gomber, P., Gsell, M.: Catching up with technology – The impact of regulatory changes on ECNs/MTFs and the trading venue landscape in Europe. Competition and Regulation in Network Industries 1(4), 535–557 (2006)
15. FIX Protocol: What is FIX? (2007), http://www.fixprotocol.org/
16. SEC: Regulation NMS. Final Rule Release No. 34-51808 (June 9, 2005)

Transferring Portfolio Selection Theory to Customer Portfolio Management – The Case of an e-Tailer

Dennis Kundisch[1], Stefan Sackmann[2], and Markus Ruch[2]

[1] University of Freiburg, Germany
Institute of Economic Research, Department of Information Systems
Platz der Alten Synagoge, 79085 Freiburg, Germany
dennis.kundisch@vwl.uni-freiburg.de
[2] University of Freiburg, Germany
Institute of Computer Science and Social Studies, Department of Telematics
Friedrichstr. 50, 79098 Freiburg, Germany
{sackmann,ruch}@iig.uni-freiburg.de

Abstract. Investing in existing customers is widely accepted as a promising strategy because it is believed to be less costly than attracting new ones. Recent research by Reinartz et al. [38] provides indications, however, that it could also be profitable to simultaneously focus on a customer segment being more transaction-oriented. In this contribution – using the example of an e-tailer – we specifically look at the question regarding the optimal mix of different customer segments within a customer portfolio. Portfolio Selection Theory is applied to develop a model to determine the optimal proportion of the different customer types from a value-based risk management perspective. A first evaluation is realized with a publicly accessible set of empirical data from the e-tailer CDNow. The results of the model provide a basis for the alignment of future CRM-activities.

Keywords: Customer Portfolio Management, Risk Management, E-Commerce, Portfolio Selection Theory, eCRM, integrated Risk-Return Management.

1 Introduction

In competitive economies, the main goal of every company is to maximize shareholder value [30]. Many authors, e.g. [20], [23], [35] argue that the basis of a company's profitability is its customers. Hence, an increase in shareholder value firstly requires an increase in customer value and a building of strong customer relationships as a means of gaining competitive advantage as proposed, e.g., by [40]. This insight has led to some fundamental changes in marketing theory as well as in practice towards a customer-centric view and the emergence of Customer Relationship Management (CRM). CRM centers on the valuation, selection, acquisition, retention, and development of durable customer relationships with the objective of allocating limited resources in order to maximize the value of a company.

The empirical findings regarding acquisition and retention strategies are mixed. A rule of thumb in marketing says that maintaining an existing customer relationship is

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 32–49, 2008.

far less expensive than acquiring a new one [2]. There are quite a number of contributions that suggest focusing on existing customers, assuming a positive lifetime-profitability relationship [33], [36], [37], [49]. However, Dowling et al. [12] question such a lifetime-profitability relationship by analyzing customer loyalty programs and have suggested a different examination. Garbarino et al. [17] as well as Ganesan [16] have shown that a differentiated treatment of transaction-oriented (short-term) and relationship-oriented (long-term) customers with appropriate marketing tools is advisable. Reinartz et al. [38] find strong evidence that transaction-oriented customers may be a very profitable segment, which should not be excluded from strategic considerations. Still the question remains about the optimal mix of these segments within a customer portfolio.

In this contribution, we suggest a model that helps in determining the optimal configuration of a customer portfolio of transaction- and relationship-oriented customers. The model is based on Portfolio Selection Theory according to [31] and particularly exploitable for e-tailers due to the potentials of the Internet for adequate relationship management [48]. Basically, the Internet as a communication channel enables e-tailers to vary the interaction according to the preferences of their customers. Results of a current online survey among German corporations indicate that most companies (76%) already rely on CRM software to support their CRM activities – and another 9% plan to do so by the end of 2008 [44]. However, the use of CRM software for controlling the overall customer portfolio according to superordinated quantitative goals like, e.g., risk management has been devoted little attention so far.

The paper is organized as follows. The next section provides an overview of recent research in customer relationship management with respect to quantitative customer portfolio management. Subsequently, we present our customer portfolio model and apply it to a publicly accessible data set of the online retailer CDNow. After a brief discussion, the basic, static model is expanded to get a dynamic view. Finally, the results of the paper are summarized and directions of further research are discussed.

2 Customer Relationship Management and Portfolio Theory

One basic condition for efficiently managing customer relations is the ability to estimate a customer's value for a company. A customer valuation concept, compatible with the principle of shareholder value that has gained broad attention in marketing literature and CRM research, is the Customer Lifetime Value (CLV). However, whereas marketing literature discusses the concept of CLV in detail, it still lacks practicability, since the estimation of future profitability is uncertain [50]. Thus, in a non-contractual relationship between a company and customers, which is particularly true of e-tailers selling, e.g., books or CDs on the Internet, the assumption of a deterministic world, where future cash flows are known with certainty, seems to be fairly unrealistic [28]. The consideration of risk, i.e. the deviation of cash flows from their expected value, is therefore crucial for a risk-averse decision maker.

Traditional customer valuation concepts concentrate on assessing individual customers [24]. However, it is not enough to evaluate customers one by one and consequently decide on acquisition and retention strategies. Not just the risk within a

single customer relation, but the risk *contribution* of each customer to the customer portfolio should be taken into account in a customer portfolio valuation. For reasons of simplification and the ability to better predict individual customer behavior, as well as strategic target group considerations, customers are often grouped into segments. These segments should be formed by trying to group together customers exhibiting similar buying behavior [51]. E.g. students that always buy a lot of books at the beginning of each term could form one segment,. Consequently, these segments can be addressed by specific marketing campaigns. Based on an appropriate segmentation, portfolio effects in terms of risk diversification – just like in a portfolio of financial assets – can be assumed and should be taken into account [11]. Interestingly, there are only a few contributions on this issue.

Incorporating risk and at the same time a portfolio view into customer valuation, some authors propose the usage of the weighted average cost of capital (WACC) of a company – based on the capital asset pricing model (CAPM) – as minimum rate of return [11], [20], [30], [25], [23], [28], [41], [42]. It is argued that the security market line may be used to adjust the specific WACC of any risky investment alternative and thus the beta value of a customer (segment) reflects the systematic business risk of the segment and the systematic financial risk of the company itself [30]. Consequently, the net present value (NPV) of the customer segment is given by the expected cash flows, discounted with the segment-specific risk-adjusted WACC. However, CAPM just takes systematic risks into account, whereas it is assumed that unsystematic risk can be neglected due to perfect diversification. Moreover, defining a market portfolio with respect to customers is quite a challenging task. Ryals [42] as well as Dhar et al. [11] define the "market portfolio" in the CRM context as the company's current customer base. In our view, this is not adequate if the company is following a growth strategy. This will most likely result in changing segments weights and the overall structure or the customer portfolio over time. Another shortcoming of the CAPM is the assumption of homogeneous expectations of all investors. This assumption is crucial for the existence of the market portfolio and the equilibrium on capital markets [8].

To take the risk of future cash flows as well as the customer portfolio into account, other authors, e.g. [3], suggest applying Portfolio Selection Theory according to [31]. The transformation of financial theory to other disciplines is not new though. Cardozo et al. [5], in comparison, suggest applying Markowitz's theory in product portfolio decisions. Although applying Markowitz's theory in a non-financial context also brings some shortcomings – which have led to an interesting discussion, see [6], [9] –, it nevertheless seems well-suited for providing reasons for customer portfolio decisions for the problem in question. This is discussed and argued in further detail in the next section.

3 A Model for Analyzing Customer Portfolios from an Integrated Risk-Return Perspective

"Customers as Assets" is the title of an often cited contribution by Gupta and Lehmann [19] – and this is exactly the viewpoint that is taken in the approach proposed below. Furthermore, it has been widely acknowledged in literature (e.g. [23], [43]) that customer relationships indeed share characteristics with other business assets: They

generate a risky cash flow. Thus, just like financial assets, they provide for an expected return, measured, e.g., as a CLV. Some of them provide a larger CLV but their cash flows may be unsteady and therefore more risky, whereas the CLV of others may be smaller but more stable [159]. However, there are also some major differences between customers and financial assets. This limits the transferability of Portfolio Selection Theory to customer relationships and is therefore discussed in the following section.

3.1 Transferability of Portfolio Selection Theory to Customer Relationships

The basic condition that an investment generates some kind of return which is uncertain is fulfilled for financial assets as well as for customers considered as assets. Besides this, there are a number of assumptions that come with the application of Portfolio Selection Theory [31] and have to be taken into consideration:

- *Investor's decision*: Risk-averse investors seek to maximize expected utility. They regard investment alternatives as being presented by a probability distribution of expected returns over some holding period. Decisions are solely based on return and risk, thus their utility curves are a function of expected return and the expected variance of returns only. There seems to be no substantial difference between financial assets and customers as assets with respect to the decision rational. For a given level of risk, investors prefer higher returns to lower returns and vice versa for a given level of expected return, investors prefer less risk to more risk. In order to make rational decisions according to the Bernoulli principle, one either has to assume a quadratic utility function – which will be rarely the case in reality – or assume normally distributed returns. The authors are not aware of any study that deals with typical distributions of CLVs within specific customer segments. This appears to be an open issue for further research.[1]
- *Given and stationary values for return, variances and co-variances*: The parameters that go into the optimization model are assumed to be given and stationary. Since objective probabilities are in general not known, "probability beliefs" or so-called subjective probability distributions are used [32]. At this point, there is again no major difference between financial assets and customers as assets. One might argue that there is much more information available on financial assets compared to customers and customer behavior. However, with powerful data warehouses being in place, customer behavior is no longer a "black box" and it is an open research question as to whether the estimated values for financial assets are far more accurate compared to the estimations that have to be made for customers. Equally, the assumption about stationary values is a limitation that hits the applicability of the model in both cases. A company unexpectedly selling off large parts of its business will alter the risk-return characteristics of its stock, just as a customer experiencing a change in his circumstances, e.g. through marriage, hence changing his buying behavior, which in turn influences his CLV. As long as the chosen time horizon for the application for the model is not too long, such unexpected changes should be not too severe.
- *One-period optimization*: The basic model of Markowitz just considers a one-time optimization. The weights of the assets are determined at the beginning so as to

[1] However to assume a normal distribution seems not only to be common in financial markets but also with respect to customer buying behavior, see e.g. [46].

maximize expected utility at the end of the period. The use of CLV as measure for return does not offend this assumption.

- *Complete and frictionless market*: In financial theory it is assumed that the market is without frictions, i.e. no taxes, no transaction costs and arbitrary divisibility of assets. Obviously, distortions can be found in financial markets as well as in the market of customers. Taxes diminish the expected CLV of each customer by affecting the underlying profit margin. Based on a tax system with constant tax rates on profits, taxes can be easily incorporated in the model.[2]

  However, with respect to transaction costs, there are clear differences between financial assets and customers as assets because there is no liquid market for customers. Although transaction costs can be lowered by new means of electronic communication and a high level of automation, acquiring and also "selling", i.e. getting rid off[3], of customers is still affiliated to transaction costs that may vary heavily according to the current market circumstances. This quality generally limits the applicability of portfolio theory in this context. Thus, we do not claim to apply our model with the objective of adapting the current customer portfolio according to the model results right away. The aim, in fact, is to obtain valuable information about the status quo of the current portfolio and to give an idea of where a theoretical optimum would lie. Thus, the information provided by the model shall be complementary to the use of other decision tools for strategic management and, in a first step, transaction costs are not of eminent importance here.[4]

  The last issue, the divisibility of assets may be a problem for companies in markets with just a few customers, like component suppliers for the automotive industry.[5] However, when talking about e-tailers like Amazon that globally addresses millions of private clients one may attest an approximate arbitrary divisibility.

In practice, a company needs both: An optimization of the customer portfolio [43] and, at the same time, an efficient management of individual customer relationships [38]. Ideally, these two issues should be simultaneously optimized. Due to the complexity of this optimization, it seems appropriate to split this process up into two steps: Firstly, a company decides on the overall customer portfolio. Here, just broad customer segments are considered. Secondly, the customer relationships within the (optimal) customer portfolio are managed as an individual basis as possible. Since a lot has already been written on the second step, we concentrate on the analysis of customer portfolios from an integrated risk-return perspective. This focus is also advisable because the application of Portfolio Selection Theory to customers as assets is, as discussed above, more suitable to the identification of the optimal overall

[2] However, since different companies, e.g. based in countries with different tax schemes, generally face different tax rates, there should be opportunities for tax arbitrage which will have an effect on market prices.

[3] This issue is known as "selective demarketing" according to Kotler et al. [27]. In the context with CLV, it has recently gained increasing importance in literature, see e.g. [52].

[4] In a further evolution of the model proposed below, transaction costs should of course be incorporated just as in financial theory, much like in, e.g., [18] – to name just one out of numerous contributions on the issue of portfolio theory with transaction costs.

[5] Here, contributions such as [4] about key account management and respective portfolio models, from a strategic management point of view – may be more useful.

customer portfolio. We start with a simple segmentation approach just distinguishing loyal, i.e. relationship-oriented customers and transaction-oriented customers, i.e. customers that do not feel committed to a company after buying some products there. The transfer and application of Portfolio Selection Theory requires a few (additional) assumptions about the customers and the characteristics of a customer relationship.[6]

3.2 Model Assumptions

(AC) Customers

Relationship-oriented customers are customers that repeatedly buy after a successful acquisition in $t = 0$ at two or more points in time $t \in \{0, \ldots, T\}$. Transaction-oriented customers are customers that buy just after a successful acquisition in $t = 0$ and subsequently stop buying from this e-tailer.[7]

(ACS) Customer Segments

There are two ex ante observable and disjunctive customer segments $i \in \{R; S\}$ in the market: Segment R comprises all relationship-oriented customers and segment S comprises all transaction-oriented customers. A segment i yields the cash inflow $CF_{i,t}^{in}$ which is the average periodic revenue per capita at time t, with $t \in \{0,\ldots, T\}$, as well as the average cash outflow per capita $CF_{i,t}^{out}$ that consists of direct costs, i.e. costs for acquisition, service and advisory as well as transaction costs. The segments are stable over the planning horizon, i.e. there is no transaction-oriented customer that becomes a relationship-oriented customer and vice versa.

(ACP) Customer Portfolio

The e-tailer has no customers in his portfolio before $t = 0$. The customer portfolio PF after the acquisition in $t = 0$ consists of $N \in \mathbb{N}$ customers. The portfolio shares r and s of the segments R and S (with $r + s = 1$ and $r \geq 0$, $s \geq 0$), given by the ratio of the number of customers in a segment and the total number N of customers in the portfolio, are the decision variables of the portfolio optimization in $t = 0$ for the whole planning horizon until $t = T$[8].

For each customer segment i, with $i \in \{R; S\}$, the average per capita cash flow Q_i is given by

$$Q_R = \left(\tilde{q}_{0,R}, \tilde{q}_{1,R}, \ldots, \tilde{q}_{T,R}\right) \quad \textit{and} \quad Q_S = \left(\tilde{q}_{0,S}, 0, \ldots, 0\right) \tag{1}$$

---

[6] The static model presented below is also applied in [29].

[7] If the e-tailer wants to perform the optimization process once again at a later point in time, it follows that transaction-oriented customers have to be regained in order to make a repeated purchase. Obviously, the two defined segments are idealized in order to allow for a first simple analysis.

[8] The planning horizon for the CLV calculation as well as for the optimization is assumed to be equal here. Of course, one may also model different planning horizons, however with the risk that the major cash inflows included in the CLV calculation may lie beyond the planning horizon for the optimization, leaving the decision maker potentially with an unintended result at the end of the planning horizon of the optimization.

The components $\tilde{q}_{t,i}$ are the average net cash flows per customer in customer segment *i* and represent the delta of cash in and outflows at time $t \in \{0,\ldots T\}$:

$$\tilde{q}_{t,i} = CF_{t,i}^{in} - CF_{t,i}^{out} \tag{2}$$

$\tilde{q}_{t,i}$ are independent and identically distributed random variables, which are given at the decision time $t = 0$. The average per capita Customer Lifetime Value CLV_i of segment *i*, which is also normalized to the number of customers in segment *i* at $t = 0$, is given by the expected NPV of Q_i (*z* denotes the risk free market interest rate):[9]

$$\mu_i = E(CLV_i) = \sum_{t=0}^{T} \frac{E(\tilde{q}_{t,i})}{(1+z)^t} \tag{3}$$

For the following model, we define the expected return per capita μ_i of customer segment *i* as $E(CLV_i)$ at time $t = 0$, as is done in equation (4). Hillier et al. [21] showed that if the net cash flows are supposed to be independent and identically distributed random variables, it may be concluded that the expected return per capita μ_i is asymptotically normally distributed. On the basis of assumptions *ACP* and *ADM*, the expected NPV per capita of the customer portfolio $E(CLV_{PF})$, shortly denoted as μ_{PF}, may be calculated as the sum of the weighted NPV of both segments' μ_i:

$$\mu_{PF} = E(CLV_{PF}) = r \cdot E(CLV_R) + s \cdot E(CLV_S) = r \cdot \mu_R + s \cdot \mu_S \tag{4}$$

(ACA) Customer Acquisition

Customers can be acquired only at $t = 0$. The customer segments are large enough and the e-tailer is comparably small enough so that there will never be a shortage of acquirable customers in a customer segment.

(ADM) Decision Maker

The risk-averse decision maker aims to maximize the utility per capita of the portfolio alternatives. The risk[10] of the expected return per capita of customer segment *i* is quantified by the standard deviation $\sigma_i = \sqrt{Var(CLV_i)}$. The risk σ_{PF} of the expected portfolio return per capita involves the standard deviation σ_i of the portfolio segments as well as their covariance Cov_{RS}, i.e. $\sigma_{PF} = \sqrt{r^2\sigma_R^2 + s^2\sigma_S^2 + 2rsCov_{RS}} = \sqrt{r^2\sigma_R^2 + s^2\sigma_S^2 + 2r\sigma_R s\sigma_S\rho_{RS}}$. The correlation coefficient $\rho_{RS} < 1$ is given in time period $t = 0$ and is constant over the planning horizon. For all possible values x assumed by the random variable CLV_{PF}, their utility is given by $u(x) = 1 - e^{-ax}$. The parameter *a* denotes the Arrow-Pratt measure that indicates the individual level of risk aversion[11] [1].

---

[9] A current survey of the market for CRM software according to Hippner et al. [22] shows that 58 of 78 CRM standard software products support the evaluation of customers. A study by Sackmann et al. [44] shows – at least for Germany – that monetary approaches of customer evaluation are increasingly applied. Thus, the necessary data may be already available in many companies.

[10] Risk includes both systematic risks (e.g. macroeconomic shocks, competitive environment) as well as unsystematic risks (e.g. product and pricing strategies of the e-tailer).

[11] For risk-averse decision maker: $a > 0$; for risk-neutral decision maker $a = 0$.

Schneeweiß showed that the only rational preference relation that meets assumption ADM, i.e. in case of normally distributed random variables, the utility function given in (ADM) and compatibility with the Bernoulli-Principle, is given by the following equation [47]:

$$\Phi_u(\mu_{PF}, \sigma_{PF}) = \mu_{PF} - \frac{a}{2}\sigma_{PF}^2 = U_{PF} \quad \rightarrow \quad \max \tag{5}$$

The parameters μ_{PF} and σ_{PF} both depend on the portfolio shares r and s of the two customer segments, which have to be chosen so that $\Phi_u(\mu_{PF},\sigma_{PF})$ is maximized. In the context of relationship valuation, $a/2$ is defined as a monetary factor that reflects the price per unit of risk, i.e. the reward asked by a risk-averse decision maker for carrying the risk σ_{PF}. Since the portfolio shares of the two customer segments sum up to one, the expected portfolio utility U_{PF} is a monetary per capita amount.

3.3 Portfolio Optimization – Static View

For the optimization, firstly μ_{PF} and σ_{PF} of all efficient portfolio alternatives have to be derived, i.e. the so-called efficient frontier (EF) is calculated, and secondly, the optimal portfolio based on the utility function provided in assumption (ADM) can be determined using equation (5).

Generally, the decision maker has to choose between portfolios with higher expected return accompanied by higher variance and portfolios with lower expected return and variance. Furthermore, he will only select a portfolio *PF*, which meets the following conditions and is therefore referred to as efficient [31]:

- The portfolio is a feasible portfolio, i.e. all portfolio weights are part of the feasible interval of $r,s \in [0; 1]$ and the portfolio shares sum up to one.
- If any feasible portfolio has a greater expected return, it must also have a greater variance of return than the portfolio *PF*.
- If any feasible portfolio has a smaller variance of return, it must also have a smaller expected return than the portfolio *PF*.

Analytically, this may be written as:

$$\min_r \quad \sigma_{PF}^2 = r^2 \cdot \sigma_R^2 + s^2 \cdot \sigma_S^2 + 2rs\rho_{RS}\,\sigma_R\sigma_S \tag{6}$$

The objective function (6) has to be minimized subject to the following constraints:

$$\overline{\mu}_{PF} = r \cdot \mu_R + s \cdot \mu_S \text{ (see equation (4))} \tag{7}$$

$$r \geq 0,\ s \geq 0 \text{ (see (ACP))} \tag{8}$$

$$r + s = 1 \text{ (see (ACP))} \tag{9}$$

Talking about the EF, one might think that with just two risky assets there may be four different cases to be differentiated, where relationship-oriented customers (*R*) have a lower or higher return and at the same time a lower or higher risk compared to transaction-oriented customers (*S*). From portfolio selection theory, we know that the set of efficient portfolios, where for a given level of risk no other investment opportunity offers a higher return, form the EF. Generally, the EF in a no-short-sales

setting corresponds to a concave curve[12] in a risk-return diagram starting at the minimum-variance portfolio (MVP) and ending at the asset with the highest return (see bold line in the left hand side of Figure 1).

In the financial markets, higher risk is generally associated with a higher return. Obviously this need not be true if the "assets" are customers – an issue that deserves more thought in future research. However, even though the individual asset *R* dominates asset *S*' (see left hand side of Figure 1), i.e. *R* offers a higher return for less(!) risk, due to a sufficiently small correlation coefficient it can transpire that a mixture of these two assets is still superior in comparison to a full investment in asset *R* –depending on the risk preferences of the decision maker. In case the correlation coefficient is sufficiently high and the segment with the higher return carries sufficiently lower risk compared to the other segment, the EF is just the point R.

Remarkably, all possible mixtures of the two customer segments within the customer portfolio lie *on* the line that connects *R* and *S*. Since the optimal portfolio has to be efficient, the point of tangency of the indifference curve (dotted line on the right hand side of Figure 1) and the EF, i.e. the point where the slopes of both functions are equal, represents the locus of the optimal portfolio at the given risk preference. Thus, after the EF has been determined, equation (5) may be used to determine the preferred customer portfolio weights.[13]

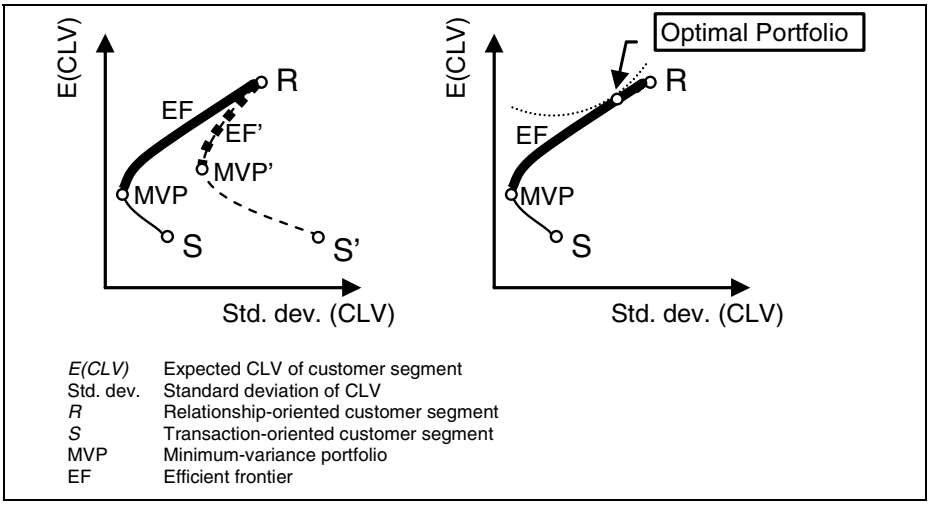


Fig. 1. Efficient Frontier with two customer segments

The knowledge of the preferred customer portfolio weights allows an e-tailer to identify the deviation between the optimal customer portfolio from an integrated risk-return perspective and the customer portfolio at the status quo. It should be noted that

[12] In the special cases where $\rho_{RS} = 1$ or $\rho_{RS} = -1$, the efficient frontier is a straight line.

[13] Finally, we may check if the utility per capita of the optimal portfolio covers the average NPV of direct and indirect fixed costs per capita. Buhl et al. [3] suggest a heuristic to account for these costs already within the optimization.

in the context of customer portfolios, it is generally not possible to adopt the portfolio by "selling" some shares of one segment in order to "buy" some of the other segment like in financial markets (see section 3.1). However, the result from the analysis is rather a hint as to where the focus of prospective acquisition efforts should lie: The customer segment where a shortage was identified, consequently leading to a (relative) increase in the portfolio weight of this segment.

With these first results, some limitations of the presented model should be noted. Obviously this is a simple model with just two segments and a one period optimization. Even though current data warehousing and data mining techniques are already very powerful, the parameter estimation, particularly with regard to the correlation coefficient, is a challenge. The segments are defined as idealized representations of customers in real life and it is assumed that with targeted marketing efforts it is possible to acquire the "right" customers. This may have become easier for e-tailers compared to traditional retail outlets, since much more computing power as well as data about (potential) customers are available and can be processed.[14] In addition, individualization and personalization is achievable at acceptable costs on the Internet. Nevertheless, there is still a long way to go until such targeted and successful marketing campaigns are in place. Considering all the limitations mentioned above, results of the analysis have to be handled with care. Besides these limitations, Markowitz's Portfolio Selection Theory still gets by with much less restrictive assumptions compared to CAPM proposed in many other contributions (see above).

Moreover, there is another limitation that has to be taken into account. An additional way to optimally manage the customer portfolio is given by taking up initiatives to work on the parameters themselves. This constitutes a major difference between financial markets and the "market of customers": While the return and volatility estimations of stocks, bonds and other financial assets are an exogenous input to the Markowitz model, CLV and its standard deviation are by far from being given.[15] In contrast, these are the parameters companies are trying hard every day to improve. A huge body of literature specifically deals with measures to increase CLV and to lower the risk of churning. Suggested measures include:

- Introduction of loyalty programs or customer clubs, often associated with customer cards (an example are frequent flyer programs of all major airlines),
- Service-related measures: E.g. professional complaint management or preferred service for existing customers,
- Price-related measures: E.g. discounts or kick-back payments,
- Product-related measures: E.g. using data mining techniques to determine the next product to buy [26],
- Communication-related measures: E.g. apologies, explanations, additional information, arguments of benefits [48].

[14] See [34] for a discussion about the potential resources and the value of having complete information about a customer for CRM purposes.

[15] This issue has been already discussed by [6], [9]. (Devinney et al. [10] call this issue with respect to product portfolio *managerial control*. Investors seldom have control over risk/return characteristics of an investment, while this is generally not true for risk/return characteristics of products.

Particularly for e-tailers, the Internet offers possibilities to combine these more traditional marketing instruments (product, price, place, and promotion) with precision, payment, personalization, and push and pull [7].

3.4 First Evaluation with an e-Tailer's Data Set

For a better clarification of the model, the analysis of an optimal customer portfolio will be performed by using a publicly available, exemplary data set from the online retailer CDNow[16]. The data set contains 2357 customers, who made their first purchases at the CDNow-website in the first quarter of 1997 and were observed over a period of 39 weeks. For further details of the data set see [13], [14]. The sold goods are CDs and the relationship between customers and retailer is non-contractual. In total, 1411 customers bought only once and 946 up to 30 times.

The Pareto/NBD-Model is applied as method for segmenting the overall customer portfolio into the segments of relationship- and transaction-oriented customers. Subsequently, the customer segments are evaluated based on the segments' CLVs and their respective standard deviations before the EF is determined.

Segmentation of the overall customer portfolio
The first step in optimizing a customer portfolio from a risk management perspective is to differentiate between relationship- and transaction-oriented customer groups. In doing this for the CDNow data set, the Pareto/NBD-Model from [45] – particularly suitable for the analysis of situations with non-contractual customer relationships – may be used, since it includes, among others, the transaction frequency. The model generates a probability $P(alive) \in [0,1]$ per customer revealing whether a customer is still active or not. After calculating $P(alive)$ for every customer, a cutoff value c has to be defined that separates according to their activity relationship- from transaction-oriented customers. Due to a sharp kink of the curve of the $P(alive)$-values, c is set to 0.21 as the value to divide the two customer segments (see Table 1).[17] The customers with $P(alive)$ higher c represent the relationship-oriented segment R and those with a lower $P(alive)$-value – who bought only once or after several times in the beginning nothing more for the rest of the observation period – the transaction-oriented segment S.

Valuating the different customer segments
The base for the average per capita Customer Lifetime Value CLV_i of segment i are the individual CLVs of the customers. To calculate these and to get the NPV, it is necessary to know the exact purchasing dates so that the values can be discounted to t_0. Unfortunately, this information is not included in the data set of CDNow, so the needed purchasing dates are modeled by using random numbers following an equal distribution. An average cash flow per purchase has also been defined and was normalized to 1. This enables a simple calculation of the CLVs by summing up the number of purchases per quarter and discounting these to t_0 with an assumed risk-free interest rate z of 1 % per quarter. To operate the portfolio optimization, simply an average CLV for each consumer segment is needed together with the standard deviation. The results for the CDNow data set are shown in Table 1. Apparently, at least for this data set, a higher return goes hand in hand with a higher risk.

[16] The data set is available at http://brucehardie.com/notes/004/

[17] Using a value of 0.5 does not change the general results of the following evaluation. For a more detailed discussion about the calculation see [29].

Table 1. Characteristics of CDNow customer segments *R* and *S*

	Number of customers	Average CLV_i	Standard deviation of CLV_i
Segment *R* (relationship-oriented customers)	698 (30 %)	3.8	2.8
Segment *S* (transaction-oriented customers)	1659 (70 %)	1.3	1.1

Calculating the efficient frontier and determining CDNow's customer portfolio

Figure 2 shows alternative EF for three assumed correlation coefficients for the two identified segments, namely $p = 0$ (assuming stochastic independence), $p = 0.5$ (assuming a positive correlation), $p = 1$ (assuming a perfect positive correlation). As expected, the higher the risk diversification potential is, the smaller the correlation of the expected CLV distribution of these two segments is.

Assuming independence of the CLV-development of the two segments ($p = 0$), the minimum variance portfolio (MVP) is characterized by a CLV of 1.6 and a respective standard deviation of 1.0. In this MVP, 86% of the customers are transaction-oriented and 14% are relationship-oriented. Thus, even extremely risk-averse decision makers would choose a portfolio consisting of shares of both segments instead of just addressing transaction-oriented customers.

The current situation for the data set of CDNow is depicted by the triangles in Figure 2. The expected CLV per capita of the portfolio is 2.0. If the portfolio construction was a deliberately managed process at CDNow this would imply a quite risk-averse management. Otherwise, the result may suggest that CDNow rethinks their customer acquisition and retention strategy in the future; given similar customer behavior.

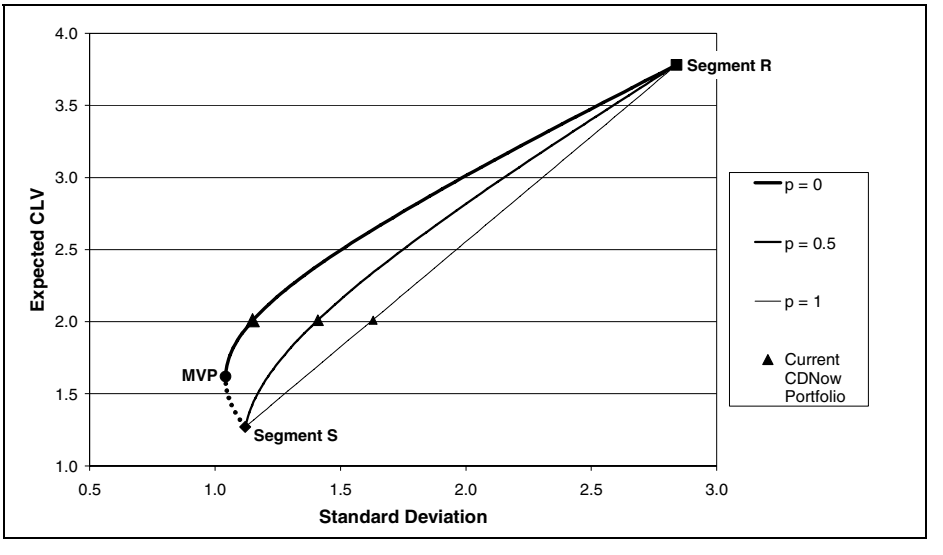


Fig. 2. Alternative efficient frontiers for CDNow Data set

3.5 Portfolio Optimization – Dynamic View

Of course, a one period optimization is not appropriate for businesses that permanently want to operate in a specific market. Also, the market environment might change leading to changing parameter estimations, which should also be reflected in the optimization process over time. One simple extension of the presented model is the application of the model after each period. This changes the optimization process a little bit, since by definition relationship-oriented customers stay with the e-tailer for several periods. These loyal customers do not have to be acquired again in the next period. Hence, their expected CLV will change due to the acquisition costs being regarded as sunk costs.

Moreover, there are at least two more reasons why CLV calculation for these customers has to be adapted. Firstly, the estimations may be adapted to the actual behavior in the first observed period. With this data, it should generally be possible to reach more accurate estimations. Secondly, if the CLV calculation is not primarily based on the expected termination of the relationship but – as can be observed in practice quite often – just on the near and midterm future cash flows (e.g. one to four years), the CLV estimations will tend to go up – if the relationship is appropriately managed – with increasing overall relationship duration (e.g. [36], see Figure 3).

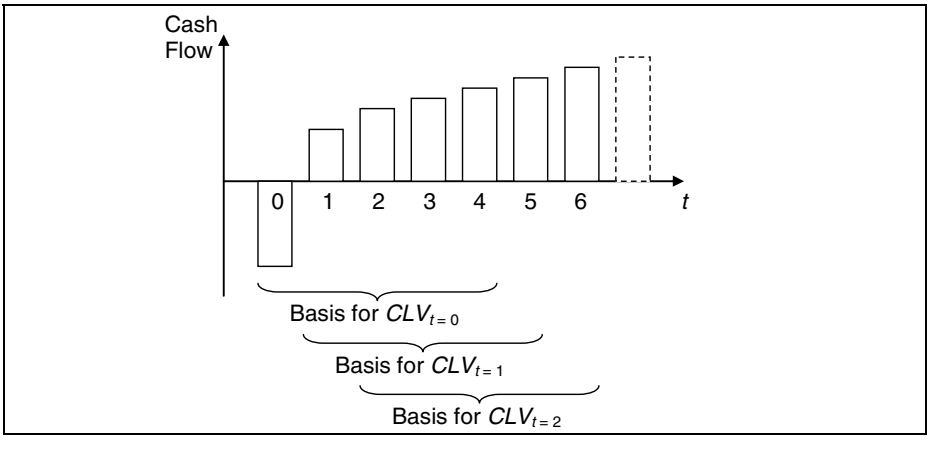


Fig. 3. Periodical CLV calculation for relationship-oriented customer (segment)

However, not only the parameter estimations have to be altered in this optimization process over time. Obviously, *existing* relationship-oriented customers are not a decision variable since it is not the question of whether they should be acquired or not. They are existing customers providing for future cash flows which are (still) risky and which will deviate from cash flow estimations for customers that shall be acquired (see above). Hence, these customers shall not be "fired" but contribute to the overall return and risk position of the e-tailer. Thus these customers form another *constraint* in the optimization process. Analytically, equations (6 - 9) change to

$$\min_{r_{new}} \quad \sigma_{PF}^2 = \sum_k k \cdot \sigma_k^2 + \sum_k \sum_{l \neq k} kl \rho_{kl} \sigma_k \sigma_l \quad \text{with} \quad k, l \in \{r_{new}, s, r_{exist}\} \tag{10}$$

$$\mu_{PF} = r_{new} \cdot \mu_{R_{new}} + s \cdot \mu_S + r_{exist} \cdot \mu_{R_{exist}} \tag{10}$$

$$r_{new} \geq 0,\ s \geq 0\ (r_{exist} = \text{const. from optimization in previous period}) \tag{11}$$

$$r_{new} + s + r_{exist} = 1 \tag{12}$$

This situation is depicted in Figure 4. Each period, the existing customers tend to move to the upper left in the risk-return diagram (illustrated by the dotted arrow in the diagram to the right). The parameters for already acquired segments of relationship-oriented in each period can be determined prior to a new optimization process. Since these parameters are an exogenously given input to the optimization model, for reasons of clarity, they may also be merged to form just one segment of existing relationship-oriented customers as a constraint for the optimization. At the same time, the MVP will also move up and to the left in each period and thus the EF and also the optimal portfolio possibly changes from period to period.

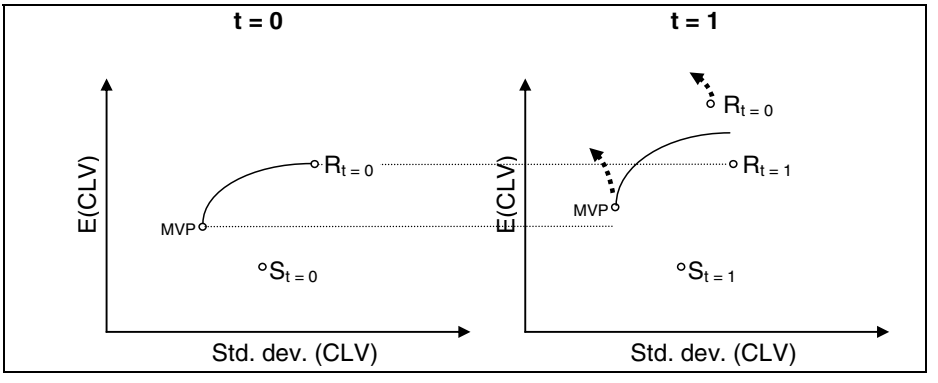


Fig. 4. Optimization over time

So what can we learn from this dynamic view? Firstly, relationship-oriented customers (*R*) already in the customer base of an e-tailer form a new constraint that has to be determined and considered in each period. Secondly, the segment of already existing relationship-oriented customers (*R*) tends to drift upwards and to the left also shifting the MVP. Thus, the EF also tends to drift upwards in each period. Consequently, a higher utility level may be realized. This does not seem surprising, since new customers are acquired after each period while relationship-oriented customers from previous periods stay with the e-tailer (see assumption AC). Hence, the e-tailer is growing. However, if management is at least slightly risk averse, sooner or later a state of saturation is reached: If the portfolio of existing (relationship-oriented) customers is comparably large in relation to the amount of money that may be invested to acquire a small number of new customers, it may become perfectly rational to focus just on transaction-oriented customers in order to realize risk diversification effects.[18] The higher the risk aversion of management, the sooner such a "steady state" will be reached.

[18] It is assumed though that the correlation between newly acquired and already existing relationship-oriented customers equals 1.

Customer portfolio management is an ongoing process. Acquisition efforts and measures to increase customer loyalty have to be adapted and adjusted in each period. Still one has to be aware that we perform a sequential set of one-period-optimizations, always taking the results of the optimization of the previous period into account. There is no guarantee that following this sequential approach leads ex post to an overall maximum. Nevertheless, the extended version of the basic model is a first step in the direction to more realistic assumptions.

4 Conclusion and Outlook

After years of proposing a focus on turning customers into loyal customers, it seems that in the last couple of years a more differentiated view has made its way in literature as well as in practice. Moreover, recent studies suggest that disloyal, transaction-oriented customers who do not expect and ask for bonus programs or other forms of customer loyalty measures may still constitute a quite profitable customer segment in the overall customer portfolio of an e-tailer. Acknowledging these issues, new questions arise – if management is considered as being risk-averse – concerning the appropriate mix of two or more customer segments with different buying behavior. Traditional customer evaluation methods mostly focus on evaluating customers one by one. Including risk into this consideration most often comes as a mere premium on the risk free discount rate for the CLV calculation. We propose a different approach in this contribution.

With regard to customers as assets, we transfer and apply Portfolio Selection Theory due to Markowitz [31], to the issue of customer portfolio optimization. With our novel model, it is possible to account for different buying behavior of two (or potentially more) customer segments while having an integrated view on return (CLV) and risk (here measured as the standard deviation of the CLV within a customer segment). In a quite simple analysis, just looking at relationship-oriented and transaction-oriented customers, it turns out that an optimal mix of these two customer segments within a customer portfolio will typically include shares of both segments. Hence, focusing just on (potentially) loyal customers may be optimal with respect to the expected CLV, but not necessarily with respect to the overall expected utility of a customer portfolio, incorporating the risk of future cash flows. Moreover, we showed the applicability of our model using a publicly available data set of the e-tailer CDNow and provided for a first extension to get a more dynamic view into the process of customer portfolio optimization. The results of the extended analysis showed that a point of saturation may be reached leading to the situation that from this period onwards just transaction-oriented customers are targeted in the acquisition efforts.

The results of the analysis contribute to the understanding of customer portfolio management and optimization. It may facilitate a decision maker concerning the allocation of marketing budget for customer acquisition. Marketing measures – particularly when talking about e-tailers operating primarily on the Internet – may be designed to address specific customers segments. Still the results above have to be carefully applied. Obviously, there will be market circumstances where an unlimited number of customers in each defined segment cannot be assumed and an ex ante segmentation just based on some limited data will (often) be associated with risk.

Nevertheless, customer portfolio optimization and management are interesting and demanding issues not only in research. A current study among German corporations reveals that there is substantial demand for quantitative approaches in customer portfolio management within the companies' CRM-activities [44]. However, a mere quarter of the respondents in that survey already have quantitative customer evaluation methods, such as the CLV, in place. So the basis for the application of the approach suggested above is still comparably small – but growing.

There are a lot of open research questions in these areas. Based on this contribution, we identified four major topics for future steps. First, an empirical investigation with an e-tailer going beyond the limited data set of CDNow is envisaged. Second, the presented model should be expanded to incorporate an arbitrary number of segments and issues like that of the customer retention rate should be included. Third, the issue of managing customer relationships in order to increase the expected CLV and decrease the risk of churning on the one hand has to be formally combined with a simultaneous optimization of the overall customer portfolio. Fourth, the dynamic character of customer portfolio management as an ongoing process should be incorporated in the model beyond the first step discussed in this contribution.

Acknowledgements

The authors wish to thank two anonymous reviewers for their assistance and also the participants of the INFORMS International Puerto Rico 2007 Meeting and the FinanceCom 2007 Workshop in Montreal. Finally, the authors extend their thanks the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) for the financial support during the EVENT project as part of the "Internetökonomie" research program.

References

1. Bamberg, G., Coenenberg, A.: Betriebswirtschaftliche Entscheidungslehre, 12th edn., Vahlen, Munich (2004)
2. Bitran, G., Mondschein, S.: A Comparative Analysis of Decision Making Procedures in the Catalog Sales Industry. European Management Journal 15, 105–116 (1997)
3. Buhl, H.U., Heinrich, B., Singbartl, A.: Customer Portfolio Management in the Financial Services Industry. Discussion Paper, University of Augsburg, Augsburg (2005)
4. Campbell, N.C.G., Cunningham, M.T.: Customer analysis for strategy development in industrial markets. Strategic Management Journal 4, 369–380 (1983)
5. Cardozo, R.N., Smith, D.K.: Applying Financial Portfolio Theory to Product Portfolio Decisions: An Empirical Study. Journal of Marketing 47, 110–119 (1983)
6. Cardozo, R.N., Smith, D.K.: On the Use of Financial Portfolio Theory in Marketing Decisions: A Reply to Devinney, Stewart, and Shocker. Journal of Marketing 49, 113–115 (1985)
7. Chen, C.Y.: The comparison of structure differences between internet marketing and traditional marketing. International Journal of Management and Enterprise 3, 397–417 (2006)

8. Copeland, T., Weston, J., Shastri, K.: Financial Theory and Corporate Policy, 4th edn. Pearson Addison Wesley, Boston (2005)
9. Devinney, T.M., Stewart, D.W., Shocker, A.D.: A Note on the Application of Portfolio Theory: A Comment on Cardozo and Smith. Journal of Marketing 49, 107–112 (1985)
10. Devinney, T.M., Stewart, D.W.: Rethinking the Product Portfolio: A Generalized Investment Model. Management Science 34, 1080–1095 (1988)
11. Dhar, R., Glazer, R.: Hedging Customers. Harvard Business Review 81, 86–92 (2003)
12. Dowling, G.R., Uncles, M.: Do Customer Loyalty Programs Really Work? Sloan Management Review 38, 71–82 (1997)
13. Fader, P.S., Hardie, B.G.S.: Forecasting Repeat Sales at CDNOW: A Case Study. Interfaces 31, 94–107 (2001)
14. Fader, P.S., Hardie, B.G.S., Lee, K.L.: Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model. Marketing Science 24, 275–284 (2005)
15. Ford, D., Gadder, L.-E., Hakansson, H., Snehota, I.: Managing Business Relationships, 2nd edn. John Wiley & Sons, West Sussex (2003)
16. Ganesan, S.: Determinants of Long-Term Orientation in Buyer-Seller Relationships. Journal of Marketing 58, 1–19 (1994)
17. Garbarino, E., Johnson, M.S.: The Different Roles of Satisfaction, Trust, and Commitment in Customer Relations. Journal of Marketing 63, 70–78 (1999)
18. Goldsmith, D.: Transactions Costs and the Theory of Portfolio Selection. Journal of Finance 31, 1127–1139 (1976)
19. Gupta, S., Lehmann, D.: Customers as Assets. Journal of Interactive Marekting 17, 9–24 (2003)
20. Gupta, S., Lehmann, D., Stuart, J.: Valuing Customers. Journal of Marketing Research 41, 7–18 (2004)
21. Hillier, F.S., Heebink, D.V.: Evaluating Risky Capital Investments. California Management Review 8, 71–80 (1965)
22. Hippner, H., Hoffmann, O., Wilde, K.D.: CRM Studie 2007. Katholische Universität Eichstätt-Ingolstadt (2007)
23. Hogan, J., Lehmann, D., Merino, M., Srivastava, R., Thomas, J., Verhoef, P.: Linking Customer Assets to Financial Performance. Journal of Service Research 5, 26–38 (2002)
24. Hogan, J., Lemon, K., Libai, B.: What is the True Value of a Lost Customer? Journal of Service Research 5, 196–208 (2003)
25. Hopkinson, G., Lum, C.Y.: Valuing customer relationships: Using the capital asset pricing model (CAPM) to corporate relationship risk. Journal of Targeting, Measurement and Analysis for Marketing 10, 220–232 (2002)
26. Knott, A., Hayes, A., Neslin, S.: Next-product-to-buy models for cross-selling applications. Journal of Interactive Marketing 16, 59–75 (2002)
27. Kotler, P., Levy, S.: Demarketing, Yes, Demarketing. Harvard Business Review 49, 74–80 (1971)
28. Kumar, V., Ramani, G., Bohling, T.: Customer Lifetime Approaches and Best Practice Applications. Journal of Interactive Marketing 18, 60–72 (2004)
29. Kundisch, D., Sackmann, S., Ruch, M.: CRM and Customer Portfolio Management for E-Tailers. In: Proceedings of the Forty-First Annual Hawaii International Conference on System Sciences (HICSS-41), IEEE Computer Society Press, Los Alamitos (2008)
30. Lumby, S., Jones, C.: Investment appraisal and financial decisions. Thomson Learning, London, England (2001)
31. Markowitz, H.: Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons, New York (1959)

32. Markowitz, H.: Foundations of Portfolio Theory. Journal of Finance 46, 469–477 (1991)
33. Morgan, R.M., Hunt, S.D.: The Commitment-Trust Theory of Relationship Marketing. Journal of Marketing 58, 20–38 (1994)
34. Padmanabhan, B., Zheng, Z., Kimbrough, S.: An Empirical Analysis of the Value of Complete Information for eCRM Models. MIS Quarterly 30, 247–267 (2006)
35. Rappaport, A.: Creating Shareholder Value: A Guide for Managers and Investors. The Free Press, New York (1998)
36. Reichheld, F.F., Sasser, E.W.: Zero Defections: Quality Comes to Service. Harvard Business Review 68, 105–111 (1990)
37. Reichheld, F.F., Teal, T.: The Loyalty Effect. Harvard Business School Press, Boston (1996)
38. Reinartz, W.J., Kumar, V.: On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. Journal of Marketing 64, 17–35 (2000)
39. Reinartz, W.J., Kumar, V.: The Mismanagement of Customer Loyalty. Harvard Business Review 80, 86–94 (2002)
40. Rust, R., Lemon, K., Narayandads, D.: Customer Equity Management. Prentice Hall, Englewood Cliffs (2005)
41. Ryals, L.: Are your customers worth more than money? Journal of Retailing and Consumer Services 9, 241–251 (2002)
42. Ryals, L.: Measuring risk and returns in the customer portfolio. Journal of Database Marketing 9, 219–227 (2001)
43. Ryals, L.: Making customers pay – measuring and managing customer risk and returns. Journal of Strategic Marketing 11, 165–177 (2003)
44. Sackmann, S., Kundisch, D., Ruch, M.: CRM, Kundenbewertung und Risk-Return-Steuerung im betrieblichen Einsatz. HMD – Praxis der Wirtschaftsinformatik 259 (2008)
45. Schmittlein, D.C., Morrison, D.G., Colombo, R.: Counting your Customers: Who are they and what will they do next? Management Science 33, 1–23 (1987)
46. Schmittlein, D.C., Peterson, R.A.: Customer Base Analysis: An Industrial Purchase Process Application. Marketing Science 13, 41–67 (1994)
47. Schneeweiß, H.: Entscheidungskriterien bei Risiko. Springer, Heidelberg (1967)
48. Scullin, S.S., Fjermestad, J., Romano Jr., N.C.: E-relationship marketing: changes in traditional marketing as an outcome of electronic customer relationship management. Journal of Enterprise Information Management 17, 410–415 (2004)
49. Sheth, J.N., Parvatiyar, A.: Relationship in Consumer Markets: Antecedents and Consequences. Journal of the Academy of Marketing Science 23, 255–271 (1995)
50. Srivastava, R., Tasadduq, S., Fahey, L.: Driving Shareholder Value: The Role of Marketing in Reducing Vulnerability and Volatility of Cash Flows. Journal of Market Focused Management 2, 49–64 (1997)
51. Wedel, M., Kamakura, W.: Market Segmentation: Conceptual and Methodological Foundations. Kluwer, Dordrecht (2000)
52. Zeithaml, V.A., Rust, R.T., Lemon, K.N.: The Customer Pyramid: Creating and Serving Profitable Customers. California Management Review 43, 118–142 (2001)

Trends in European Cross-Border Securities Settlement – TARGET2-Securities and the Code of Conduct

Torsten Schaper

Johann Wolfgang Goethe-University Frankfurt, Chair of e-Finance,
Robert-Mayer-Strasse 1,
60054 Frankfurt, Germany
Schaper@wiwi.uni-frankfurt.de

Abstract. Securities settlement in Europe is still said to be highly inefficient for cross-border transactions. This paper provides an overview of the market, regulation, and recent approaches that aim to improve the efficiency of the cross-border settlement in Europe. The European code of conduct and TARGET2-Securities are identified as primary concerns. The different approaches to improve cross-border clearing and settlement are presented and compared with each other.

Keywords: Securities Clearing and Settlement, Central Counterparty, TARGET2-Securities, European Code of Conduct for Clearing and Settlement.

1 Introduction

During the last decades trading on securities markets increased considerably. Moreover, the international trade of securities strongly increased. This implies that not only more transactions need to be settled, but more of these transactions need cross-border settlement [1]. The importance of an efficient securities settlement system lies in the safer transfer of ownership of assets against payment. The significance of settlement derives from the fact that it must be viewed as a subset of transaction costs facing an investor in effecting a trade. Such systems must be developed in a way to minimise the risks involved in securities transactions and it must offer lower costs, which do not hinder the intention to trade securities [2, 3]. In general, the arrangements between actors in trading, clearing, and settlement within any country have been organised on the basis of direct or indirect access to the local central securities depository (CSD) and by accounts hold at the central bank. Gaining access to this local market by a foreign investor involves costs in establishing a relationship to a local agent, which financial institutions pass over to the investor. The costs of cross-border securities settlement within Europe are significantly higher than in domestic settlement[1] [2]. A study by NERA showed that there is a significant gap

[1] The pricing of post-trading services is complex. Several studies have attempted to analyse the pricing. Most studies concluded that cross-border prices and costs are considerably higher than the corresponding costs and prices for domestic transactions. None of the results have been accepted as providing an accurate description of the prices or costs incurred by investors in acquiring post-trade services in Europe. Example studies are [3, 4, 5, 6].

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 50–65, 2008.

between the costs of settling securities transactions in the United States and in Europe. Whereas a securities transaction settled on a net basis costs around €0.10 in the United States, the costs in European markets range between €0.35 and €0.80 [5]. The main reason is the fragmented structure of the European settlement industry.

Technological innovations and a changing regulatory environment are fundamental catalysts behind the past and the future changes in settlement. In the past the consolidation of the post-trade infrastructure and the introduction of the central counterparty (CCPs) have improved the efficiency of post-trade services. Once implemented in 2008, the European code of conduct is expected to cut the costs of trading shares in the member states by enhancing price transparency and increasing competition. Furthermore, TARGET2-Securities (T2S) as technical platform for securities settlement could have a significant impact on the settlement of securities in Europe.

The aim of this paper is to present and analyse recent approaches for improving efficiency of cross-border securities settlement in Europe. The different approaches are compared with each other regarding the capability to improve the efficiency of the European securities settlement.

The paper is organised as follows. Section 2 introduces functional definitions of clearing and settlement and the performing institutions. It is followed by an overview of the consolidation that has taken place in the settlement industry and the introduction of the CCP in European securities markets and its impact on the settlement of securities. In section 4 the code of conduct is presented and the current status of implementation analysed. Section 5 shows T2S and presents the current status of the project. In section 6 a comparison of the presented approaches to improve efficiency of securities settlement is made. Section 7 concludes.

2 Clearing and Settlement of Securities Transactions

Clearing and settlement are required after two parties have decided to the transfer ownership of a security. Clearing and settlement services deal with the execution of a trade. The need arises after any trade, regardless of whether the parties trade over an exchange or over the counter, and whether the trade involves domestic or international securities. Clearing of a securities transaction confirms the legal obligation from the trade. Clearing involves the calculation of mutual obligations of market participants and determines what each counterpart receives. Clearing houses, CSDs, or international central securities depositories (ICSDs) are the providers of clearance [2]. Central counterparty (clearing) is not included in the definition of clearing. A central counterparty is an entity that interposes itself between the transactions of the counterparties in order to assume their rights and obligations, acting as buyer to every seller and seller to every buyer. The original legal relationship between the buyer and the seller is thus replaced by two new legal relationships: between the CCP and the buyer and between the CCP and the seller. The substitution of the original counterparty by a new contractual counterparty is called a contract novation. The CCP thus takes over the counterparty risk and guarantees the clearing and settlement of the trade. A CCP is a service offered by a clearinghouse. Following the clearing stage the second operation is settling a trade. Settlement is the exchange of cash or assets in return for other assets or cash and transference of ownership of those assets and cash. A CSD is the organisation

that performs these functions. Some post-trade services are not related to a securities transaction, but are needed on an ongoing basis to administer securities on behalf of the owner. The process of settlement is typically linked with custody and safekeeping.

A stylised domestic equity transaction instruction flow consists of six steps [4]. For a cross-border trade investors rarely access a foreign system directly, but instead typically rely on intermediaries for this purpose. A cross-border transaction normally involves one of three basic models: using a link between two CSDs (the local and the foreign), using a network of local brokers (who have access to the local CSDs), or using an ICSD. As a cross-border trade increases the number of clearing and settlement intermediaries that have to be accessed to complete the trade. This is likely to increase the costs of a cross-border trade. In a cross-border transaction two CSDs and two central banks are involved in the process. A stylised cross-border equities transaction instruction flow consists of nine steps [4].

3 Development of Clearing and Settlement in Europe

3.1 Consolidation of Settlement Market Infrastructure

Compared with the United States, the settlement industry in Europe is fragmented. Until recently, each country in Europe had its own stock exchanges. In addition, individual countries have their own distinctive legal and regulatory apparatuses. Settlement in Europe shows its origins in a patchwork of national systems. At the beginning, there were several institutions offering settlement services at the national level. At the national level, the consolidation has taken place and in most countries only one CSD has prevailed [8]. The domestic settlement systems are efficient within the national boundaries. The costs per transactions in domestic settlement are as expensive as in the United States, but European CSDs realise higher margins [9]. In contrast, the settlement of cross-border transactions in Europe is not efficient [8]. The main reason for the fragmented European settlement industry is that securities were traded at national level, partly as result of the existence of different currencies. As result, several CSDs at the European level continue to coexist and only recently consolidation has taken place [1]. In the EU the number of settlement engines declined from 23 in 1999 to 18 in 2004 [10]. The consolidation of the European financial market is not only taking place on a horizontal level. Different types of integration and consolidation in trading, clearing, and settlement industry can be observed [10]:

- The horizontal integration involves mergers of institutions or systems providing similar services in different markets, such as the merger of securities settlement systems.
- The vertical integration involves mergers of institutions providing different, but integrated services which are processed along the securities transaction chain within a single entity or group of entities.

Fig. 1 shows the integration in selected European countries.

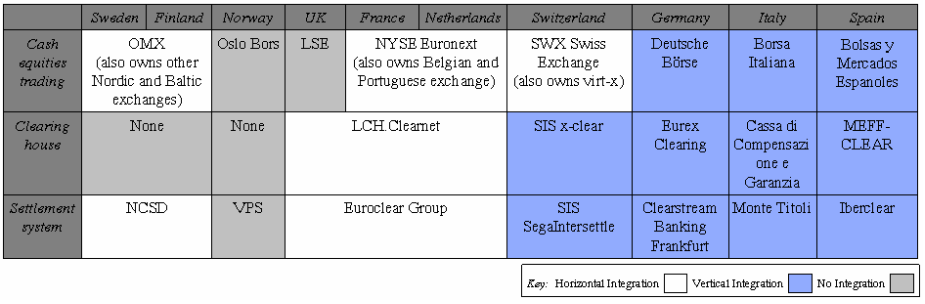


Fig. 1. Overview of equities market infrastructure in selected European countries

3.2 Central Counterparty

CCPs have been first introduced in derivative markets and, at least in the EU, only recently in securities markets. Today, a CCP is established in nearly all major securities markets. An important driver for the increased use of CCP services in cash markets is the increased use of electronic order books to match trades. The anonymity of the electronic order books complicates the risk management of market participants, because counterparty risk cannot be managed through their choice of counterparty. A CCP is a useful service to clear and settle anonymous trades, since the market participant can manage its counterparty risk towards the CCP. Besides performing the CCP-clearing function, most clearing houses perform other functions as well, like collateral management and netting. They enhance the efficiency of securities clearing and settlement and redistribute the risk between market participants. A CCP creates the possibility to net on a multilateral instead of a bilateral basis, since all debit and credit positions of a particular clearing member can be netted, resulting in only one single net position towards the CCP. Settlement costs are substantially reduced due to the reduction of the number of settlements. In general, the introduction of a CCP is followed by increased liquidity, smaller spreads, and lower cost [7].

Like in the settlement infrastructure, a consolidation of clearing houses in Europe has taken place. Between 1999 and 2004 the number of CCPs declined from 14 to 8 [10]. Two main clearing houses for securities have established in Europe, LCH.Clearnet and Eurex Clearing [2]. In addition to these larger CCPs there are a number of smaller ones.

In May 2006, the London Stock Exchange and SIS x-clear have announced that they will provide market participants with a choice of the clearing provider for equity trades. Beside the London Stock Exchange, virt-x is the only European exchange offering two CCPs to their customers. The intended business model is expected to deliver price reduction and service amendments to the markets through competition.

4 The European Code of Conduct for Clearing and Settlement

The efforts to improve the efficiency and safety of cross-border post-trading arrangements in the EU date as far back as the 1970s. The main reason behind the

growing interest in post-trade services in the last few years has been the explosive growth in securities trading [6].

The Financial Services Action Plan (FSAP) is the European Commission's (EC) main instrument for achieving the single market in financial services. It was published in 1999 and endorsed by the Lisbon European Council in 2000. It consists of a set of measures to remove barriers, so as to provide a legal and regulatory environment that supports the integration of the European financial markets. The Committee of Wise Man was appointed by the Economic and Financial Affairs Council (ECOFIN) in 2000 in order to support the extensive program of FSAP. Specific recommendations of the final report, which refers to clearing and settlement, target to the necessary systems restructuring, to avoid the high costs of cross-border settlement. Although the consolidation should be driven by the private sector's forces, public policy will be needed to remove the obstacles to consolidation. Among the most important policy requiring assessments are: open and non-discriminatory access to CSDs, exclusive agreements, the soundness of technical linkages, and the implication of a single CCP. The report also mentions the possible need to separate clearing system issues from settlement and understanding that an efficient clearing is of public benefit. Additionally, it shall be considered whether the EU needs to establish a framework for clearing and settlement activities or not. Finally, competition among market participants must also be addressed in order to avoid practices which are not consistent with the community's competition policy [11].

In 2001, the Giovannini Group was assigned by the Commission to analyse the current situation for cross-border clearing and settlement in the securities markets, to consider the requirements against which the efficiency of possible alternative arrangements for clearing, settlement, and depository services can be assessed, and to identify some possible alternative arrangements for clearing, settlement, and depository functionalities [4]. The first report on EU cross-border clearing and settlement arrangements was published in 2001, describing the functions of clearing and settlement systems, the arrangements in Europe, and the existing 15 barriers[2] hindering integration. The second report in 2003 brought proposals to overcome the barriers evidenced in the previous work, followed by an assessment of current models of consolidation. A time framework of three years was recommended (to market participants and regulators) for the execution of the necessary steps to create a level playing field, where efficient clearing and settlement systems could coexist or consolidate [12].

Later in 2006, the EC published a working paper on post-trade activities in which the EC set out measures to achieve the objective of integrated, competitive, safe, and cost-effective clearing and settlement in the EU [6].

The European Commissioner for Internal Markets and Services, Charlie McCreevy, announced in July 2006, that he would not propose an EU directive on post-trade services, but that he would strongly encourage the sector to adopt a code of conduct, which delivers price transparency, interoperability, and unbundling of post-trade services. Finally on 7th November 2006, the European code of conduct for clearing and settlement was presented by the post-trade industry. The intention of the code of conduct is to

[2] The barriers are differentiated into technical requirements and market practice, taxation, and legal certainty.

establish a strong European capital market and to allow investors the choice to trade any European security within a consistent, coherent, and efficient European framework. The aim is to offer market participants the freedom to choose their preferred provider of services separately at each layer of the transaction chain (trading, clearing, and settlement) and to make the concept of cross-border redundant for transactions between EU member states. The code of conduct is a voluntary self-commitment[3] and will adhere to a number of principles on the provision of post-trading services for cash equities. It includes clearing and CCP, settlement and custody services, and some of the elements of the code also apply to trading activities. The supporting organisations are represented by the Federation of European Securities Exchanges (FESE), the European Association of Central Counterparty Clearing Houses (EACH), and the European Central Securities Depositories Association (ECSDA). The implementation of the code of conduct consists of three phases (see Fig. 2). To monitor the implementation of the code of conduct, the Commission has set up the Monitoring Group.

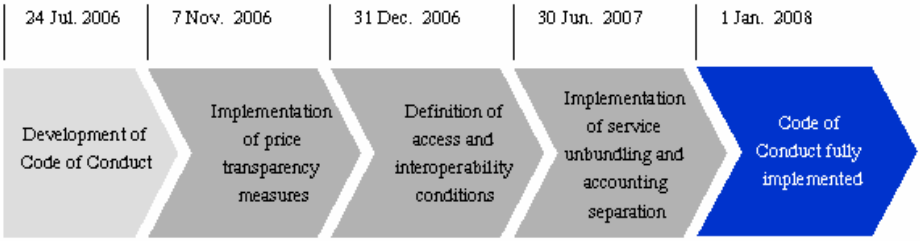


Fig. 2. Implementation of the European code of conduct for clearing and settlement

First Phase of the Code of Conduct

The objective of the first phase is to enable the customers to understand the services they will be provided with and to understand the prices they will have to pay for these services, including discount schemes. The intention is also to facilitate the comparison of prices and services and to enable customers to reconcile ex-post billing of their business flow against the published prices and the services provided [14]. The following principles shall apply to all organisations supporting the code of conduct and all prices the organisations charge, including one-time and periodic fees, prices of transaction-related services, prices of custody services, and prices of additional services to customers. Every organisation agrees to the publication of:

- All offered services and their respective prices including applicable terms and conditions,
- All discount and rebate schemes and the applicable eligibility criteria, and
- Examples that explain prices, as well as discount and rebates schemes for different types of customers.

All information shall be made available on the organisations' websites. The published price lists shall contain all services and prices, a brief description of each service, and

[3] According to Mr. Nava (EC, DG Internal Markets and Services) the code of conduct is a mixture of self-regulation and soft law [13].

the relevant price basis. The organisations are committed to work on further comparability of prices within each layer of the value chain.

Second Phase of the Code of Conduct
The aim of the second phase is to define access and interoperability conditions. The Market in Financial Instrument Directive (MiFID) already grants some access rights in the post-trade area to regulated markets and to investment firms. The code of conduct is not intended to contradict any of those rights. In particular, MiFID grants certain access rights in articles 34 and 46 [14]:

- The right of market participants to remotely access a foreign CCP and/or CSD;
- The right of market participants to choose the settlement location for their trades;
- The right of regulated markets to choose a particular CCP and/or CSD to clear and settle their transactions.

The MiFID has to be applied in November 2007. The second phase of the code of conduct addresses the effective extension of these principles to additional relations in the clearing and settlement sector, addressing mainly the relationships between infrastructures. Organisations from a member state should be able to access organisations in the same or another member state and the responding organisations should provide such access:

- CCPs should be able to access other CCPs;
- CCPs should be able to access CSDs;
- CSDs should be able to access other CSDs;
- CCPs and CSDs should be able to access transaction feeds from trading venues;
- CSDs should be able to access transaction feeds from CCPs;
- A trading venue should be able to access a CSD and/or CCP for its post-trading activities.

Third Phase of the Code of Conduct
The aim of the third phase is to unbundle services and to implement accounting separation. The unbundling of services and accounting separation are important to improve the transparency and efficiency of European capital markets [14]. Service unbundling gives customers flexibility in choosing which services to purchase. Accounting separation grants information on the provided services. The following measures are part of the third phase of code of conduct:

- To realise transparency of the relation between revenues and costs of different services in order to facilitate competition;
- To provide transparency on potential cross-subsidies between the different services;
- To provide users with choices, regarding the services available to purchase.

The prices and services of trading, CCPs, and CSDs shall be unbundled from each other. Furthermore, each CSD needs to unbundle the following services:

- Account provision, establishing securities in book entry form, and asset servicing,
- clearing and settlement (including verification),
- credit provision,
- securities lending and borrowing, and
- collateral management.

Unbundling means that the organisations will allow any customer to purchase an unbundled service without compelling that customer to purchase another service. Each unbundled service will be available at a price applicable to this service.

Any group that includes one or more trading venue, CCP or CSD shall disclose to the national regulators the annual non-consolidated accounts separately upon request from the national regulator. Organisations which offer trading, clearing, and/or settlement services in a single corporate structure shall disclose to the national regulators the costs and revenues of these services separately upon request from the national regulator. Each organisation shall disclose to the national regulators its costs and revenues for each unbundled service in order to make transparent potential cross-subsidies. The organisations will task their external auditors, or another external auditor of the organisation's choice, to verify their compliance with the code. The organisations are committed to ensure adequate monitoring of compliance and are ready to engage in further discussions on the exact mission statement of this committee, on the access to confidential data, and on how the committee will interact with national authorities and regulators, the EC's Clearing and Settlement Advisory and Monitoring Expert Group (CESAME), and market participants [14].

The implementation of the code of conduct is still ongoing. Only the first two phases have been completed. FESE, EACH, and ECSDA have sent a status report on price transparency of their members to the EC. The implementation is managed through task forces of the associations and supervised by the FESE board [16]. The process is also supervised by the CESAME Group.

To analyse the implementation status of the first phase of the code of conduct, we reviewed the organisations' websites on 1st January 2007 and again on 1st February 2007 [15]. It was analysed if the transparency of prices was implemented and the required fees were published on the websites. It was also investigated if examples explaining the fees schedules were published. As the code mentions comparability of prices, in the investigation only fees and pricing examples available in English were considered. This study showed that not all organisations have completed the first phase of the code of conduct in time and published their fees on their websites. 32 CSDs, 11 clearing houses, and 20 exchanges have signed the code of conduct (or a letter of intent). 9 CSDs, 3 clearing houses, and 4 exchanges have not published the information for the first phase in time. On 1st February 2007, still 6 CSDs, 1 clearing house, and 2 exchanges have not published the required information in English. Still 17 CSDs, 5 CCPs, and 7 exchanges have not published examples that explain prices, as well as discount and rebates schemes for different types of customers. Furthermore, structure, format, range, and scope of the published fee schedules differed essentially. For example one CSD charged the same price for every transaction to be settled, not charging any membership fees or fees for technical infrastructure and not considering any discount or rebates. In contrast, another CSD charged fees for different kind of memberships and for technical infrastructure and provides discounts and rebates, which need to be considered if prices are compared. Fundamental knowledge of the local market is necessary for the assessment of costs. At the current status a comparison of the costs for clearing and settlement is only

possible if many assumptions are made. Great efforts of the ECSDA, EACH, and FESE are necessary to improve the comparability of prices.

A study conducted by the European Association of Cooperative Banks (ECSA) User Task Force on the implementation status of the first phase of the code of conduct by CSDs and ICSDs comes to more optimistic results [17]. According to this study in all four analysed areas progress was achieved[4]. In detail in the area of:

- Prices and services: 29 CSDs have achieved a good process and only 3 a partial progress;
- Discounts and rebates: 27 CSDs have achieved a good progress and 5 a partial progress;
- Price comparability: no CSD has achieved a good process and all CSDs achieved a partial progress;
- Billing reconcilability: 23 CSDs have achieved a good progress and 9 a partial progress.

In a first report of the EC to the ministers of finance it says that the code of conduct had a promising start. A great effort has been made by the industry to achieve price transparency. But the EC asks the organisations to improve transparency of services and pricing. The main areas for improvement are seen in the explanation of discounts and in the comparability of prices [18, 19].

Since the 3rd July 2007 the guidelines for access and interoperability have been published. These define principles and conditions for access and interoperability in line with the 2nd phase of the code of conduct. The aim is to offer market participants the freedom to choose their preferred provider of services separately at each layer of the transaction chain [20].

We have analysed public available information such as press releases and articles in the financial press until 15th December 2007. We found out that already 69 requests for access and interoperability have been addressed (see Fig. 3). It remains to be seen what the results of these requests will be.

Market	Belgium			France			Netherlands			Portugal			Germany			Switzerland			UK			Italy		
Recieving Entity / Requesting Entity	Exchange	CCP	CSD	Exchange	CCP	CSD	Exchange	CCP	CSD	Exchange	CCP	CSD	Exchange	CCP	CSD	Exchange	CCP	CSD	Exchange	CCP	CSD	Exchange*	CCP	CSD
Cassa di Compensazione e Garanzia																			X	X	X			
Eurex Clearing				X	X	X	X	X	X							X	X	X	X	X	X			
LCH.Clearnet Ltd	X		X	X		X	X		X	X		X	X	X	X							X	X	X
LCH.Clearnet SA													X	X	X				X	X	X	X	X	X
SIS x-clear	X	X		X	X		X	X		X	X		X	X					X	X	X			
Clearstream Banking Frankfurt		X			X			X			X	X					X			X	X			
Clearstream Banking Luxembourg		X			X			X			X						X		X					
Euroclear Bank														X										
SIS SegaInterSettle													X	X	X									

Key: X New Request; Connection already existent

* virt-x

Fig. 3. Request for access and interoperability

[4] Benchmark was the 31st December 2006.

5 TARGET2-Securities

T2S is a different approach to improve efficiency of European securities settlement, by establishing an integrated technical platform for the settlement of cash and securities, developed and operated by the Eurosystem[5]. The platform would base on the payment systems TARGET and TARGET2.

In 1999, the European System of Central Banks introduced the Trans-European Automated Real-time Gross settlement Express Transfer system (TARGET), a wholesale payment system. TARGET consists of a connection of national payment systems of the European System of Central Banks. With TARGET, European CSDs normally settle the central bank money leg of securities transactions that are denominated in Euro through the local real-time gross settlement component of the TARGET system. The securities accounts managed by the CSDs are linked to the cash accounts managed by the national central banks via a delivery versus payment link. Fig. 4 illustrates the current market infrastructure with TARGET. To facilitate the settlement, the members of a CSD must also participate in the national payment system of the country where the CSD is located [21].

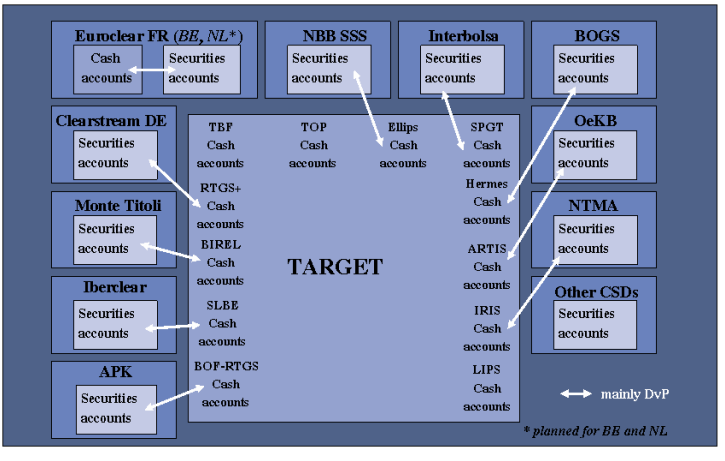


Fig. 4. Current market infrastructure with TARGET [21]

On 24th October 2002, the Governing Council of the ECB took a strategic decision on the direction of the next generation of TARGET. TARGET2 is a multiple platform system based on the principles of harmonisation, a single price structure, cost effectiveness, and no competition among its components. With the introduction of TARGET2 it is possible for each participant to settle, through a single TARGET2 account, transactions effected via any CSD that provides settlement in central bank money in Euro. Additionally, there is the proposal that some CSDs will be able to complete cross-border transactions between themselves without involving TARGET2.

Fig. 5 shows the market infrastructure with TARGET2. TARGET2 was successfully introduced on 19th November 2007 [22].

[5] Consisting of the European Central Bank (ECB) and the national central banks.

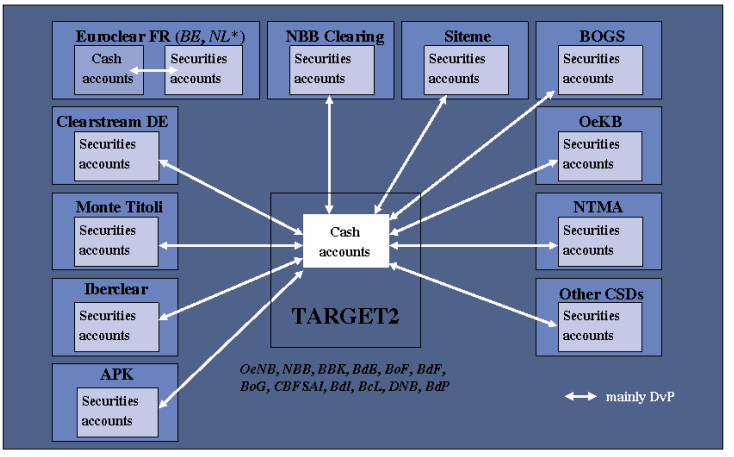


Fig. 5. Market infrastructure with TARGET2 [23]

On 7th July 2006, the ECB issued a press release, stating that the Eurosystem was evaluating opportunities to provide efficient settlement services for securities transactions in central bank money, leading to the processing of both securities and cash settlement on a single platform through common procedures. The platform, called T2S, is the proposal to the CSDs to transfer their securities accounts to a common technical platform. The main benefits of this platform would be the reduction of settlement engines and therefore the reduction of costs for CSD-infrastructure and for custodians' back offices. The ECB announced to explore the setting up of the new service for securities settlement in the Euro area. The background of T2S is the technical debate about the best way to synchronise delivery of securities with payment when settling a transaction in securities. There is general agreement that the most efficient approach for both security and cash movements is to be managed by the same platform. In some countries this process is managed by the securities settlement system, which determines when settlement takes place (e.g. in France and in the UK). As a result, the CSD effectively controls some payments across the books of the central bank: when the CSD determines that a transaction has settled, this causes the money to move on the books of the central bank. In other countries the central bank is unwilling to outsource control of central bank payments to another organisation. To maintain an integrated system, if the CSD cannot manage the money, the central bank has to manage the securities. Then T2S is the only way to reach the integration of the settlement of cash and securities. The settlement of securities and cash would be realised within a single integrated platform. Fig. 6 shows the aimed market infrastructure with T2S.

At the start of every day, participating CSDs would transfer their securities balances and outstanding transactions to T2S. During the day, T2S would settle these transactions and report to the CSDs at the end of the day. One consequence of T2S would be the separation of operation of settlement from the other functions performed by CSDs, such as asset servicing, asset financing, and provision of collateral. These

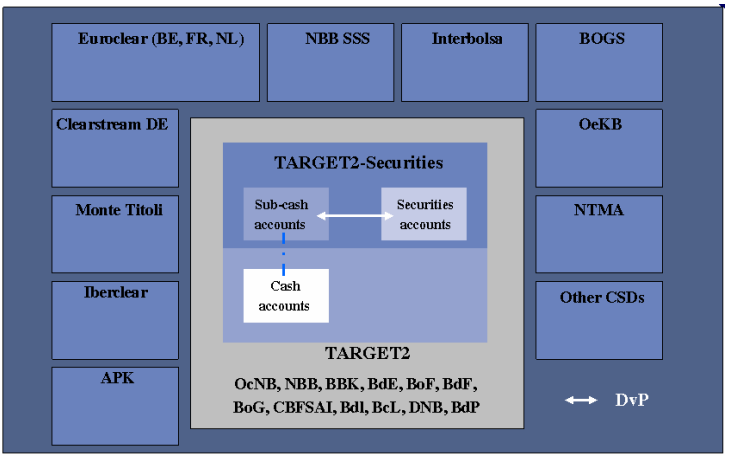


Fig. 6. Market infrastructure with TARGET2-Securities [23]

other functions require access to real-time, intraday information on the securities balances held by participants in the systems, and the ability to control those balances. To realise this, sophisticated linkage between T2S and the systems from the CSDs is required [21].

On 15th January 2007, the ECB presented details on the economic, technical, operational, and legal feasibility of T2S. According to the economic feasibility study, T2S could reduce the average costs for securities settlement to €0.28 per transaction. According to this study, costs for settlement in Europe range between €0.45 and €2.30. The total investment costs for T2S are calculated with €166 million. Additional running costs and costs for telecommunication are calculated with €62 million per year. As investment period for the development, six years are planned [24]. The partaking CSDs will face one-off costs for adapting their systems to transfer settlement functions to T2S. They will still need to maintain and update all information they hold. In addition they will have to maintain an interface to the T2S platform. According to the feasibility report, these additional costs are exceeded by the reduction of operation costs. The platform will also affect custodians who will face one-off costs for system adjustment. These costs are exceeded by the reduction of operation costs [24]. The feasibility report focuses on costs, revenues are not part of the consideration. If revenues are considered, the results from the perspective of CSDs and custodians are not that obvious. For CSDs and custodians settlement represents an important source of revenues.

For the success of T2S the participation of all relevant CSDs is essential. The economic feasibility report assumes that all CSDs in the Euro area participate and thus calculates with a number of 161 million transactions to be settled per year (number of transactions settled by relevant CSDs in 2006). If the participation in T2S is not mandatory, the number of transactions could be significantly lower and the costs per transaction would increase significantly. In the following we present different scenarios for the participation in T2S. One scenario is the aim of the ECB, where all 19 CSDs participate. Beside this, we simulated how the feasibility of the

ECB study would look like, if one of the two largest CSDs[6] or even both would not participate (see Table 1).

In case all CSDs participate in T2S, the ECB calculates the costs per transaction with €0.28. If one of the largest CSDs, Clearstream Banking Frankfurt or the Euroclear Group, would not participate in the system, the costs per transaction would increase to €0.38/€0.36 per transaction. If both would decide not to participate, the costs per transaction would even increase up to €0.60.

Table 1. Costs per transaction of TARGET2-Securities for different scenarios of participation

	All 19 CSDs participating	Clearstream not participating	Euroclear Group not participating	Clearstream and Euroclear Group not participating
Number of annual transactions	160,773,700	110,373,700	116,476,200	66,076,200
Total annual[7] costs in Euro	89,900,000	84,257,286	84,940,513	79,297,800
Costs per transaction in Euro	0.28	0.38	0.36	0.60

On 17th January 2007, T2S was presented to the Financial Services Committee. The settlement platform had the same number of supporters as opponents. Support was indicated by Germany, France, and Finland. England, the Netherlands, and Belgium formed the opposition to T2S. There were a number of details to be clarified, like the supervision of the platform, governance, questions on competition, the effects on the private enterprise infrastructure, and alternatives to integrate the different national infrastructures [25]. On 8th March 2007, the Governing Council of the ECB has concluded that it is feasible to implement T2S and therefore decided to go ahead with the next phase of the project, namely the definition of user requirements on the basis of market contributions. A final decision on the implementation is expected in summer 2008.

6 Analysis of Approaches to Improve the European Cross-Border Securities Settlement

The presented approaches to improve cross-border settlement of securities transactions differ essentially. The code of conduct is a market driven approach, supervised by the EC, while the technical platform T2S represents a centralised

[6] Clearstream Banking Frankfurt had 50,400,000 relevant transactions in 2006 while the Euroclear Group had 44,297,500 transactions in 2006 [24].

[7] The study of the European Central Bank assumes that only telecommunication costs are variable and depending on the number of transactions. The other costs, consisting of costs for infrastructure, costs for application development, and running cost, are seen as fixed costs with a value of €72.2 million per year [24].

approach for securities settlement, provided by the Eurosystem. An EU directive could be another approach, but the EC has favoured the code of conduct, mainly because a directive would have lasted longer and a market driven solution was preferred [11]. These three approaches differ according to the time needed for the implementation, the scope of the projects, the flexibility to adapt to a changing environment, and the legal certainty. With legal certainty the legal implementation is meant. The uncertainty of a potential dispossession of CSDs or of parts of CSDs and the legal uncertainty of the competition with a supranational institution as the Eurosystem were identified as the main concerns.

The code of conduct is an approach that is fast to realise, shows a high legal certainty, and extends clearing, settlement, and parts of trading. The main advantage of the code of conduct is that it is a market driven solution. The main disadvantage is the lag of sanctions in case of malpractice and the possibility of strategic behaviour of the market participants[8].

In contrast, T2S is a centralised technical platform for the settlement of securities. The main benefit is the integration of securities and cash settlement in one single technical platform and thus the reduction of settlement engines and the reduction of technical barriers in European cross-border settlement. The main disadvantages are the long time needed for the implementation and the uncertainty of the participation of the CSDs, and the governance of the platform. T2S would be competing with existing integrated settlement platforms as the Single Settlement Engine of the Euroclear Group.

An EU directive is another possible approach to improve efficiency of clearing and settlement. The process to adopt legislation is often slow. It takes three years on average to agree a directive. Such a timescale is unacceptable when legislation is

Table 2. Analysis of different approaches to improve clearing and settlement in Europe

	European code of conduct	T2S	EU directive
Description	Market driven solution, supervised by EC	Centralised technical platform by Eurosystem	Collective legislative act
Scope	Clearing, settlement, and partly trading	Settlement	Clearing and settlement
Time for implementation	1 year	6 years	> 4 years
Hierarchy	Decentralised (market driven)	Centralised (by ECB)	Centralised or decentralised (implemented by market or central)
Flexibility for adjustment	High	Medium	Low
Legal certainty	High	Medium	High

[8] E.g. a participant could act not conform to the code and force the EC to implement a directive. Due to the missing sanctions the participant would not be punished. The other participant would have invested to act conform to the code conduct and would thus have a disadvantage.

meant to bring an appropriate response to a fast changing world. Because of the long implementation time, the low flexibility to adjustment to a changing environment a directive should only be used, if no market driven solution is available [26]. If self-regulation by the code of conduct is not achieving the wanted results, the directive could be an alternative approach.

Table 2 shows the main differences of these approaches for improving the efficiency.

7 Conclusion and Outlook

In the last years integration and consolidation has taken place, but the European settlement industry is still a fragmented industry, which shows inefficiencies in the cross-border settlement. The introduction of CCPs in Europe has improved the securities settlement significantly by reducing risk and the number of transactions to settle.

The implementation of the European code of conduct for clearing and settlement by the post-trade industry aims to improve the price transparency, interoperability, and service unbundling. The code of conduct is currently being implemented. As of now the first two stages of the code are implemented. This paper shows that the first phase of the code of conduct is not fully implemented by several organisations. Furthermore, it shows that the comparison of the prices of the participants at the current status is not possible due to differences in structure, format, range, and scope.

T2S, as central technical settlement platform for securities could have another essential impact on the settlement of European securities. The platform could help to create an integrated settlement process and remove most of the existing technical barriers and difficulties in cross-border securities settlement. For the success of T2S the participation of all major CSDs is essential for achieving the aimed scale effects. The presented simulation of different scenarios of participation shows, that if not all major CSDs participate in the platform, the costs per transaction of the new platform could be higher that the present domestic prices of some CSDs.

The platform T2S stands in contrast to the efforts of the EC and the industry to bring more competition into the securities markets with the implementation of the MiFID and the adoption of the settlement industry through the code of conduct. The settlement platform would create a monopoly for securities settlement leading into the opposite direction of more competition. It also has to be considered that the settlement industry is making great efforts to improve technical platforms and to increase usage of standards for communication, as the Single Settlement Engine of the Euroclear Group and SWIFT, to remove technical barriers for cross-border settlement.

Acknowledgments. The author gratefully acknowledges the support of the E-Finance Lab, Frankfurt for this work.

References

1. van Cayseele, P., Wuys, C.: Cost Efficiency in the European Securities Settlement and Safekeeping Industry. In: Clearing and settlement of financial markets (2005)
2. de Cavalho, C.: Cross-Border Securities Clearing and Settlement Infrastructure in the European Union as a Prerequisite to Financial Markets Integration: Challenges and Perspectives. HWWA Discussion Paper 287 (2005)

3. Schmiedel, H., Malkamaki, M., Tarkka, J.: Economies of scale and technological development in securities depository and settlement systems. Journal of Banking & Finance 30, 1783–1806 (2006)
4. Giovannini Group: Cross-Border Clearing and Settlement Arrangements in the EU (2001)
5. NERA Economic Consulting: The direct costs of clearing and settlement: an EU-US comparison. Corporation of London City Research Series, vol. 1 (2004)
6. European Commission: Draft working document on Post-Trading (2006)
7. Wendt, F.: Intraday Margining of Central Counterparties: EU Practice and a Theoretical Evaluation of Benefits and Costs. Netherlands Central Bank Working Paper 107 (2006)
8. Börse, D.: The European Post-Trade Market – An Introduction (2005)
9. Lannoo, K.: The ECB should not monopolise securities settlement. CEPS, http://www.ceps.be/wp.php?article_id=555 (accessed, February 8, 2007)
10. Schmiedel, H., Schöneberger, A.: Integration of Securities Market Infrastructure in the Euro Area. European Central Bank 33 (2005)
11. Committee of Wise Men: Final Report of Committee of Wise Man on Regulation of European Securities Markets (2001)
12. Giovannini Group: Second Report on EU Clearing and Settlement Arrangements (2003)
13. European Central Bank: Summary of the 15th Meeting of the Contact Group on Euro Securities Infrastructure (2006)
14. FESE, EACH, and ECSDA: The European Code of Conduct for Clearing and Settlement (2006)
15. Schaper, T.: Trends in European Clearing and Settlement – The European Code of Conduct and TARGET2-Securities. Working Paper 570 (2007)
16. FESE, EACH, and ECSDA: Implementation Progress towards the Code of Conduct (2007)
17. Sabatini, D.: Price transparency: Feedback on implementation by CSDs, ICSDs and CCPs (2007)
18. Börsen-Zeitung: Börsen nehmen Kodex ernst 4 (February 2, 2007)
19. Börsen-Zeitung: EU will transparente Börsenabwickler 3 (April 4, 2007)
20. FESE, EACH, and ECSDA: Access and Interoperability Guidelines (2007)
21. European Central Bank: TARGET2-Securities (2006)
22. Deutsche Bundesbank: TARGET2: Single Shared Platform (2007)
23. Tumpel-Gugerell, G.: TARGET2-Securities: from vision to reality. The Eurosystem's contribution to an integrated securities market. In: EU Commission's Conference on: The EU's new regime for clearing and settlement in Europe (2006)
24. European Central Bank: TARGET2-Securities – Economic Feasibility Study. In: 2nd TARGET2-Securities meeting with market participants (2007)
25. Frühauf, M.: EZB ist bei Abwickungsplattform im Zugzwang. Börsen-Zeitung, 8 (January 25, 2007)
26. Committee of Wise Man. Final Report of Committee of Wise Man on Regulation of European Securities Markets. Brussels: European Commission (2001)

General Requirements of Banks on IT Architectures and the Service-Oriented Architecture Paradigm

Stefan Schulte[1], Nicolas Repp[1], Julian Eckert[1], Rainer Berbner[1], Korbinian von Blanckenburg[2], Ralf Schaarschmidt[3], and Ralf Steinmetz[1]

[1] Multimedia Communications Lab (KOM)
Technische Universität Darmstadt, Germany
schulte@kom.tu-darmstadt.de
Phone: +49-6151-166187;
Fax: +49-6151-166152
[2] Institute of Public Finance
Westfälische Wilhelms-Universität Münster, Germany
12kobl@wiwi.uni-muenster.de
[3] IBM Global Business Services, Germany
ralf.schaarschmidt@de.ibm.com

Abstract. The Service-oriented architecture (SOA) paradigm has been gaining momentum over the last few years. Although the banking industry has often been mentioned as an early adaptor of service-oriented technologies, there is still a lack of knowledge concerning bank requirements on IT architectures and whether an SOA is suited to meet them. In this paper, we present the results from an empirical study which quantifies the qualification of service-oriented technologies for the German banking industry. By using data collected from the German banking industry, it turns out that SOA is apparently suited to meet predefined needs of this industry. However, there are differences in the expectations among various groups of banks. In addition, this paper presents the status quo of SOA adaptations in German banks.

Keywords: Service-oriented Architecture, Banking Industry, Requirements on IT Architectures, IT value creation.

1 Introduction

Although service-orientation is based on well-known concepts like autonomy and the loose coupling of software components, the adaptation of Service-oriented architectures (SOA) in the research community, as well as, in the software industry has been stimulated in recent years by the standardization of Web service technologies. When considering SOA for the banking industry, most of the academic discussions address how SOA can be applied to predetermined domain-specific problems like the management of capital market systems [1,2] or the core banking system [3,4]. Other research approaches focus on quality aspects and other related concepts (e.g., [5]).

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 66–80, 2008.

In contrast, only a few empirical studies (cp. Sect. 2.3) have been conducted regarding to the impact of SOA on the banking industry in general. Therefore, we decided to extend our previous research in the field of SOA by analyzing the German banking industry based on an empirical survey [6].

In this paper, selected results from a survey we conducted among CIOs, CTOs, IT architects, and enterprise architects from Germany's 1001 largest banks (with respect to their balance sheet totals) are presented. The rate of return was 5.19%.

The key focus of this survey was to access whether SOA is a major trend or mere hype for the German banking industry – thereby determining whether SOA is suited to meet the requirements of the German banking industry. Furthermore, the status quo of SOA adaptation in German banks was also investigated. Consequently, the following research question was the key motivation behind the survey:

Is SOA a major trend or hype for the German banking industry?

The remaining part of this paper is structured as follows. In the next section, the basics and common benefits of SOA, its potential impact of SOA on banks, and relevant related work are introduced. Based on this theoretical foundation, nine statements were identified which formed the basis of our questionnaire (Sect. 3). Following this, the results of our empirical study are presented in Sect. 4. Finally, the paper concludes with a summary of the findings in this paper and an outlook of our future work.

2 Theoretical Foundation

In order to answer our research question, it is necessary to define SOA and evaluate its potential role in the banking industry. Hence, we introduce SOA with regard to the banking industry in the following paragraphs. In Sect. 2.1, we briefly explain the SOA paradigm. Subsequently, Sect. 2.2 introduces the potential benefits and effects of SOA on the banking industry. Relevant related work is also included in Sect. 2.3.

2.1 The Service-Oriented Architecture Paradigm

Since the term SOA was coined in 1996 by Gartner [7], several publications have redefined it. In recent years, SOA has often been used synonymously with Web service technologies, even though there are great differences between the actual implementation of an SOA with a certain technology (i.e., Web services) and the underlying concepts which constitute the SOA paradigm. As the term "paradigm" implies, SOA is not a technology but a holistic approach to design an application architecture. By using service-oriented concepts, it is possible to model business processes independent of actual technologies or tools [2].

In order to define the architectural part of SOA, we make use of the following principles [8]:

1. All functions (e.g., business functions) are defined as services.
2. All services are independent and can be used without paying attention to the actual implementation.
3. Services can be accessed by an invokable interface without any knowledge of its location.

Accordingly, an SOA is *"an application architecture 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"* [8].

By using this definition, it is possible to apply the SOA paradigm to an application architecture in general or to map exactly one aspect of a company's business model [9]. Business-oriented services may be mapped to (parts of) business processes, thus allowing new internal and external users to access processes, replace business functions, reorganize processes, or build new business functionalities from existing services.

In order to design services or processes in an SOA, it is necessary to identify and comprehend the business aspects of an organization [10]. Thus, the IT perspective of SOA is strongly related to its business perspective. Otherwise, there is also a strong relationship between the business side and SOA. SOA enhances the agility and flexibility of companies, making it possible to offer new products and services. As a result, business processes might have to be adapted in order to tap the full potential offered by service-oriented technologies [11].

Today, the technology most commonly associated with the implementation of SOA are Web services. With standardized Web service technologies like SOAP [12], WSDL (Web Service Description Language [13]) or UDDI (Universal Description Discovery and Integration [14]) it is possible to apply service-oriented concepts on the Web [15].

2.2 Current State of Research on SOA and the Banking Industry

The banking industry is often recognized as a technology leader in terms of its early utilization of new information technologies (e.g., [2,16]). In the following, we briefly present effects of SOA on banks.

Most commonly mentioned benefits of implementing an SOA include the ability to build agile enterprise systems architectures, which are able to support business flexibility and organizational speed. Adaptation and active application of service-oriented technologies are the foundation for transforming a business model (e.g., by realizing new outsourcing strategies) [9]. Furthermore, service-oriented technologies are considered to be able to solve strategic challenges like application integration, value reconfiguration processes, value preservation after mergers and acquisitions, and enable more agile forms of IS development [17].

Rabhi et al. state that banks especially benefit from the implementation of an SOA due to the reusability of services across several business processes and the ability to provide legacy system functionalities without exposing underlying logics. On the other hand, SOA involves performance drawbacks and requires extra development time due to the need to develop additional service wrappers [2].

Homann et al. dwell especially on the evolution from the formerly monolithic value chain towards a more fragmented value net in which separate activities (i.e., services) have to be fulfilled by specialized entities. Accordingly, information systems have to be coupled in order to enable communication within companies and between an organization and external partners. SOA provides an approach which reduces the complexity and costs of these requirements [18].

Apart from these academic considerations, there are several examples where banks have already implemented an SOA. For example, *Credit Suisse* began deploying service-oriented concepts in 1998 in order to uncouple their platforms and functional groups of applications [19,20]. The primary objective was to reduce the complexity of Credit Suisse's IT ecosystem, thereby increasing its comprehensibility. The system landscape was partitioned into 90 components, hence, instead of administrating a large and complex landscape, it is now possible to manage smaller and less complex components on the one hand and clearly defined interfaces on the other hand.

While Credit Suisse is a large organization with its own internal IT management, a large portion of the German banking industry consists of local savings banks and credit unions which outsource most of their IT infrastructure to data processing service centers established by their umbrella associations. *Sparkassen Informatik* for example, provides such IT services supporting more than 230 German savings banks. As a result of its large number of customers, the application landscape is highly distributed and heterogeneous. In order to tackle the challenges due to the centralized character of the application landscape (e.g., the fast and inexpensive business process integration between Sparkassen Informatik and the savings banks or a highly heterogeneous front end landscape) a Web services-based SOA was implemented in order to meet IT strategy requirements, e.g., minimization of interfaces required, decrease of data transferred, and reduction of development efforts by minimizing interface complexity [3,4].

2.3 Related Work

To the best of our knowledge, this is the first ever survey to analyze the impact of SOA on the German banking industry in detail. However, surveys have been conducted which are related to our study regarding content. *ibi research* (University of Regensburg, Germany) interviewed 21 IT architecture experts from German credit institutions and other related industries (e.g., Deutsche Bank, Credit Suisse, and SAP) in 2005 [21]. While there are certain similarities between this survey and our approach, the focus of these surveys differ especially concerning IT architectures, which is examined in more detail within our study. Furthermore, ibi research interviewed employees of large banks and companies primarily, while we invited CIOs, CTOs, IT architects, and enterprise architects from the 1001 largest (including smaller and medium-sized) German banks to participate in our survey. Major results of ibi research include the following findings [21]: 1. Cost reductions are the primary goal of SOA. 2. SOA is the foundation for an efficient collaboration between business and IT departments. 3. SOA exceeds the

pure technological aspects and includes functional architecture, organizational structure and a governance model.

While these findings helped to describe the current implementation efforts of German financial institutions, the study did not evaluate whether SOA is the most qualified application architecture paradigm to fulfill the requirements of the financial industry.

Another brief analysis of SOA in the German financial industry was carried out by *Fraunhofer IRB* in 2005 [22]. Furthermore, non-academic surveys – e.g., by *Infoworld* and *Gartner* – investigated SOA among other topics, in 2005 and 2006 respectively [23,24].

3 Methodology and Sample Characteristics

Within this section, the methodology used is discussed. In addition, we present the characteristics of the target audience of this survey.

3.1 Methodology

As mentioned in Sect. 2, we identified nine statements which formed the foundation for our questionnaire [25]:

- A company has to adopt an SOA if it wants to stay competitive and achieve continuous growth.
- The active adaptation and use of an SOA enforces innovation in a company's processes and products.
- Flexible and agile business processes are only possible if an SOA is adopted.
- The SOA paradigm is a holistic approach not limited to IT or business only.
- Companies have a need for action regarding SOA.
- The SOA paradigm will have an impact on both the development of custom software and standard software.
- Service-orientation is a critical success factor for future outsourcing activities.
- SOA offers protection of investments.
- The adaptation of service-oriented technologies leads to cost reductions.

For this survey, the 1001 largest banks in Germany were chosen, based on their total assets as reported officially in 2003. All banks were contacted by phone to identify possible participants. This group of contacts included CIOs, CTOs, IT architects, and enterprise architects of the banks selected. We identified one contact per bank. As a large number of banks does not participate in surveys in general and a number of people contacted were not able to participate due to time constraints and other reasons, we invited 288 out of 1001 banks to fill out our online questionnaire made up of 27 questions with more than 120 variables.

In total, 52 analyzable questionnaires were returned. This equals a response rate of 18.06% among the invitees and 5.19% among the 1001 largest banks.

Table 1. Participants

	1001 largest banks		Rate of return	
German banks	Number	Percentage	Number	Percentage
Commercial banks	51	5.09%	5	9.62%
Credit unions	495	49.45%	32	61.54%
Savings banks	455	45.45%	15	28.85%
Total	1001	100%	52	100%

3.2 Sample Characteristics

The German banking industry is divided into three types of banks: commercial banks, credit unions, and savings banks. While credit unions and savings banks often supply their products to customers within a certain geographical area, commercial banks are not limited to any particular region.

As shown in Table 1, the rate of return of commercial banks and credit unions is higher than expected while the rate of return of savings banks is lower than expected. A contingency analysis showed that the distribution of participating banks did not match the distribution in the basic population of the 1001 largest German Banks. Hence, the distribution of the banking groups is not representative (cp. Sect. 4.4).

In order to analyze the data with reference to the previous knowledge of the participants, we asked for a self-assessment regarding the concepts/terms service-orientation and SOA. The evaluation scale spans from "not familiar" to "familiar" on a five-point Likert scale. As shown in Fig. 1, almost 77% of the participants are "to some extent familiar" with the concept of service-orientation,

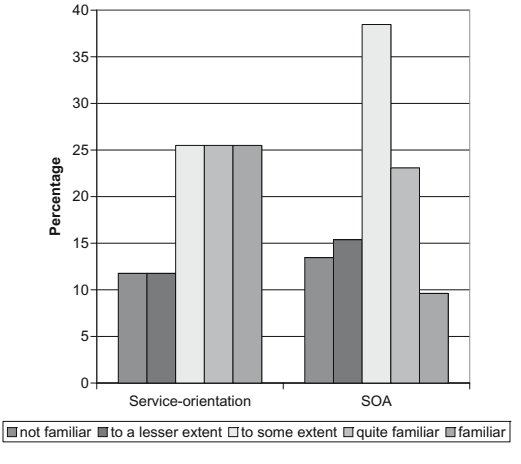


Fig. 1. Familiarity with the concepts Service-orientation (n_1=51, μ_1=3.41, σ_1=1.31) and SOA (n_2=52, μ_2=3.00, σ_2=1.15)

about 51% are even "quite familiar" or "familiar" with this concept. By examining the results of the SOA concept, the figures are similar, with nearly 71% of the participants being familiar with this concept "to some extent". Nearly 32% are "quite familiar" or "familiar" with the concept SOA.

4 Empirical Results

In the following results from our survey are presented. First, the current and future relevance of requirements on IT architectures are assessed in Sect. 4.1. Subsequently, the status quo of SOA adaptation in the German banking industry is presented. Section 4.3 introduces restrictions for the adaptation of SOA. Limitations and the transferability of our results are considered in Sect. 4.4.

Table 2. Classification of criteria

	Criterion is met by an SOA	Criterion is partially met by an SOA	Effect of SOA implementation on criterion is difficult to measure
Flexibility of business processes	X		
Turnover increase			X
Cost savings			X
Reduction of time-to-market		X	
Integration potential	X		
Scalability		X	
Reduction of risks			X

4.1 Assessment of Current/Future Relevance of Requirements on IT Architectures

Seven core requirements of banks on IT architectures were identified by a panel of experts before the actual empirical study. As not all criteria can be met to the same degree by an SOA, the requirements were classified by their feasibility to be met by an SOA (cp. Table 2).

Participants of our study separately assessed the current and future relevance of the seven requirements. The evaluation scale spans from "not relevant" to "most relevant" on a five-point Likert scale.

As shown in Table 3, the criterion *cost savings* has the highest mean of all specified requirements. The relatively low standard deviation of 0.65 (current relevance) and 0.61 (future relevance) establishes the particular relevance of this criterion. More than 90% of the survey participants deem cost savings "relevant" or "most relevant". There are no significant deviations of means between the different groups of banks presented in Sect. 3.2. While the criteria *flexibility of business processes* (μ_1=3.96, μ_2=4.31) and *reduction of risks* (μ_1=3.92, μ_2=4.33) have significant impact, *turnover increase* (μ_1=3.60, μ_2=4.04) and *reduction*

Table 3. Relevance of requirements on IT architectures (n=52)

	Current relevance		Future relevance		Difference	
	μ_1	σ_1	μ_2	σ_2	Δ_μ	Δ_σ
Cost savings	4.35	0.65	4.46	0.61	+0.11	-0.04
Flexibility of business processes	3.96	0.79	4.31	0.73	+0.35	-0.06
Reduction of risks	3.92	0.95	4.33	0.73	+0.41	-0.22
Turnover increase	3.60	0.98	4.04	0.84	+0.44	-0.14
Reduction of time-to-market	3.58	1.00	3.87	1.03	+0.29	+0.03
Scalability	3.40	0.87	3.62	0.93	+0.22	+0.06
Integration potential	3.33	1.00	3.25	1.23	-0.08	+0.23

of time-to-market (μ_1=3.58, μ_2=3.87) are of secondary importance. *Scalability* (μ_1=3.40, μ_2=3.62) and *integration potential* (μ_1=3.33, μ_2=3.25) feature the lowest means amongst the requirements observed.

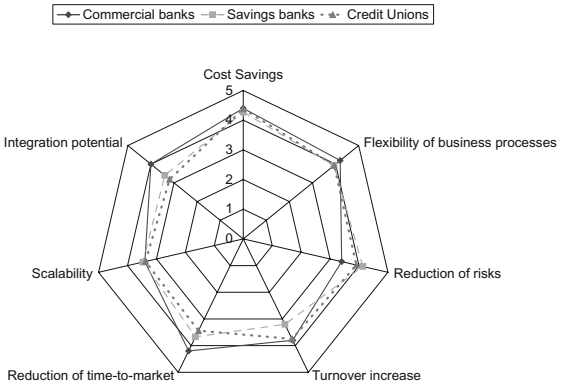
Except for *integration potential*, all means observed are significant deviations from the expected value "to some extent relevant (3)" at the 0.01 level. The mean of *integration potential* is a significant deviation at the 0.05 level for the current relevance, but there is no significant deviation from the expected value for the future relevance of this requirement.

When comparing the results of this part of our survey with the classification presented in Table 2, the results have to be analyzed critically. *Integration potential*, which is one of the both criteria presented that is definitely met by an SOA, is rated as less relevant than any other criterion. The criteria *reduction of time-to-market* and *scalability*, which are partially met by an SOA, are relatively unimportant, too.

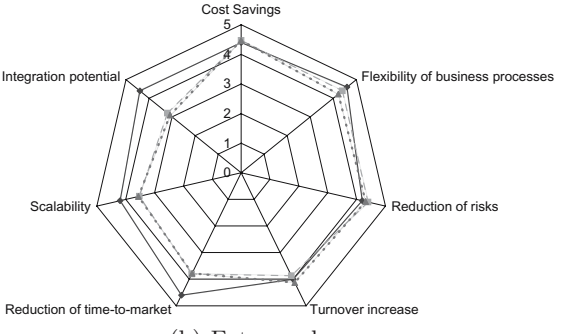
Beside *integration potential*, *flexibility of business processes* is the second requirement definitely met by an SOA within the context of this study. Only the requirement *cost savings* is rated more relevant than *flexibility of business processes*. This shows that a large part of the banking industry attaches great importance to flexible business processes. Consequently, this requirement has to be met by an application architecture.

Otherwise, *integration potential* is rated as more relevant by commercial banks than by savings banks and credit unions, with means of 4.00 (current relevance) and 4.40 (future relevance). The mean of future relevance has a significant deviation at the 0.05 level from the mean of all savings banks and credit unions (3.13). Regarding the future relevance, *reduction of time-to-market* is also assessed more relevant by commercial banks (mean of 4.60) than by all other banks (mean of 3.79) with a significant deviation at the 0.10 level. The attitude of different banking groups towards individual requirements is depicted in Fig. 2.

The higher need for *integration* in commercial banks might be explained by the different IT strategies of savings banks and credit unions. While last-mentioned groups of banks often outsource most of their IT department to data processing service centers established by their umbrella associations, commercial banks mostly operate their own IT infrastructure and do not outsource to the same



(a) Current relevance



(b) Future relevance

Fig. 2. Relevance of requirements on IT architectures

extent. However, outsourcing is part of the IT strategy of commercial banks, too.

Therefore, IT architecture experts from commercial banks have an increased recognition for the need to integrate (legacy and other) systems. Furthermore, *integration potential* is especially needed if a bank has to deal with mergers and acquisitions. This explains the higher requirement for integration in commercial banks, too, as mergers and acquisitions are rather a characteristic of the business model of commercial banks than of credit unions or savings banks.

4.2 Status Quo of SOA Adaptation in the German Banking Industry

Apart from general requirements on IT architectures and reasons against SOA adaptation (cp. Sect. 4.3), we examined the current status of SOA implementations in the German banking industry. The participants were requested to estimate the progress of the adaptation of SOA in their company. As shown in Fig. 2, 30.77% of the participants state that their organizations *plan an SOA*

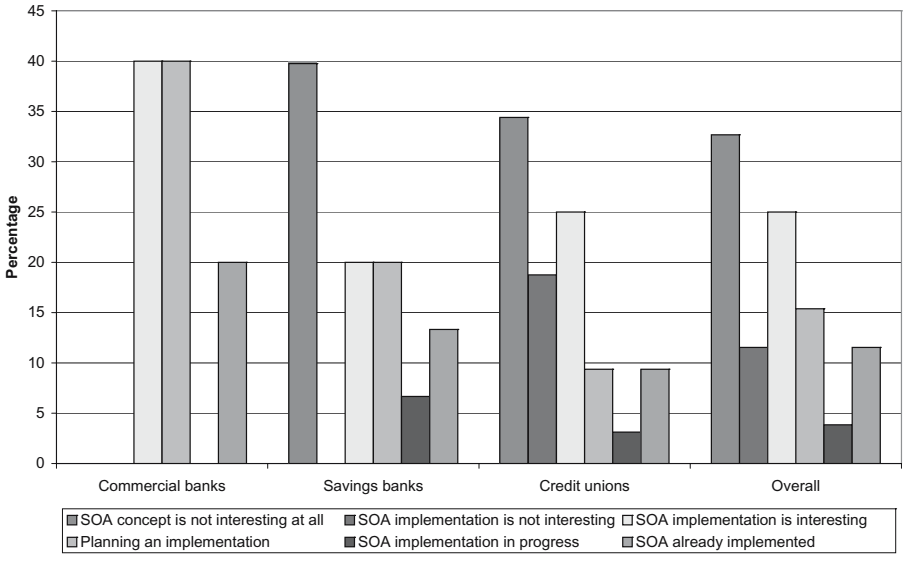


Fig. 3. Status quo of SOA adaptation in the German banking industry

implementation, *are currently implementing an SOA* or *have already implemented an SOA*. Further 25% *consider an implementation interesting*.

Although the acceptance of SOA in the surveyed banks seems to be surprisingly high with nearly one third of the banks bringing an SOA into operation or already operating an SOA, it should be taken into account that the propagation of service-oriented technologies in the German Banking Industry is promoted either by external parties (e.g., data processing service centers or consulting companies) or by internal IT departments. For example, Sparkassen Informatik (see before) supports one tenth of the German banking industry (or half of all savings banks) with its Web services-based SOA [3,4]. Otherwise, it should be noted that only 13.33% of the participating savings banks claim that they already have an SOA in production.

This data conflicts with the assumptions we made earlier and should be further observed in the future. One possible explanation for this discrepancy between estimated and actual values could be the low visibility of technologies in use in the data processing service centers. Nevertheless, the results show the extensive distribution of SOA in German banks with 11.54% of all participating banks operating an SOA.

Concerning the different banking groups, our presumptions from Sect. 4.1 regarding the higher need for SOA-related requirements in commercial banks are confirmed (cp. Fig. 3). 60% of the commercial banks *plan an SOA implementation* or *have already implemented an SOA*. The percentage of credit unions planning or currently performing an SOA implementation or running an SOA is considerably lower (21.88%). With 40%, the percentage of savings banks ranges between the values for the other banking groups.

4.3 Restrictions for the Adaptation of SOA

While the results from Sect. 4.1 and Sect. 4.2 indicate that SOA meets the requirements of German banks and is already well distributed in this sector, a large percentage of banks also have no plans to or interest in adapting service-oriented technologies at the time of the survey.

Prior to the actual empirical study, the following possible reasons for the lack of interest in SOA were identified by a panel of experts:

- SOA-technologies are not standardized
- Lack of experience in the SOA field
- Short-/medium-term benefits are not identifiable/measurable
- Insufficient budget
- Assumptions about the future market environment do not apply
- Organizational restrictions
- Strategic restrictions

Table 4. Restrictions for the adaptation of SOA

	Overall n=52		Commercial banks		Savings banks		Credit unions	
	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4
Lack of experience	3.52	1.13	3.00	1.22	3.20	1.37	3.75	0.95
Organizational restrictions	3.23	0.90	3.60	0.89	2.93	0.88	3.31	0.90
Technologies are not standardized	3.13	1.01	2.80	0.84	3.27	1.10	3.13	1.01
Insufficient budget	3.06	0.98	2.60	0.55	2.67	0.90	3.31	1.00
Short-/medium-term benefits are not identifiable	3.04	1.01	3.00	1.41	2.73	0.88	3.19	1.00
Strategic restrictions	2.94	0.89	2.60	0.89	2.67	0.90	3.13	0.87
Assumptions about the future market environment do not apply	2.62	0.93	2.00	1.00	2.33	0.82	2.84	0.92

The participants of our study assessed these cases on a five-point Likert scale spanning from “does not apply” to “applies”.

As shown in Table 4, *lack of experience in the SOA field* has the highest mean (μ_1=3.52) of the identified obstacles. *Organizational restrictions* (μ_1=3.23) and the assumption, that *SOA-technologies are not standardized* (μ_1=3.13) show relatively high means, while *insu cient budget* (μ_1=3.06) and *non-identifiable short-/medium-term benefits* (μ_1=3.04) are of secondary importance. *Strategic restrictions* (μ_1=2.94) and *incorrect assumptions about the future market environment* (μ_1=2.62) feature the least means of all reasons identified.

The observed means for *lack of experience* and *incorrect assumptions about the future market environment* are significant deviations from the expected value “neutral (3)” at the 0.01 level, the mean of *organizational restrictions* is a significant deviation at the 0.10 level. There is no significant deviation for the means of the remaining criteria.

IT experts in the field of SOA (i.e., participants that considered themselves to be familiar with SOA in Sect. 3.2) assess the reasons of *non-standardized technologies* (mean of 2.40) and *non-identifiable short-/medium-term benefits* (mean of 2.00) significantly lower at the 0.05 (*technologies*) respectively 0.10 (*benefits*) level than all other participants (means of 3.21 and 3.15).

The different banking groups also assessed the particular reasons very distinctly. Credit unions tend to rate most criteria higher than savings banks, while commercial banks tend to evaluate most criteria lower. It should be noted that *organizational restrictions* are exceptional as commercial banks rate them higher than any other banking groups. Furthermore, this criterion possesses the highest mean of all criteria for commercial banks (cp. Table 4).

As a result, the means of credit unions for the reasons of *non-standardized technologies* and *incorrect assumptions about the future market environment* are significantly higher at the 0.05 level compared to other banks. Furthermore, the reasons *lack of experience* and *insu cient budget* are significantly higher at the 0.10 level. Especially the *lack of experience in the SOA field* confirms the finding from Sect. 4.2 that SOA is less distributed in credit unions compared to other banking groups.

4.4 Limitations and Transferability of Results

There are two limitations on the research presented. First, the survey data was collected at a specific point of time (i.e., at the end of 2006 and the beginning of 2007) and therefore provides a snapshot perspective. Future rollouts of this survey will show the development of SOA in the banking industry and if the SOA adaptations planned in banks have been realized.

Furthermore, generalization made from the data collected is limited due to the low response rate of 5.19% from the 1001 banks in scope. Besides, the distribution of participating banks within the banking groups does not match the distribution of banking groups for the 1001 largest German banks (as mentioned above).

However, we still believe that the results are able to provide insight into application architectures within the German banking industry. This industry is largely characterized by its organization into three banking groups, but continues undergoing massive changes due to regulations and adjustments in the market. As we analyzed the results within each banking group, it is possible to transfer the results for commercial banks to countries that are more characterized by, e.g., commercial banks. Nevertheless, the results are only indicative for other industries or countries.

5 Summary and Future Work

In this paper, we have presented results from a survey we conducted among CIOs, CTOs, IT architects, and enterprise architects from Germany's 1001 largest banks. The goal of this survey was to identify whether SOA is regarded as a major trend or mere hype within the German banking sector.

It has been shown that the criteria requested by banks are difficult to measure, i.e., *cost savings*, *reduction of risks* or *turnover increase*. *Flexibility of business processes*, which is a typical characteristic of SOA, is also requested. Otherwise, SOA-typical requirements like *integration potential* and *reduction of time-to-market* are requested by commercial banks in particular, while these criteria only play a minor role for the German banking industry by large. This indicates that SOA is especially suited to fulfill the requirements of commercial banks. This assumption is further supported by the fact that SOA is already more widespread in commercial banks compared to credit unions. Furthermore, credit unions tend to assess reasons against SOA adaptation more strongly than all other banks.

Therefore, we interpret that SOA is already *more than just a hype* for commercial banks.

However, SOA is not relevant for every banking group to the same extent. While commercial banks seem to be technology leaders [2,16], especially credit unions are currently not adopting the SOA paradigm. It is most likely that the divergences between the banking groups result from the different expectations these groups have towards an IT architecture. Credit unions and savings banks often do not operate an extensive application architecture. Hence, these banks do not have a need for *integration potential* or *scalability* of a particular IT architectur paradigm. This assumption is supported by the presented data. As a result, in order to answer the question if SOA is a major trend or hype for the *whole* German banking industry, it should be investigated in what way credit unions and savings banks might benefit from the SOA paradigm.

It seems reasonable that SOA-related projects have already been set up or will be initiated in credit unions (and savings banks) by the service organizations established by their umbrella associations. In order to investigate the propagation of service-oriented technologies in these banking groups, we will set up a multi-participant case study in cooperation with the E-Finance Lab e. V. and IBM. This case study addresses the service organizations of both the credit unions and savings banks. The goal of this case study is to analyze the different strategies that service organizations follow while implementing SOA, as well as the impact of SOA implementation on affiliated credit unions and savings banks.

As aforementioned, the survey presented provides a snapshot perspective. Therefore, we intend to conduct further surveys in the following years in order to show the development of SOA within the German banking industry. While the intention of the presented survey was to get an overview of SOA in the German banking industry, future surveys and case studies will have a different focus, e.g., the actual value creation of SOA.

Acknowledgements

This work is supported in part by the E-Finance Lab e. V., Frankfurt am Main, Germany (`www.efinancelab.com`).

References

1. Rabhi, F.A., Benatallah, B.: An Integrated Service Architecture for Managing Capital Market Systems. IEEE Network 16, 15–19 (2002)
2. Rabhi, F.A., Yu, H., Dabous, F.T., Wu, S.Y.: A service-oriented architecture for financial business processes. Information Systems and E-Business Management 5, 185–200 (2007)
3. Brandner, M., Craes, M., Oellermann, F., Zimmermann, O.: Web services-oriented architecture in production in the finance industry. Informatik Spektrum 27, 136–145 (2004)
4. Zimmermann, O., Milinski, S., Craes, M., Oellermann, F.: Second Generation Web Services-Oriented Architecture in Production in the Finance Industry. In: Companion to the 19th Annual ACM SIGPLAN Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA 2004), pp. 283–289. ACM Press, New York (2004)
5. Berbner, R., Grollius, T., Repp, N., Heckmann, O., Ortner, E., Steinmetz, R.: An approach for Management of Service-oriented Architecture (SoA)-based Application Systems. In: Proceedings of the Workshop Enterprise Modelling and Information Systems Architectures (EMISA 2005), Klagenfurt, Austria, pp. 208–221 (2005)
6. Repp, N., Schulte, S., Eckert, J., Berbner, R., Steinmetz, R.: An Approach to the Analysis and Evaluation of an Enterprise Service Ecosystem. In: Proceedings of the 1st International Workshop on Architectures, Concepts and Technologies for Service Oriented Computing (ACT4SOC 2007), Barcelona, Spain, pp. 42–51 (2007)
7. Schulte, R.W., Natis, Y.V.: "Service-Oriented Architectures", Part 1, SPA-401-068. Gartner Group (1996)
8. Channabasavaiah, K., Holley, K., Tuggle Jr., E.: Migrating to a service-oriented architecture, Part 1. IBM DeveloperWorks (2003), http://www-128.ibm.com/developerworks/library/ws-migratesoa/ (accessed at 2007-08-19)
9. Krafzig, D., Banke, K., Slama, D.: Enterprise SOA: Service-Oriented Architecture Best Practices (The Coad Series). Prentice Hall PTR, Upper Saddle River (2004)
10. Cherbakov, L., Galambos, G., Harishankar, R., Kalyana, S., Rackham, G.: Impact of service orientation at the business level. IBM Systems Journal 44, 653–667 (2005)
11. Woods, D., Mattern, T.: Enterprise SOA: Designing IT for Business Innovation. O'Reilly Media, Inc., USA (2006)
12. Mitra, N., Lafon, Y.: SOAP Version 1.2 Part 0: Primer (2nd edn.). W3C Recommendation (2007), http://www.w3.org/TR/2007/REC-soap12-part0-20070427/ (accessed at 2007-08-18)
13. Booth, D., Liu, C. K.: Web Services Description Language (WSDL) Version 2.0 Part 0: Primer. W3C Recommendation (2007), http://www.w3.org/TR/2007/REC-wsdl20-primer-20070626 (accessed at 2007-08-18)
14. Clement, L., Hately, A., von Riegen, C., Rogers, T. (eds.): UDDI Version 3.0.2 – UDDI Spec Technical Committee Draft. OASIS Standard (2004), http://uddi.org/pubs/uddi-v3.0.2-20041019.htm (accessed at 2007-08-16)
15. Papazoglou, M.P., Georgakopoulos, D.: Service-Oriented Computing. Communications of the ACM 46, 25–28 (2003)

16. van Hillegersberg, J., van Oosterhout, M., Valkenier, R., Waarts, E.: Business Agility Requirements in Financial Services: Implications for IT Architectures. In: Second International Workshop on Enterprise, Applications and Services in the Finance Industry (FinanceCom 2005), Regensburg, Germany (2005)
17. 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: Proceedings of the Thirteenth European Conference on Information Systems (ECIS 2005), Regensburg, Germany, pp. 761–772 (2005)
18. Homann, U., Rill, M., Wimmer, A.: Flexible Value Structures In Banking. Communications of the ACM 47, 34–36 (2004)
19. Heutschi, R., Schemm, J. W.: Fallstudie: Serviceorientierte Architektur bei der Credit Suisse. St. Gallen: Institut für Wirtschaftsinformatik, Universität St. Gallen (2005), http://www.alexandria.unisg.ch/publications/29504 (accessed at 2007-08-13)
20. Ploom, T., Kurmann, A., Hagen, C.: Kombination der MDA und SOA als Mittel zur IT-Komplexitätsreduktion bei Credit Suisse. In: MDD, SOA und IT-Management (MSI 2007), Workshop, Oldenburg, Germany (2007)
21. Koch, M., Rill, M.: Serviceorientierte Architekturen bei Finanzdienstleistern. ibi research an der Universität Regensburg GmbH, Regensburg, Germany (2005)
22. Spath, D. (ed.), Engstler, M., Praeg, C.-P., Vocke, C.: Trendstudie Bank & Zukunft 2006 – Wettbewerbsfähigkeit durch Innovationen im Vertrieb und industrialisierte Prozesse. Stuttgart, Germany, Fraunhofer IRB (2006)
23. Gartner Group: User Survey Analysis: SOA, Web Services and Web 2.0 User Adoption Trends and Recommendations for Software Vendors, North America and Europe, 2005-2006 (2007)
24. InfoWorld Market Research: An Overview of the SOA Market, March 2005. Conducted by IDG Research Service (2005)
25. Schulte, S., Repp, N., Berbner, R., Steinmetz, R., Schaarschmidt, R.: Service-oriented Architecture Paradigm: Major Trend or Hype for the German Banking Industry? In: 13th Americas Conference on Information Systems (AMCIS 2007), Keystone, Colorado, USA (2007)

IT Capabilities and Organizational Change: Digging Deeper into the Banking Industry

Marco Marabelli and Federico Rajola

Catholic University of Milan, Italy
CeTIF – Research Center of Technology, Innovation and Finance
Largo Gemelli, 1 – 20123 Milan, Italy
Marco.Marabelli@unicatt.it, Federico.Rajola@unicatt.it

Abstract. The importance for organizational performance of aligning IT and organizational change is well recognized in the empirical literature [1], [2], [3] and there are many theoretical approaches that focus on this subject. In this study we use an emergent perspective [3], [4], [5] to show how IT capabilities can help management in strategic planning involving organizational change. The organizational change is studied according to process-level research [6] and the findings describe how IT capabilities can drive this change. The context for our study is the Italian financial industry in its bid to be compliant with the MiFID directive. The method combines quantitative (questionnaire) and qualitative (focus groups) analysis to achieve reliable evidence and results. The sample is composed of 37 Italian financial institutions and the study focuses on theoretical and empirical work.

Keywords: IT capabilities, organizational change, strategy, banking industry, financial markets, MiFID.

1 Introduction

This paper, which contributes to the literature in information technology (IT) studies and organizations [7], aims to analyze organizational change in the Italian finance industry, related to IT capabilities. In recent years this industry has become very competitive or, to use Jeffrey Williams's expression [8], has become a "Schumpeterian market". The occurrence of mergers and acquisitions (M&A) provides strong evidence of the rapid process of internationalization that has taken place in European banks since the early 1990s. The need to be ever more competitive has driven the European Union (EU) to standardize some of the more typical financial services in order to compete with US and Asian banking structures. The institutional "tool" typically applied is the EU directive. These changes in Europe's financial systems have had and are having obvious consequences in terms of organizational changes in processes, practices and business. This article focuses on these changes. In particular, we investigate what has been or what might be the contribution of IT to the activity of planning organizational change and determining a long run strategy. The need for banks to react quickly to the changes

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 81–96, 2008.

in the environment and the competition reinforces the importance of IT in their bid to maintain competitiveness in a complex and dynamic scenario.

1.1 The Scenario

Since 2000 the European financial system has undergone major structural changes as a result of efforts to harmonize national currencies (introduction of the Euro), credit rates according to Basilea II, new payment system (SEPA – Single Euro Payment Area) and financial markets (MiFID – Market in Financial System). Each of these changes has had a different impact depending on the types of organizations involved, on the way that particular financial institutions have chosen to manage the change, and on the strategic value they have imputed to these changes. In this paper we examine the implications of MiFID for the financial sector from an organizational perspective. We aim to study how organizations are managing the process implications of MiFID in the new dispositions of the directive, which involve several functions and departments (e.g. information systems, trade, compliance, risk management, etc.).

The main effects of the introduction of MiFID have been felt in the financial market system and inside-institution trading activity.

In terms of the financial market system, we can identify three elements that represent a break with the old trading system: a) the introduction of different market venues, such as MTFs – Multilateral Trading Facilities - and internalizers, that can act as "official venues" for exchanging financial instruments; b) the end of the "concentration rule": the directive gives all financial institutions the opportunity to trade financial instruments directly, with no need to involve the national "official" stock exchanges; c) the need for *pre trade* (quotations availability in the different venues) and *post trade* (execution price) transparency of the negotiation. The objective of this practice is to avoid enfeeblement of the price discovery, in order to shelter the investors, especially retail investors.

The aim of b) is to create a freely integrated infrastructure of financial instruments exchange. One of the most important effects is that banks and traders will be allowed to choose from a panel of preferred trading venues to which they can direct orders. However, because of positive network externalities, the macro-economic scenario changes only slowly [9], [10], [11]. Externalities are the costs or benefits arising from an economic transaction that accrue to a third party not directly involved in the transaction. In financial markets the concentration of such transactions produces a benefit for users who can take advantage of economies of scale. In fact, the more that financial instruments are exchanged at one concentration point (e.g. a stock exchange), the more that scale decreases the costs of each transaction [12].

The paper is structured as follows: Section 2 presents the theoretical assumptions and hypotheses of this research. Section 3 describes the methodology, types of analysis, and data correlations. Section 4 focuses on correlation and analysis of the data. Section 5 provides a discussion of the questionnaire results, illustrated by comments from members of the focus group. It concludes by outlining the contribution to the literature, areas for future research and the limitations of the present study.

2 Theoretical Framework and Research Questions

2.1 Background

Following Leavitt and Whistler's perspective [13] many studies have tried to relate and measure IT [14] and organizational change [15] and there are many discussions in the literature on their alignment. Organizational performance based on IT is well recognized in literature ([1], [2], [3] and the importance of strategic planning and the role of IT in making it effective has been highlighted by several scholars [16], [17]. Also IT capabilities are well known to be drivers of innovation [18], [19], which provide major contributions to the managing of ambidextrous organizations [20], [21], [22] in turbulent environments, where organizations need flexibility and agility in order to react quickly to fast-changing markets/industries [23], [24], [25], [26].

Markus and Robey [27] examined IT and organizational change in studying the causal structure of organizational change. Theories have become categorized into those privileging the technology (technological determinism, discrete entity tool view, autonomous technology), those favoring the social (strategic choice view, web-based ensemble models), and more middle ground approaches that take account of socio-technical and emergent perspectives [3], [28], [29], [30], [4], [31]. In this paper we study how IT can influence organizational change according to the emergent perspective, [32], [3], [4] looking at both technology and human resources in information systems. We consider IT capabilities, which we define according to Bharadwaj et al. [1999:379], as a "multidimensional construct encompassing both the technical and organizational dimensions". In detail, and again following Sambamurthy and Zmud [33], we contend that IT capabilities are combinations of IT-based assets and routines that support the conduct of business in value-adding ways [34], [35], [36].

The operationalization of this (dependent) variable also follows Sambamurthy and Zmud [33], and we measure IT capabilities by looking at a) value innovation, b) knowledge work improvements, c) IT-enabled business performance, d) operational excellence, e) value-chain extension, and f) solution delivery. We understand IT capabilities as being comprised of highly interdependent core assets and routines that take on distinctive profiles in their situational execution [37].

In measuring IT capabilities we consider "strategic options" (see below) from MiFID (independent variable), which involve organizational change.

Mintzberg [38] identifies three patterns of strategy formulation: deliberate, emergent and exercised strategy. In this study we look at deliberate strategy. We investigate the projects designed by managers planning process changes, compliant actions and re-organization designs in order to align IT resources and new activities [39], keeping in mind the operative units and the business needs [40], [41].

The context we analyzed (MiFID directive, i.e. a modification of the trade process) provides scenarios, identified as strategic options from which the banks can choose. The impacts on organizational processes and information systems depend on the type of change chosen. According to Anolli and Petrella [42], validated by the interviews with banks in the first focus group, we can identify four scenarios:

- *broker*, a firm that passes orders exchanging financial instruments via primary or secondary markets;
- *systematic internalizer*, i.e. an investment firm, which, on an organized, frequent and systematic basis, deals on its own account, executing client orders outside a Regulated Market or an MTF;
- *MTF collaboration*[1] *(joins an existing MTF)*;
- *MTF (creates own MTF).*

Each choice has a consequent organizational change on trade processes (high or low organizational impact - measurement is described later in this paper). Our aim was to verify that the four strategic options represent an increase in the intensity of change in organizational process and IT infrastructure, starting from the first (slight change) increasing to the highest level (creating own MTF), which has all the characteristics of the first three types (the third level - join existing MTF - involves all the characteristics of the first two; the second level involves all characteristics of the first).

To measure the intensity of change in processes we use existing theory and two different process definitions can be found in the literature: (1) a category of concepts or variables that pertains to actions and activities, and (2) a narrative describing how things develop and change ([43]. Within the first definition, process is typically associated with a "variance theory" [44] of change, in which a set of independent variables statistically explains variations in some outcome criteria (dependent variables). Within the second definition, process is often associated with a "process theory" explanation of the temporal order and sequence in which a discrete set of change events occurs, based on a story or historical narrative ([45], [46], [47]. We can examine the entity of change at four levels: low, medium, high and very high. The operational variables considered are: a) number of tasks involved in the change; b) level of change in each task; c) how trade process (critical in this study) is modified in the number of people employed, extension of the process, importance and role compared to other contiguous processes, and elements that impact on it.

In this research we analyze organizational change in the trade process occurring within a bank. The main explanatory variables analyzed are organizational change and IT capabilities. We suppose the first to be dependent on the second.

In order to measure the real connection between IT capabilities and organizational change – and in order to verify not only the alignment, but also the role of IT as a "driver"– we consider two independent variables that correlate with IT capabilities. These are size of institution and relevance of trading in financial instruments for the institution.

Size (S) is measured according to the Bank of Italy classification. The Bank of Italy is the central national bank, whose main functions concern banking and financial supervision; its objective is to ensure the stability and efficiency of the system and compliance with the various banking rules and regulations (MiFID, for instance). This is achieved by the bank through secondary legislation, controls and cooperation with governmental authorities. The Bank of Italy classifies Italian banks into major, big,

[1] MTFs are regulated market and multilateral systems They can be operated by the operator of a Regulated Market or an investment firm. MTFs do not have the same obligations concerning financial instruments in their quotations, and there are no specific requirements regarding authorization for the operation of an MTF.

medium and small institutions based on the average amounts of funds exchanged annually.

Business relevance of trading in financial instruments (T) is operationalized in six items: a) number of people employed in the trading process; b) number of operations (executed orders) in a particular time unit –1 year; c) volume of total orders executed; d) number of clients in trading area; e) growth of trade in financial instruments business in previous 5 years; and f) percentage of financial instruments traded in Italian banking industry.[2] All items were researched via the questionnaire (see method, *infra*). The data were collected from the responses to 12 questions, grouped in six categories. (See data analysis in section 4).

The assumptions made in this analysis in terms of the two independent variables mentioned above, are that a bank's organizational size (S) and relevance of trading in financial instruments (T) both, are positively related to the availability of the resources that a firm can allocate to organizational changes. In other words, the bigger the organization, the more resources it can make available to accomplish any change, and the more important trading is for the organization, the more it can invest in its "core business".

Consequently, we show that IT capabilities can drive strategic planning involving organizational change.

Consistent with these statements, the main research questions are:

- *Is organizational change related to higher or lower levels of IT capabilities in organizations?*
- *Is organizational change driven by IT capabilities? In other words, a) is there a relationship of cause and effect between IT and organizational change; and b) is IT capabilities the independent variable that most conditions strategic planning involving organizational change?*

Our aim is to verify synthetically the existence (or not) of interactions between IT capabilities (IC) and strategic organizational change (OC), and compare them to the interactions between IT capabilities and organizational size and trading in financial instruments (S, T). This allows us to correlate the variables and formulate our first hypothesis:

Hypothesis 1: IT capabilities and strategic planning of organizational change are positively related: the higher the IT capabilities, the stronger will be the change.

The second hypothesis refers to the relation between IT capabilities (dependent) and the independent variables (S) and (T).

The strategic choices of decision makers and the contribution of the main activities of a firm and their directions are more disposed to giving funds and resources to units that contribute more to general organizational performance.

We want to find whether the size (S) and the trading (T) variables are related to IT organizational change and if so, how much or how intensely. In other words, we want to discover whether strategic choice is managed by one or both of the above mentioned variables.

[2] Source: Bank of Italy.

If this second hypothesis is confirmed then we can say that, in this context (implementing MiFID), IT capabilities drive strategy and generates OC measured at the level of IT capabilities, in other words:

Hypothesis 2: a) Size (S) and b) trading in financial instruments relevant to a bank (T) are not related (or are only weakly related) to the organizational change strategy.

3 Methodology

Field of the research

The research field is the financial industry. We had data on 40 financial institutions, 37 of which agreed to take part in the study. Our methodology includes a questionnaire and two focus groups, both of which tools are described later.

The total amount of traded financial instruments represented by the institutions in the sample is the 54.5% of the total volume in Italy. The data are based on the average value of the financial instruments exchanged (46%) and the number of transactions (63%). The high number of operations compared to the amount of value exchanged is because, within our sample, most of the institutions (80%) are retail banks, so the amount-per-transaction is low and the number of transactions is high. We did not deliberately choose retail institutions over other kinds of financial institutions, but retail banks seem to have higher impacts in organization and process within the MiFID due to the need to respect best execution dispositions[3] (art. 21, Directive 2004/39/CE; art 44, Directive 2006/73/CE). Best execution dispositions must be respected when brokering with small clients (retail and professional clients), but not when brokering with institutions, referred to in the Directive as eligible counterparties (art. 50 Directive 2006/73/CE).

The survey participants were mainly IT professionals, trading officers and compliance officers[4]. All are responsible for decisions related to organizational change in implementing MiFID and within their firms have the greatest influence on investment directions and strategic choices (see further the "MiFID scenarios").

Analysis tools

The data are derived from the responses to a semi structured questionnaire and notes taken at the three focus groups the first of which was held before the questionnaire was administered.

1) *Questionnaire*: the 12-page semi-structured survey included 45 questions, 36 of which were multiple choice questions. The questionnaire provided a brief description of the main research activities of our research center in order to familiarize the

[3] Best execution (BE) can be defined as follow: In order to give effect to that policy (BE rules), an investment firm should select the execution venues that enable it to obtain on a consistent basis the best possible result for the execution of client orders. An investment firm should apply that policy with a view to obtaining the best possible result in terms of the total consideration, representing the price of the financial instrument and the cost related to execution (art 44, Section 2, Title II).

[4] Compliance officers are those individuals responsible for implementing new laws/directives of the EU dispositions in financial institutions.

respondents with our approach, to inform them that we are a university research center interested in scientific methods for eliciting evidence that can be generalized. The scale for the multiple choice questions was 1 to 5; nine questions were controls to validate responses on the aspect of coherence. The tool we used to decide about the level of consistency of the responses was a pre-designed control decision tree system. We rejected questionnaires where up to two of the nine control questions were unanswered. All questionnaires with responses to all nine control questions were included in our analysis.

The questionnaire had three sections: a) a first group of questions designed to extract personal data from the respondent and also information about the firm (e.g. institution, function in the institution, type of firm – retail bank, investment bank etc… size, according to the ABI classification – see above, par. 2.1). Employees of banks belonging to a group were asked to respond for the group in terms of politics, strategy and market/business decisions and trends. This first section aimed at establishing the status of the bank's technology and organizational processes, focusing especially on data integration systems, customer registry management, database architecture and the connection between processes and information systems. A specific objective was to extract information about the level of automation of the IT infrastructure.

b) the second section was also related to the technology, but was focused on the impacts of MiFID on organizational processes (e.g. matters concerning best execution, client classification, etc. …) and on the organizational departments and offices involved in managing MiFID: e.g. first and foremost compliance, but also risk management, internal audit, information systems.

c) the third section asked about strategic options/strategic settings resulting from implementation of the directive. We wanted to find out about the bank's choices.

In addition to the choices made by their banks, respondents were asked about each scenario and possible strategies, expected impact on organizational structure, functions, offices, departments and business. Many of these (qualitative) answers required clarification in the second focus group meeting.

2) *Focus groups* were facilitated by the authors and were structured as follows:
a) the first meeting provided us with knowledge about the primary needs and difficulties of financial institutions in implementing MiFID –organizational and IT setting problems. We used this information to identify the variables and to build the questionnaire. A presentation from a technology firm (Datamat Spa) regarding the importance of knowledge management of client data, helped us and the sample to focus on the main organizational problems involved in changing the structure in order to be compliant with the directive;

b) in the second meeting we presented the main evidence derived from the questionnaire. The IT capabilities and strategic choices of the banks were discussed. This second meeting helped us to operationalize the qualitative answers to items;

c) the third focus group allowed us to make a final validation of the results.

Data

1) *data were collected* from the responses to the questionnaire and discussions in the focus groups. The questionnaire was administered by email a few days prior to the telephone interviews, to allow respondents time to obtain specific information (e.g.

percentage of securities portfolios, number of employees in a particular department, etc. …). The focus groups were taped.

2) *Data processing and analysis* consisted of three steps:

a) *editing*: we conducted quality control of the questionnaires to ensure that the information on the questionnaire schedule was in the right form to be transferred to the computer for analysis [48].

b) *coding*: we assigned numbers and symbols to the variables categories.

c) once the data were entered electronically, they were cleaned, i.e. we identified and resolved errors in data coding and transmission to the computer. We regarded this as an essential process that would also identify any respondent-related errors. We used a consistency checking process á la Sonquist and Dunkerberg [49:215]. The idea was to see whether the responses to certain questions were related in reasonable ways, to responses to particular other questions.

3) *tools of analysis*, which included: a) the mean used as the "balancing" point in a distribution (arithmetical average, calculated by adding the responses and dividing them by the total number of respondents);. b) median (midpoint in the distribution – the value of the middle response; half of the responses were above it and half were below); c) mode (the value or category with the highest frequency); d) range, the statistical difference between the lowest and highest values, usually reported by identifying these endpoints; d) standard deviation – a measure of the average spread of observations around the mean; e) cluster analysis, to analyze the effect of one ratio variable on another.

This generated a total of 86 graphs and diagrams, based on three data tables, the biggest of which had 159 rows and 846 columns. The instruments (software) we used were Excel and SPSS.

4 Data Analysis and Findings

Our data analysis involved iterative comparisons of theory and data. The first step included editing and summarizing the responses (coding), data entry, and error checking (cleaning). Some of data processing occurs during data collection in computer-assisted surveys. The second step, as in Broh's [48] analysis, involved inspection and modification: inspection is designed to provide a clearer picture of the data in order to determine appropriate statistical analysis and necessary data modification. The third step involved empirically testing the relationships theorized.

The two hypotheses in this study compare IT, size, and trading relevance to organizational change. We argue that only IT capabilities is strongly connected to strategic planning for organizational change.

4.1 Data Presentation

The main results are presented below:

IT capabilities

IT capabilities are measured by features of IS and human resource competences and skills, as described above (see section 2). The groups represent four different degrees of IT capabilities, identified by the cluster analysis, the results of which are presented in Table 1.

Table 1. Cluster analysis of IT capabilities

labels/ var.	**V.1**	**V.2**	**V.3**	**V.4**	**V.5**	**V.7**	**Avg**	**Var (*)**	**Valid**
IC_1	0,27	0,21	0,39	0,21	0,17	0,41	0,27	0,00848	OK
IC_2	0,31	0,31	0,48	0,33	0,27	0,55	0,363	0,00875	OK
IC_3	0,64	0,61	0,77	0,63	0,71	0,61	0,661	0,00348	OK
IC_4	0,75	0,79	0,82	0,69	0,77	0,68	0,75	0,00256	OK

(*)Validation < 0,01

Below, the percentage of banks for each IT capabilities degree:

- IC_1 is low level of IT capabilities (39%);
- IC_2 is medium level of IT capabilities (31%);
- IC_3 is high level of IT capabilities (15%);
- IC_4 is very high level of IT capabilities (15%).

Our aim is to verify whether the strategic choices made to be compliant with MiFID are (or are not) related to the clusters described above.

Strategic choice
Level of planned organizational change is represented by the strategic choice made by the financial institutions. We identified four main scenarios related to compliance with MiFID (see before), which represent the accumulation of OC involved. The distribution of the financial institutions is synthesized below:

- 62% low level of organization change (broker);
- 15% medium level of organizational change (systematic internalizer);
- 8% high level of organizational change (MTF collaboration);
- 15% very high level of organizational change (MTF – create own).

Below we provide the results of the other two variables, in order to discuss their relation with IT capabilities and strategic choice.

Size
As already mentioned, we used Bank of Italy size classifications. The size distribution was as follows:

- 47% small;
- 18% medium sized;
- 20% large;
- 15% very large.

Relevance of Trading
The relevance of the trading-business (T), whose value was explained above in the methodology, is represented in detail in Table 2.

Table 2. Cluster analysis of the relevance of trading activity (business)

labels/ var.	**V.1**	**V.2**	**V.3**	**V.4**	**V.5**	**V.6**	**Avg**	**Var**	**Valid**
T_1	0,28	0,41	0,4	0,31	0,28	0,27	0,225	0,00439	OK
T_2	0,41	0,5	0,44	0,54	0,38	0,38	0,433	0,00530	OK
T_3	0,57	0,67	0,71	0,79	0,79	0,74	0,71	0,00753	OK
T_4	0,82	0,79	0,81	0,94	0,85	0,89	0,850	0,00316	OK

(*) validation >0,01

V.1, V.2, V.N are operationalizations of the questions/responses (see section on variables for further details) we drew on to achieve the B variable measure and percentage:

- low relevance 39%;
- medium relevance 31%;
- high relevance 15%;
- very high relevance 15%.

4.2 Verification of our Hypotheses

Validation of HP1 is in the relation between IT capabilities and strategic choice (merging Table 1 and strategic choice, from the results shown above) depicted in Fig. 1, which we can see are strongly related. The matrix in Fig. 1 shows that the distribution of financial institutions is in the main split between the first and third quadrants.

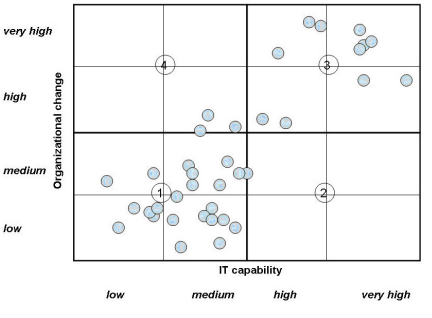


Fig. 1. Validation of HP 1

HP1 (supported): as expected, low and medium IT capabilities correspond to low and medium planned organizational change (strategic choice). The second and fourth quadrants of the matrix are almost empty. Our expectation was that it would be likely that institutions with high IT capabilities could choose a strategic option involving

high organizational change, and vice-versa. IT capabilities and strategic choice are strongly related.

In terms of HP2 we argue that IT capabilities drive organizational change, for the reasons outlined below. To verify this hypothesis we compared the strategic choice with the independent variables (S) and (T).

HP2-a is supported: i.e. size is not related to strategic choice. In other words we assume that major or large financial institutions do not necessarily choose strategic options that involve wide ranging organizational change. However, small organizations may be involved in major organizational change. Fig. 2 depicts the dispersion and sizes of institutions. The bigger the circle, the larger is the organization.

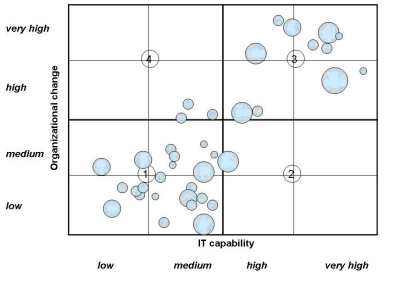


Fig. 2. Validation of HP 2a

HP2-b is supported: i.e. business relevance in financial institutions is not related to organizational change. We assume that strategic choice is not driven by the higher importance of the trade business in a specific organization. Fig. 3 depicts the dispersion of the different kinds of institutions whose profits are more or less based on financial trading. The bigger the circle, the more important is financial trading for the firm.

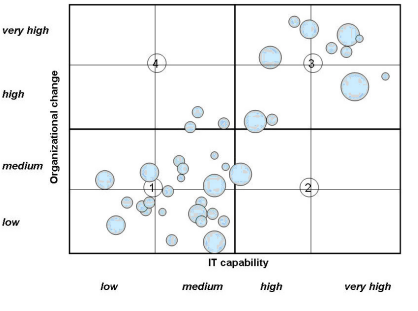


Fig. 3. Validation of HP 2b

5 Discussion and Conclusions

We can state that in the banking industry, IT capabilities has more influence on an organizational change strategy than size and specific business interest. There are

many reasons why exploiting MiFID (strong change, high strategic choice) is important: in a long-term perspective, the end of the concentration rule on national exchange authorities provides a free market, enabling first movers to gain advantages. Within this scenario, international alliances and agreements (such as MTF) can create market exchange channels with lower transaction costs. Also, the bigger the firm (in the sense of being a market maker) the greater is its contracting clout. And, finally, banks can increase their reputation and acquire more clients from trading, which will increase the use of other banking services.

But, the choice of how to operate within the MiFID is not easy and it seems that availability of funds is not a deciding factor. Choices must be made based on resources and "core competences" [50]. There are many problems involved in decision making, most of which – as argued by the authors – are determined by higher or lower levels of IT capabilities.

First there is a restriction on time in decision making related to organizational change [51]. In our study firms were forced to comply with the MiFID dispositions in the short run. The quality of decision making depends heavily on time [52]. Based on the reports from participants in the second focus group meeting, we can state that the capabilities of IT strongly influences decisions made under time pressures because: a) availability of data makes decision making easier, because it allows managers to have deeper knowledge about their own organizations and, in the banking context, provides better information about micro-processes [53]. The changes required by MiFID imply major changes to micro-processes, such as the setting up of smart order routing (SOR) platforms[5] to execute financial transaction; b) decision makers can get a clearer picture of the status of their businesses (e.g. types of clients, their ratings and aggregate values, type of transaction favored by clients, etc…) which will inform optimum choices related to organizational change. This information should be available within any organization, but the faster it can be accessed the more effective will be decisions that are made. Obviously, infrastructure alone is not enough to manage good decisions but, according to Orlikowski [54], [28], both IT and its usability are fundamental for the best performance in decision making in time constrained conditions. This is in line with our data analysis and hypotheses. Evidence demonstrates that firms lacking a clear picture of their firms, and important firm data and knowledge, will not make radical changes to their organizations. Lastly, but no less important, IT governance processes (IT business value, business IT strategic thinking, business process integration) are fundamental to planning infrastructure investments, in order not only to be compliant with MiFID, but also to exploit it to gain competitive advantage.

Second, there is the opportunity to introduce new micro-systems into existing IS. As underlined by many authors [55], [56], [57], [54] usability by human resources and especially IT capabilities in exploiting existing IS (we can call it the IT operations process), are fundamental components that contribute to the easier implementation of new software. In fact, despite their large size, organizations with high IT capabilities

[5] SOR software is an automated system to locate a suitable trading venue in order to be compliant with best execution principles, according to MiFID dispositions. SOR are able to map available venues for a financial institution and to choose the one that best fits a particular client.

are much more likely to create processes for the rapid absorption of new skills and rejection of obsolete ones [57:427]. The better that IS work, the more users can exploit them and the higher will be the organizational change. Consequently, accumulation of IT skills will provide competitive advantage.

The results point strongly to the importance of IT capabilities in the process of organizational change. This paper contributes to the organizational research that studies the important relationships between technology and organizational change. We have argued that process transformation is very often driven by IT capabilities, the features of the technology and its usability by human resources.

The limits of this study are the small size of the sample (only a portion of the Italian financial institutions) and the specific context (transformation in response to the MiFID directive).

Further research could involve cross country analysis (comparing the Italian banking industry to other European country banking industries) and multilevel analysis (process transformation vs macro-structure transformation). Also useful would be a longitudinal analysis (for a two year period) comparing the results in this paper with those from other analyses using the same sample.

Acknowledgement

Thanks to CeTIF – Research Center for Technology, Innovation, and Finance that promoted the research projects, giving all support needed for survey, analysis, and writing. Thanks to FinanceCom Workshop that gave the authors the opportunity to present this paper. The authors would like to thank also, in alphabetical order, Mauro Bello, Elizabeth Davidson, and Massimo Mamino for their contribution to the paper contents, developing of ideas an methodology issues, while Cynthia Little, helped in the linguistic editing and formatting of the manuscript. Special thanks to Chiara Frigerio, for her determinant help in developing the whole paper.

References

1. Croteau, A.M., Bergeron, F.: An Information Technology Trilogy: Business Strategy, technological deployment and Organizational Performance. Journal of Strategic Information Systems 10(2), 77–99 (2001)
2. King, W.R., Teo, T.S.H.: Integration between business planning and information systems planning: an evolutionary-contingency perspective. Journal of Management and Information Systems 14(1), 185–214 (1997)
3. Sambamurthy, V., Zmud, R.W.: Arrangements for Information Technology Governance: A Theory of Multiple Contingencies. Mis Quarterly 23(2), 261–290 (1999)
4. Orlikowski, W.J., Iacono, C.S.: Research Commentary: Desperately Seeking the "IT" in IT Research—A Call to Theorizing the IT Artifact. Information Systems Research 12(2), 121–134 (2001)
5. Mitchell, V.L., Zmud, R.W.: Endogenous Adaptation: The Effects of Technology Position and Planning Mode on IT-Enabled Change. Decision Science 37(3), 325–355 (2006)
6. Van de Ven, A.H., Poole, M.S.: Handbook of Organizational Change and Innovation, pp. 261–264 (2004)

7. Rose, J., Jones, M., Truex, D.: Socio-Theoretic Accounts of IS: The Problem of Agency. Scandinavian Journal of Information Systems 17(1), 133–152 (2005)
8. Williams, J.R.: Technological Evolution and Competitive Response. Strategic Management Journal 4(1), 55–65 (1983)
9. Katz, M.L., Shapiro, C.: Network Externalities, Competition and Compatibility. The American Economy Review 75(3), 424–440 (1985)
10. Arthur, B.W.: Competing Technologies, Increasing Returns, and Lock-In by Historical Events. Economic Journal 99(394), 116–131 (1989)
11. Shapiro, C., Varian, H.L.: Information Rules. Harvard Business Press, Boston (1998)
12. Grandori, A.: Organization and Economical Behavior. Routledge, London (1995)
13. Leavitt, H.J., Whisler, T.L.: Management in the 1980s. Harvard Business Review 36(6), 41–48 (1958)
14. Bakopoulos, J.A.Y.: Toward a More Precise Concept of Information Technology. In: Proceedings of the Sixth International Conference of Information Systems, Indianapolis, IN, pp. 17–24 (1985)
15. Fry, L.W.: Technology-Structure Research: The Critical Issues. Academy of Management Journal 25(3), 532–552 (1982)
16. Bowman, B., Davis, G., Wetherbe, J.: Three Stage Model of MIS Planning. Information and Management 6(1), 11–25 (1983)
17. Earl, J.: Experiences in Strategic Information Systems Planning. Mis Quarterly 17(1), 1–24 (1993)
18. King, W.R.: Developing Global IT Capabilities. Information Systems Management 23(4), 78–79 (2006)
19. Bharadwaj, A.S., Sambamurthy, V., Zmud, R.W.: IT Capabilities: Theoretical Perspectives and Empirical Operationalization. In: Proceedings of the 20th International Conference of Information Systems, Charlotte, NC, pp. 378–385 (1999)
20. Prieto, E., Revilla, E., Rodriguez, B.: Information Technology and the ambidexterity Hypotheses: an analysis in product development. IE Working Paper 06-07 (2007)
21. Lyytinen, K., Rose, G.M.: Information System Development Agility as Organizational Learning. European Journal of Information Systems (2006)
22. Clark, C.E., Cavanaugh, N.C., Brown, C.V., Sambamurthy, V.: Building Change-Readiness Capabilities in the IS Organization: Insight from the Bell Atlantic Experience. MIS Quarterly 21(4), 425–445 (1997)
23. Rogers, E.M.: Diffusion of Innovation, 5th edn. Free Press, New York (2003)
24. Leidner, D., Beatty, R., Mackay, J.: How CIOs manage IT During Economic Decline: Surviving Thriving Amid Uncertainty. MIS Quarterly Executive 2(1), 1–14 (2003)
25. Shafiri, H., Zhang, Z.: Agile Manufacting in practice: Application of a Methodology. International Journal of Operations & Production Management 21(5/6), 772–794 (2001)
26. Haeckel, S.H.: Adaptive Enterprise: Creating and Leading Sense-and-Respond Organizations. Harvard Business School Press, Boston, Massachusetts (1999)
27. Markus, M.L., Robey, D.: Information Technology and Organizational Change: Casual Structure in Theory and Research. Management Science 34, 583–599
28. Purvis, R.L., Sambamurthy, V., Zmud, R.: The Assimilation of Knowledge Platforms in Organizations: An Empirical Investigation. Organization Science 12(2), 117–135 (2001)
29. Mitchell, V., Zmud, R.W.: Endogenous Adaptation: The Effects of Technology Position and Planning Mode on IT-Enabled Change. Decision Sciences 37(3), 325–355 (2006)
30. Mitchell, V., Zmud, R.W.: The Effects of Coupling IT and Work Process Strategies in Redesign Projects. Organization Science 10(4), 424–438 (1999)

31. Orlikowski, W.J.: Material Knowing: The Scaffolding of Human Knowledgeability. In: Proceedings of the Sixth European Conference on Organizational Knowledge, Learning and Capabilities, Cambridge, MA (2005)
32. Orlikowski, W.J., Barney, S.R.: Technology and Institutions: What Can Research on Information Technology and Research on Organizations Learn from Each Other? MIS Quarterly 25(2), 145–165 (2001)
33. Sambamurthy, V., Zmud, R.W.: Research Commentary: The Organizing Logic of an Enterprise's IT Activities in the Digital Era – A prognosis of Practice and a Call for Research. Information Systems Research 11(2), 105–114 (2000)
34. Lawler, E.E.: From the Ground Up: Six Principles for Building the New Logic Corporation. Jossey-Bass, San Francisco (1996)
35. Ross, J.V., Beath, C.M., Goodhue, D.L.: Develop Long-Term Competitiveness through IT Assets. Sloan Management Review 38(1), 31–42 (1996)
36. Bharadadway, A., Sambamurthy, V., Zmud, R.W.: IT Capabilities: Theoretical Perspectives and Empirical Operationalization. In: Proceedings of the Twentieth International Conference of Information Systems, Charlotte, NC (1999)
37. Zmud, R.W.: Design Alternatives for Organizing Information Systems Activities. Mis Quarterly 8(2), 79–93 (1984)
38. Mintzberg, H.: Patterns of Strategy Formulation. Management Science 24(9), 934–948 (1978)
39. Thompson, J.D.: Organizations in Action. McGraw-Hill, New York (1967)
40. Ciborra, C.U., Schneider, L.: Transforming the Context and Routines of Management, Strategy and Technology. In: Adler, P. (ed.) The Future of Work, Oxford University Press, Oxford (1992)
41. Ciborra, C.U., Lanzara, G.F.: Formative Context and Information Technology: Understanding the Dynamics of Innovation in Organizations. Accounting, Management and Information Technology 4(2), 61–86 (1994)
42. Anolli, M., Petrella, G.: La nuova proposta di direttiva sui servizi di investimento e il mercato azionario: concorrenza, liquidità e trasparenza. Bancaria 10, 2–217 (2003)
43. Van de Ven, A.H.: Suggestions for Studying Strategy Process. A Research Note. Strategic Management Journal, S.I., 13, 169–191 (summer 1992)
44. Mohr, L.B.: Explaining Organizational Behavior. Jossey-Bass, San Francisco (1992)
45. Abbott, A.: Transcending General Linear Reality. Sociology Theory 6(2), 169–186 (1988)
46. Pentland, B.T.: Building Process Theory with Narrative: From Description to Explanation. The Academy of Management Review 24(4), 711–724 (1999)
47. Poole, M.S., Van de Ven, A.H., Dooley, K., Holmes, M.E.: Organizational Change and Innovation Processes. Oxford University Press, Oxford (2000)
48. Broh, B.: Linking Extracurricular Programming to Academic Achievement: Who Benefits and Why? Sociology of Education 75(1), 69–95 (2002)
49. Sonquist, J., Dunkelberg, W.: Survey and Opinion Research: Procedures for Processing and Analysis. Prentice-Hall, New York (1977)
50. Lipparini, A.: Le Competenze Organizzative. Carocci Editore, Roma (1995)
51. Swenson, O., Maule, J.: Time Pressure and Stress in Human Judgment and Decision Making. Plenum Press, New York and London (1993)
52. Ahituv, N., Igbaria, M., Sella, A.: The effects of time pressure and completeness of information on decision making. Journal of Management Information Systems 15(2), 153–172 (1998)
53. Mayer, R.E.: The Promise of Multimedia Learning: Using the Same Instructional Design methods Across Different Media. Learning and Instruction 13, 125–139 (2003)

54. Orlikowski, W.J.: Using Technology and Constituting Structures: A practice Lens for Studying Technology in Organizations. Organization Science 11(4), 404–428 (2000)
55. Somers, T.M., Nelson, K.: The impact of critical success factors across the stages of enterprise resource planning implementations. In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (2001)
56. Van de Ven, A.H.: Running in Packs to develop Knowledge-Intensive Technologies. Mis Quarterly 29(2), 365–378 (2005)
57. Clark, C.E., Cavanaugh, N.C., Brown, C.V., Sambamurthy, V.: Building Change-Readiness Capabilities in the IS Organization: Insights from the Bell Atlantic Experience. Mis Quarterly 21(4), 425–455 (1997)

Cash Tokens for SAML Based Federations

David J. Lutz

Rechenzentrum Universitaet Stuttgart
Allmandring 30; 70550 Stuttgart; Germany
David.Lutz@rus.uni-stuttgart.de

Abstract. This paper introduces an approach for digital cash within SAML based federations. The approach bridges between the currently established federation infrastructure and the payment applications that are not often implemented by such federations. Since a lack of integrated payment can be seen as a drawback when federations offer commercial services, the approach described in this paper offers the possibility to include the whole payment process into the federation protocol design and architecture. A SAML token together with a SAML Payment Assertion is used to transmit digital cash and payment information within the federation.

Keywords: Federation Payment, SAML Token, Payment Assertion, Payment Provider.

1 Introduction

The importance of federations and federated identity can no longer be overlooked as more and more service providers offer their content within federations. Large federations have already been established within an academic context, but commercial federations will surely following. The collaboration of several large companies in the Liberty Alliance group can be seen as evidence for that.

But regardless of which federation is selected, the academic federations that allow students access to restricted content as well as commercial federations that allow users to consume specific services and resources, the federation infrastructure and its protocol design has its focus only on authentication and authorization. For sure, it is very important to decide whether a user is allowed to access the resource, but one major inducement for commercial service providers to join the federation is still missing: federation based payment.

In this paper, the current important federations are examined together with several methods for electronic payment. A worthwile approach for combining federations with payment is to use the same protocol, design and language. Therefore, the payment approach described here is oriented towards the federation structure based mainly on SAML. This paper shows how digital cash should be designed for secure usage within a SAML federation.

The rest of the paper is organized as follows: Following the introduction, there is an overview of related work. Next, the proposed federation structure

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 97–111, 2008.

is described in detail together with the mechanisms for authentication and authorization. Then, a security analysis is made, followed by the conclusions and future outlook.

2 Related Work

Two main areas of research are to be noted: The work already carried out in federations and the current and past ideas for electronic payment (ePayment). However, since almost none of the approaches handle both these important fields together, the identity management as well as the ePayment, the current major approaches for federations are described first and then a view towards the several existing ePayment solutions is made.

2.1 Federation Approaches

The idea of a federated network is not new and a number of different approaches already exist. The Grid community started very early to set up federation management structures such as CAS [1], Akenti [2], PERMIS [3] and VOMS [4], and the web community followed suit with approaches such as Shibboleth, Liberty Alliance and WS-Fed.

SAML can be seen as a standard for transmitting security information related to the user's identity and therefore, in the following subsections, only the three SAML based or SAML oriented federation approaches are described.

SAML. The Security Assertion Markup Language (SAML) [5] is the standard that defines an XML-based framework for describing and exchanging security information between online partners. This security information is expressed in the form of portable SAML assertions that applications working across security domain boundaries can trust. Typically, SAML assertions are conveyed with SOAP [6] messages. SAML expressions in the form of authentication and attribute assertions as well as different profiling mechanisms, constitute the structure created by identity provisioning and payment mechanisms.

Shibboleth. Shibboleth [7] is an Internet2/MACE project whose goal is to develop an open, standard-SAML based solution allowing different entities to exchange information about their users in a secure fashion.

Shibboleth offers a Single Sign On (SSO) solution for large federations. It allows the users to request a resource without having being authenticated there beforehand. Also, the user does not need to have federation specific credentials. Authentication is carried out once and after that only authorization issues have to be considered since, following authentication, a browser profile or artifact based on SAML is generated for the user that can be used to authenticate him at his Identity Provider[1] later. Due to the fact that each authorization is

[1] The Identity Provider is the entity that stores the user's authentication information (often his whole identity attributes) and that therefore can authenticate him.

done separately with new requested attribute assertions, the user can decide (by defining a policy at his Identity Provider) which information should be sent to the resource provider. Shibboleth is constrained to federations that use browsers and HTTP to provide access to the resources. This may cause problems in broad federations, since they may possibly use other protocols. Like many other approaches, Shibboleth does not specify the authorization system and assigns the inspection of the attribute assertions to the resource provider. The federation, as proposed in our framework, is SAML based and follows many ideas that were developed for Shibboleth and Liberty Alliance.

Liberty Alliance. The Liberty Alliance project [8] is aims to develop open standards within the federation and SSO scope.

The Liberty Alliance is similar to the Shibboleth approach. It was also developed for web use with browsers and HTTP. The main differences between Liberty Alliance and Shibboleth exist at a level other than the basic model architecture. They were originally developed for different purposes and have a different view of trust in their federations. The schemes they have selected are also different, but it is possible to convert between them. It is expected that in future the differences between Liberty Alliance and Shibboleth will be only in the Identity Management Space. Liberty Alliance offers, unlike most of the other approaches, the possibility of a Single Log Out and provides a concept for using it in wireless federations. Like Shibboleth, Liberty Alliance is built upon the SAML Single Sign On Browser/Artifact profile, which must be seen as imprecise and vulnerable [9].

Web Service Federation. Web Services Federation (WS-Fed) [10] is part of an overall effort by IBM and Microsoft to build a Web services security framework, or WS-Security. This approach can be used for web service requestors using SOAP as well as web browser (access with HTTP) requestors.

The main components in the WS-Federation are 1: the requestor, wanting to have access to a restricted resource, 2: the resource, which may have specific restrictions related to the access, and 3: an Identity Provider combined with a Security Token Service (STS). Other components are the Attribute Services and the Pseudonym Services. The requestor uses either his webbrowser or a requestor service to ask for access at the resource. The Identity Provider performs authentication and can make identity claims in issued security tokens. Attributes are received from the Attribute Server, which can be combined with the Identity Provider. In case of wanted pseudonymity at the resource the Attribute Service is enhanced by a Pseudonym Service, which links the requestor's attributes to a pseudonym. The authorization is done by an Authorization Service, a specific instance of an STS that operates in a decision brokering process. Different types of authentication are specified and can be used in the authorization process, too.

The process of requesting and receiving attribute and authentication information is similar to that of Shibboleth and Liberty, but with some differences: WS-Fed allows access by web browser and web (requestor) services. The use

of cookies in the requestor's browser is optional and the security token is not specified to follow a specific language, it can be built using, e.g., SAML or X509.

2.2 Payment

The approaches for handling payment in electronic commerce can be split up into three different areas: the transmission of payment information, the transmission of exchange information and the transmission of digital cash. The importance of the approach used increases from the transmission of payment information up to the transmission of digital cash, because the first approach focuses only on the secure transmission of payment information, e.g., credit card data, whilst the digital cash transmission provides ideas that can be included in the SAML based federation architecture and infrastructure. In the following subsections, typical representative schemes of those approaches are discussed according to their relevance to the idea proposed in this paper.

Transmission of Payment Information. One solution for ePayment is the transmission of payment information. Although information only required for traditional payment is transmitted, this area if often seen as part of ePayment. Thus, payment ideas that only focus on the transmission of payment information are: e.g., Secure Electronic Transaction (SET) [11] and PayPal [12].

Although both approaches are often used within eCommerce and ePayment, SET for securing credit card transactions over the Internet [13] and PayPal to transfer payment information related to an online business transaction from one PayPal account to another one, these solutions are not described in detail, since this paper focuses more on the transmission of digital cash than on transmission of payment information.

Transmission of Exchange Information. The approach for transmission of exchange information handles the idea to transmit something that can be directly exchanged into real money. That solution was implemented by including the traditional cheque concept into the web area and by inventing and defining 'electronic cheques'.

NetCheque. NetCheque is a distributed accounting service supporting the credit-debit model of payment [14]. A NetCheque user has to have an account on an accounting server. When a user needs to issue a cheque, e.g., before or after an online transaction, the customer issues the cheque in a similar manner to the conventional way of issuing cheques. Using an application, the customer specifies the name of the payer, the name of the financial institution, the payer's account, the name of the payee, the amount of the cheque, the currency unit and an expiration date. This information is signed by the payer and transferred to the payee. The clearing of the cheque at the payee's side follows the same schema as for conventional cheques. The use of conventional cryptography instead of public cryptography enhanced the performance of the NetCheque implementation and enables its possible use it for micropayments.

Electronic Check. Another electronic cheque solution is proposed by CommerceNet and called eCheck [15]. This approach works in the same way as the NetCheque idea, but instead of traditional cryptography with Kerberos, here a public key infrastructure with X.509 certificates guarantees the genuineness of the signatures. When a user opens an account at a bank that supports the Electronic Check technique, he receives an electronic smart card, which is used to protect the user's private key from theft and misuse [16]. The payer uses an issued application to write data on the electronic cheque, which is illustrated in the same way as a conventional cheque. After filling out the form, the payer signs it with the key on the smart cart and sends the cheque to the payee. The payee, in turn, adds his own account information and signature to the cheque and forwards it to his bank, where it is cleared.

To enhance security, eCheck allows the payer to first send the cheque firstly to his bank. The bank validates the payer's identity, checks his account and countersigns the cheque.

Transmission of Digital Cash. The third major area within ePayment is the transmission of digital cash. In this case, an unforgeable or at least traceable token is generated. This token is handled between contractual bound entities like cash, which means that the transmission of such a token is the same as paying with real money. In the case of an unforgeable token the receiver can be sure that the token will be accepted instead of real money during other transactions or that the token-issuing entity will exchange the token into real money. Often, within some federations, the token is seen as providing enough security even if the token is forgeable, providing the falsifier can be detected should the token be misused.

NetCash. The University of Southern California developed the NetCash [17] framework for electronic currency. NetCash allows an anonymous payment without using tamper-proof hardware but with the ability to detect illegal creation, copying and reuse of the electronic currency. The central components of this approach are the currency servers, which are registered to a non-anonymous payment infrastructure, e.g., electronic Checks or other payment protocols. It is possible, that some currency servers are connected only to other currency servers and not directly to the non-anonymous infrastructure, but in that hierarchical solution the top currency servers must be connected to the infrastructure.

Electronic cash in the NetCash environment contains all the information about the monetary value of the cash token, an internet address that specifies the currency server, which has issued the cash token, a time stamp to state the cash token's lifespan and a serial number. All the information is signed with the currency server's private key. A payer that wants to pay with such a cash token gets it from a currency server. The payment is done by submitting the signed cash token to the payee. After that, the payee has to validate the cash token by sending it to the currency server, which address is indicated in the token. The

currency server checks the cash token's serial number with a list of outstanding numbers of already issued cash tokens and its lifetime. If the token is valid, the server issues a new cash token to the payee that can be used for further payment.

This approach allows electronic payment with electronic cash tokens in an anonymous way. The tokens are protected against forgery and double-spending by the use of a serial number list at the currency server. Since the cash tokens are replaced after each payment and information about the transaction is not saved, they can be seen as untraceable. The main drawback to this solution is the need to contact the currency server immediately after the transaction to check the cash token's validity. That the currency server is able to log the user to whom the cash token is issued and the payee that asks for the validity check can be seen as privacy concern.

Digital Cash. Chaum [18] proposed in 1988 the first digital-cash scheme. Although it is no longer implemented or used in other ways due to computational drawbacks, it can be seen as the foundation of many approaches and improvements, e.g., the suggestion from Brands [19] and Mu [20].

Brads' system consists of payers, payees and a bank. During the setup-phase, the payer generates a key pair and sends the public key to the bank. When the payer needs a cash token, the bank sends cryptographic information and he identifies himself at the bank by sending back a kind of signature of the information. To withdraw a cash token, the payer generates the cryptographic cash token and sends information of the cash token to the bank, where this information is signed and sent back to the payer. The cash token consists only of cryptographical material generated by the payer and the bank and countersigned by the other party. Banks does not propose an amount to be stated in the cash token. When the cash token was transferred to a payee, the payee sends the payer a receipt, which is signed and sent back to the payee. The payee checks the cryptographical token as far as he can and sends it to the bank for depositing afterwards. The bank can check, via a list, if the cash token is already spent (due to the cash token-related signatures of the payer and the bank), and thus able to discover illegal spending. An illegal usage of the cash token cannot be detected by the payee without contacting the bank immediately.

3 Payment Assertions

A first approach to link federations with payments was done by Jennings et al. [21]. They proposed a payment mechanism that works similarly to the identity provisioning within SAML federations. Whenever a service provider receives an access-request for a commercial resource, he sends a SAML payment offer to the customer, including information about the service provider, the merchant bits[2], the cost and the billing scheme and some SP information.

[2] The Merchant Bits are characters built to a string that helps to assign the payment assertion to a specific bargain.

After receiving the offer, the consumer contacts his payment provider and asks for a payment assertion. The payment assertion (PA), which is evidence of transmitted money from the user's account at his Payment Provider (PP) to the SP's account, looks like:

```
<saml:Assertion
  xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion"
  Version="2.0"
  ID="abcd"
  IssueInstant="2005-01-31T12:00:00Z">
  <saml:Issuer>www.payment-provider.com</saml:Issuer>
  <saml:Subject>
     ...
  </subject>
  <saml:Conditions
     NotBefore="2005-01-31T12:00:00Z"
     NotOnOrAfter="2005-01-31T12:00:00Z"/>
  <saml:AuthnStatement
     AuthnInstant="2005-01-31T12:00:00Z"
     SessionIndex="67775277772">
     <saml:AuthnContext>
        ...
     </saml:AuthnContext>
  </saml:AuthnStatement>
  <saml:AttributeStatement>
     <saml:Attribute
        NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:uri"
        Name="urn:ietf:params:xml:ns:payattr">
        <samlAttributeValue xsi:type="payattr:PaymentReceiptValueType"
           payattr:merchantBits="MDE1Mw=="
           payattr:pspBits="abc"
           payattr:serviceUrl="https://psp.example.com/paymentService"
           payattr:currencyNamespace="currency:namespace:..."
           payattr:currencyDivisor:1000
           payattr:currency="USD" payattr:amount="300"/>
     </saml:Attribute>
  </saml:AttributeStatement>
</saml:Assertion>
```

The consumer's PP reserves the amount of the transaction and generates the related payment assertion. After creating the PA, the PP sends a SAML URI Reference to the consumer, from where it is forwarded to the SP. The SP can now contact the user's PP, asking to resolve the reference. In that case, the PP sends the PA to the SP, where it can be examined. If the PA is valid, the customer can consume the resource.

This approach is a major step towards federation payments. Actually, it guarantees a basic payment, but there are also some disadvantages that should not been overlooked: The assertion does not contain a signature of the PP or anything else, which can prove that the assertion was issued by the PP. Although the SP can rely on the certificate that the PP provides when he transmits the token, he has no proof of that transmission, when he requests the monetary exchange. Thus, the assertion can only be seen as a receipt for payment, already done, and not as a real payment. The amount in the assertion has to be the exact amount that the SP needs. If the consumer pays too much, the SP has to use the same mechanism (contacting his PP, payment request, SAML URI, etc.) to partially or fully refund the money. And here is a privacy issue, which may lead to much concern when using this approach. The PP needs knowledge about the consumer's identity (for charging him), about the SP (for sending him the assertion) and about the merchant bits. This means that the consumer's PP knows which consumer bought what at which SP.

4 Federation Architecture

For the approach of digital cash for SAML based Federations, a 'typical' federation is assumed. This idealistic federation is oriented towards SAML based federations and consists of one or more of the following:

- Consumers: A consumer uses the federation to access services and resources. He may buy and consume the provided goods and services and has contracts with the Identity Provider and the Payment Provider. He accesses the SP's resources via a network to which he is connected by either a mobile or fixed terminal.
- Identity Providers: An Identity Provider (IdP) stores the user's attributes in a repository. The IdP is the only entity that can authenticate the user. After a successful authentication, the IdP issues a token to the user that validates his identity. It is assumed that the IdP has contracts with the customers, whose profile it hosts, and the service providers to whom it issues authentication information (or to the federation manager). Furthermore, the IdP can issue user's attributes like address, age, gender, etc. to service providers if this information is needed for authorization.
- Service Providers: The Service Provider (SP) offers services to consumers and expects payment as well as a reliable authentication. He receives user's attributes from the IdP to perform authorization.
- Payment Providers: The Payment Provider (PP) hosts the user's account. It can be a bank or part of the IdP, but does not have to be so. The PP issues payment assertions that can be seen as the basis for digital cash within the federation. It is the only entity that can validate the digital cash tokens and exchange them into real money.

Apart from those components, often a federation has a federation administrator that controls the federation, signs contracts with all members, defines the federation rules, etc.

However, three different important steps are necessary before a resource's consumption can be identified: The authentication, the authorization and the payment. All three steps will be briefly explained in the next subsections.

4.1 Authentication

When a user is requesting access to a service or resource, the SP asks for authentication information and an already authenticated user presents a secured token that states his identity. If the user is not authenticated, he is redirected to his IdP where he submits his credentials, e.g., username and password. The IdP checks the credentials, stores maybe the user's authentication information for further purposes and returns a unique token to the user proclaiming his identity, if the validation was successfully. After authentication, the user can present the token to the SP. The SP checks the token, e.g., by sending a request to the IdP, asking if the token is still valid. Thus, the token can be seen as a digital IDCard and is called an IDToken[3]. Figure 1 shows, in brief, the message flow for a typical initial authentication.

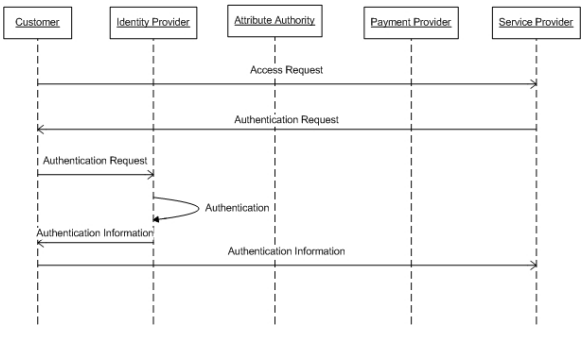


Fig. 1. Message Flow for an Initial Authentication

4.2 Authorization

Often, the SP does not only need to know whether the user is the one he claims to be, but also other information is required. At this stage of service access, the SP is already aware of the user's identity, since he has submitted his token. Now, the SP contacts the user's IdP and requests attributes that are related to the user and needed for authorization, e.g., studies information for accessing an academic service or age for an adult-content service. The IdP validates the token and checks if the user's attribute release policy allows the transmission of the requested attributes. If so, it generates a SAML Attribute Assertion and sends it to the SP, where the assertion is examined and the access request is decided. This process can be seen in Figure 2.

[3] We propose a token that contains a SAML artifact as a pointer to the SAML Authentication Assertion, a Serial Number to avoid replay attacks, a Random Number for security reasons, user's ID and the user's signature.

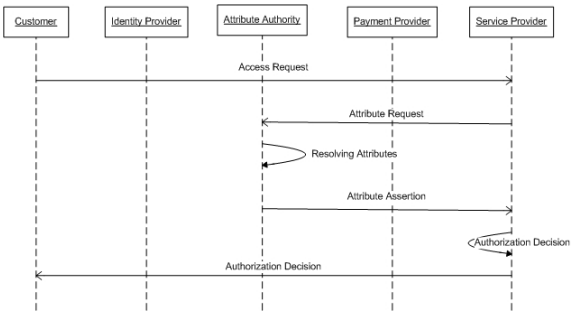


Fig. 2. Message Flow for a User Authorization based on his Attributes after Authentication

4.3 Payment

The third step, when consuming a web-resource, may be payment. Currently, this step is not integrated into the federation architecture; often, SPs need for granting service access only an authorization based on the user's attributes and/or for the payment process a different solution is used, e.g., a credit card. To bridge between the current identity provisioning and the non-integrated payment solutions, this paper introduces digital cash on top of payment assertions for SAML based federations.

The digital cash token, considered for payment in federations, is designed in a similar way as the IDToken for authentication: The digital cash token consists of

- SAML Artifact: A pointer to the Payment Assertion in the PP's database
- Payer's ID: The payer's ID, e.g., `username@domainA`
- Payment Provider ID: The ID of the assertion factory of the issuing payment provider; such as `assertionFactory@paymentProvider.com`
- Payee's ID: The payee's ID, e.g., `serviceprovider@domainB`
- Currency: The currency of the token related to ISO 4217
- Lifetime: The token's lifetime that corresponds to the lifetime of the related payment assertion
- Serial Number: A serial number to identify the token
- Transaction ID: The transaction ID assigns the token to a specific transaction
- Signature: The payer's signature of the token is done with his private key.

Whenever a user needs digital cash within the federation, he first requests a SAML Payment Assertion at his PP. After checking his account and authorizing him, the PP generates a payment assertion following the schema proposed by Jennings and sends it to the user. The digital cash token is created at the user's terminal after receiving and accepting a payment offer from an SP. The payer can decide (maybe forced by federation policy) whether to create one or more cash tokens to cover the sum amount to be paid. That basic payment process is shown in Figure 3.

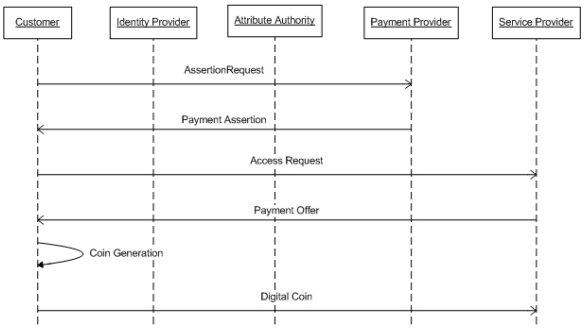


Fig. 3. Message Flow for a Payment Process after Authentication and Authorization

The SP validates the token by checking the payment information and the signature. Afterwards, the SP can directly exchange the cash token into money by contacting the PP whose address is stated in the token; the SP can do this also during the payment process to validate the token. On the other hand, the SP can collect the cash tokens and use the tokens to create new ones without contacting the PP when the SP has to pay during a transaction. In that case the new token has the same artifact pointing to the same assertion and the same lifetime but with new payment information.

The second use of the cash token is more related to the usage of real money, since this is not checked after each transaction, but it leads to major security concerns, because the token may be forgeable. Whilst a direct check by contacting the PP allows the detection of misuse during payment, the reuse of the token must be secured. This can be done in two ways: Weak security would detect the misbehaving person after misuse is detected; a stronger solution would be to secure the token in such a way that misuse is not possible.

The weak security is guaranteed in the design of the token. Since each token is bound to a payer and a payee (and both identities should be validated before the payment by checking the X.509 certificates), a transaction and an initial assertion request, a forged token can be tracked to the cheating entity. But this mechanism has some drawbacks: Although the cheating entity can be detected, the misuse has already taken place and someone has lost money. Furthermore, the detection requires the collaboration of maybe many federation participants and the possibility to find the cheating person even when he has already left the federation.

The strong security approach requires an unforgeable token. Since usually a data token can be copied without any problems, a way has to be found to avoid that. One approach for securing the token is the use of a trusted platform module [22] or other safe hardware. A trusted platform is a hardware chip that can store information securely. To do so, the module has implemented an Endorsement Key which cannot be tampered with without destroying the chip. This chip can be used to protect (by signing) the application that is used for the generation and the transmission of the token. When the application is issued by a payment

provider and it is secured by the module's key, the payee can be sure that the application is working correctly. The module's key can be further used to encrypt the token, thus, the user can only copy the encrypted information that he cannot use without the module and the correct working application. If the application logs all actions related to token usage and this information is also protected with the module's key, the token can be seen as highly secure [23].

5 Security Analysis

Due to the critical nature of Payment Tokens, it is necessary to give utmost priority to security and privacy. To fulfill the requirements regarding security, the assumption is that the federation has a reliable Public Key Infrastructure (PKI).

In the following subsections, all participants in a federation payment scheme are analyzed. It is evaluated when and how they can try attacks and how these attacks can be avoided. Although both security solutions may be accepted, this section focuses more on the use of tamper-proofed hardware, since detecting of a misbehaving element within the federation is easy, because a PKI is used.

5.1 Payment Provider

The PP has four ways to compromise the payment scheme: He can issue a false assertion, he can tamper with the user's account, he can issue false payment applications and he can refuse to exchange a token for real monetary funds.

Issuing false assertions or payment applications can be easily detected and tracked to the PP and implies violating contracts and losing trust. Thus, in this case, the PP will not behave in such a manner, especially if the contract provides for harsh penalties. Modification of a user's account is nothing special in this approach and can be caught with a good contract. The need for exchanging the token into real monetary funds lies in a secured infrastructure: Either the token is secured in such a way that it must be exchanged by the PP, because misuse is not possible, or in case of misuse defrauders can be detected.

5.2 Consumer

The consumer is the weakest link in this approach. He can try to copy or modify the token, use one token twice, copy or modify the assertion.

If only the misuse shall be detected, the attack can be tracked to the defrauder. If a tamper-proofed hardware is used, a copy of a token cannot be taken to another terminal, since each token bound to the hardware. The use of one token on one the same terminal twice is avoided by the payment application. The token is signed by the tamper-proofed hardware and, after a token is used, it's serial number is stored securely. The user has no way to use such a token more than once.

The consumer cannot tamper with the initial assertion or the token, since the tamper-proofed hardware together with the payment application would detect

the misuse when checking the PP's signature. Copying and using the assertion or the token to another terminal can be avoided by adding encrypted material to it that can only be encrypted by the tamper-proofed hardware of one terminal. The reuse of assertions and tokens is not possible, since the payment application can check a local secured database if the assertion or the token was used before.

5.3 Service Provider

The SP can try to attack the infrastructure in several ways: He can float the contract, which is not specific to this approach and would invoke penalties, since the SP and consumer are non-repudiably identified with their certificates from the PKI. The SP can try to modify the token in the same way as a consumer, but here the above security mechanisms would lead to failure.

5.4 Eavesdropper and Man in the Middle

Another problem lies in the possibility that a third party would be able to listen to the communications and obtain knowledge about the token. This can be avoided by using a secure underlying network technology, since the consumer's and SP's certificates can be validated, or by encrypting the token with the public key of the SP.

The underlying security technology, especially the use of X.509 certificates, can also avoid a Man-in-the-Middle attacks.

5.5 Hardware Theft

The infrastructure does not provide any security against hardware theft, but with passwords for accessing the secured/encrypted areas more security can be provided. Only if the thief is able to activate a payment process and pay with the stolen token, would the consumer lose money. This possible attack must be seriously considered when the token is mobile, e.g., when the user stores it not only on his terminal, but also on a USB stick to be used on different devices.

5.6 Data Loss

There are several scenarios, which can lead to a loss of data, either on the consumer's side or at the SP. The consumer can lose his assertion or a stored token. Both cases means, that he will temporarily lose money in a similar fashion as if he had damaged a banknote. However, due to the lifetime of the assertion, loss of data would not mean that he could not claim a refund. If, after the lifetime period, the token is not presented to the PP, the PP can refund the money, as the PP has had no need to honor it.

More critical is the loss of a received payment token. If the consumer is not willing to corroborate the payment again and the SP has no proof of the transaction, the SP will not be able to obtain any money. Thus, this is an obvious requirement for strong data backup at the SP.

6 Conclusions and Outlook

This paper analyzes the current major federations together with several payment approaches. The detected lack of implementations of payments in federations can be resolved by adding the payment functionality directly to the federation functions for authentication and authorization is now. This solution requires the introduction of a new federation entity, the Payment Provider and by adding the payment assertion to the assertion pool that is used in the federation. But since the payment assertion follows the same rules and language as other assertions used in the federation, its implementation is made easy. The use of payment tokens as digital cash tokens improves the payment concept again, especially when dynamical and fast composed services are considered that should not wait more than a few (milli-) seconds to deliver the service. Together with a strong security solution based on specific hardware or a weaker security that only detects the token's misuse, the token concept allows the full integration of digital cash into the federation's design.

One issue left for further research is the security of the token. Both approaches, the weak security idea only for detecting misbehavior as well as using tamper-proofed hardware such as trusted platform modules, cannot be seen as a best idea. Therefore, further research will need to concentrate on the question on how to secure the token in a way that it is as secure as real money without introducing new specific and obligatory hardware.

Acknowledgement

The approach published in this paper was partly developed in the project AKOGRIMO [24], funded by the EC under the FP6-IST programme. The Author would like to thank all the partners involved in that project.

References

1. Community Authorization Service (CAS) Documentation (June 2007), http://www.globus.org/toolkit/docs/3.2/cas/
2. Akenti – Distributed Access Control (June 2007), http://dsd.lbl.gov/Akenti/
3. PrivilEge and Role Management Infrastructure Standards Validation (June 2007), http://www.permis.org/index.html
4. Alfieri, R., et al.: From gridmap-file to VOMS: managing Authorization in a Grid environment. Future Generation Computer Systems 21(4) (2005)
5. Ragouzis, N., et al.: Security Assertion Markup Language (SAML) V2.0 Technical Overview. October 2006 (June 2007), http://www.oasis-open.org/committees/download.php/14361/sstc-saml-tech-overview-2.0-draft-08.pdf
6. W3C SOAP Specifications (June 2007), http://www.w3.org/TR/soap/
7. Shibboleth Website (June 2007), http://shibboleth.internet2.edu/
8. Liberty Alliance Project: Liberty Alliance Project Whitepaper: Personal Identity (2006)

9. Gross, T.: Security Analysis of the SAML Single Sign-on Browser/Artifact Profile. In: Proc. of the Annual Computer Security Application Conference (2003)
10. Lockhart, H., et al.: Web Services Federation Language (WS-Federation). Version 1.1, IBM Corporation, December 2006 (June 2007), http://www.ibm.com/developerworks/library/specification/ws-fed/
11. Kou, W.: Payment Technologies for E-Commerce. Springer, Heidelberg (2003)
12. PayPal Inc.: The PayPal Website, https://www.paypal.com
13. IBM International Technical Support Organization: Secure Electronic Transactions: Credit Card Payment on the Web in Theory and Practice. Redbook, IBM Corporation (June 1997)
14. Neuman, C., Medvinsky, G.: Requirements for Network Payment: The NetCheque Perspective. In: Proceedings of IEEE COMPCON 1995 (March 1995)
15. Jaffe, F., Landry, S.: Electronic Checks: The Best of Both Worlds. Whitepaper. In: Electronic Commerce WORLD (July 1997)
16. Anderson, M.: The Electronic Check Architecture. eCheck initiative (September 1998)
17. Medvinsky, G., Neuman, C.: NetCash - A design for practical electronic currency on the Internet. In: Proceedings of 1st the ACM Conference on Computer and Communication Security (November 1993)
18. Chaum, D., et al.: Untraceable electronic cash. In: Goldwasser, S. (ed.) CRYPTO 1988. LNCS, vol. 403, pp. 319–327. Springer, Heidelberg (1990)
19. Brands, S.: Untracable off-line cash in wallet with observers. In: Stinson, D.R. (ed.) CRYPTO 1993. LNCS, vol. 773, pp. 302–318. Springer, Heidelberg (1994)
20. Mu, Y., et al.: A fair electronic cash scheme. In: EMMCVPR 1997. LNCS, vol. 1223, pp. 20–32. Springer, Berlin, Heidelberg (2001)
21. Jennings, C., et al: Payment for Services in Session Initiation Protocol (SIP). Document ID draft-jennings-sipping-pay-05 (2007), http://www.ietf.org/ID.html
22. Trusted Computing Group. Trusted Platform Module (TPM) Specifications (2006), https://www.trustedcomputinggroup.org/specs/TPM
23. Lutz, D.: Federation Payments using SAML Tokens with Trusted Platform Modules. In: Proceedings of the IEEE Symposium on Computers and Communications (2007)
24. Access to knowledge through the Grid in a Mobile World (AKOGRIMO). Funded by the EC under the FP6-IST programme, http://www.mobilegrids.org/

Foreign Delisting and Domestic Stock Value: Multiple Frameworks, Different Views?

Loredana Ureche-Rangau[1] and Andrea Carugati[2]

[1] Assistant Professor, IESEG School of Management, 3 rue de la Digue, 59000 Lille, France
Tel.: +33 320 545 892
l.ureche@ieseg.fr

[2] Assistant Professor, Aarhus School of Business, University of Aarhus, Fuglesangs Allé 4, 8210 Århus V, Denmark
Tel.: +45 8948 6361
andreac@asb.dk

Abstract. The aim of our paper is to understand the foreign delisting phenomenon using a multi-method approach based on both information content analysis and event study analysis of foreign companies withdrawing their stocks from the Tokyo Stock Exchange. Our objective is to measure the impact of the delisting decision on the domestic stock price by observing the link between the pure value impact and the reasons for delisting. Our results show that, contrary to previous results, firms delisting from Tokyo can gain value depending on the reaction of the market operators to the content of the information provided in the delisting announcements. i.e. delisting can be presented either as the result of the not fulfilment of the benefits of cross-listing or as a part of a more general strategy of the company. Highlighting significant clusters, content analysis provides a valuable addition to traditional event study and useful guidance in understanding the delisting phenomena.

Keywords: Delisting, event study, stock returns, news announcements, multi-method approach.

1 Introduction

During the last twenty years, many companies have chosen to list their stocks both on the domestic and on foreign capital markets. This phenomenon is largely analyzed by the financial literature because such a decision implies both a huge challenge (e.g. organizational, financial, legal) for the listed company and an important impact on the fundamental features of the concerned stock, i.e. in terms of returns, risk, trading volume, liquidity, etc. The studies seeking to understand the net benefits of the corporate decision to list abroad specifically underline the benefits in terms of lower cost of capital, larger global shareholder base, increased trading liquidity and prestige (see [1]). Concurrently these studies point out the costs of such a decision, namely reconciliation of financial statements with home and foreign standards, direct listing costs, exposure to legal liabilities, taxes and numerous trading frictions. Nevertheless, the benefits are said to be significantly higher than the associated costs and the phenomenon largely contributed to the tremendous increase in cross-border capital flows and

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 112–135, 2008.

hence, to the global financial markets integration (for an excellent monograph on cross listings please refer to [2]). Empirical studies using foreign listings on the US market clearly show an increase in the value of the firm following the listing (see for example [3] or [4]). They confirm the theory ([5], [6] and [7]) which states that the removal of investment barriers should have a positive impact on stock prices.

In a global market where listing on foreign markets has been accepted as a positive move, the opposite operation – delisting – assumes a highly negative connotation. The case of Nasdaq is indicative of how the market professionals interpret delistings. After a record 240 listing in 2000, Nasdaq registered a record 279 delisting in 2001. Commenting on this event, delisting was considered as "embarrassing" (interview with C.J. Voss, a securities lawyer) and "a real psychological blow to investors, and the board and employees, too … It implies more than just getting off the exchange, it implies at best a long road back and at worst it's the first sign of filing for bankruptcy or selling out to another company" (interview reported in [8]).

The negative implications of delisting are therefore dominant in the mental models of market professionals as the result of more then a decade of theoretical and practitioner's reporting on the positive effects of the opposite event.

More recently, the introduction of the Sarbanes-Oxley Act of 2002 (see for instance [9]) represented for many smaller and non-US companies an additional burden to maintain public status on the American markets. As result we are observing an increasing trend to be released from the burden of SEC reporting requirements [2] even though the consequences on the stock value are still unclear. While delisting is a less-cumbersome alternative respect to going private, since it does not require a fundamental alteration of capital structure, companies must consider and explicitly weigh the costs and benefits of SEC registration, including the prestige factor, greater liquidity, and greater access to capital markets (please refer to [1]). Even in the face of cumbersome and costly administrative procedures to uphold the Sarbanes-Oxley regulations, market operators are still warning off against hasty delisting which evidences a persistent and generalized mental model that can be resumed in: listing is good, delisting is bad.

Empirical studies show that changes in the international financial environment and the market downturn of the recent years have initiated a decreasing trend in cross-listings. In [2] there are 2,300 cross-listings reported by the end of 2002 compared to 4,700 in 1997. Not only a decreasing number of companies are listing on multiple markets but increasingly companies take action to delist their stocks from foreign markets. This operation will be called in this article a foreign delisting: a company registered in a country that delists its shares from a foreign market and continues trading them on – at least – the home market.

In this case, if foreign cross-listings contribute to an increase of the stock value, then foreign delistings should have an opposite effect. This intuitive relation is in reality the result of an extremely complex web of actions carried out by the players on the stock market. A simple causal relation cannot be accepted without further study because the reasons for foreign delisting are multiple and depend on the company's history as well as the market condition. Professionals in the market, e.g. traders, whose aggregate behaviours (within the structure and constraints of the trading system itself) determine the stock valuation (see [10]), react both to the situation that a company faces and the reasons it gives for delisting. Finally the professionals filter this information through their own mental models (e.g.: delisting is bad).

Understanding the impact and uses of information regarding foreign delisting is therefore not only extremely useful for companies but also a complex topic that cannot be tackled by traditional financial methods or traditional qualitative methods but which requires a combination of the two.

The aim of our paper is to understand the foreign delisting phenomenon using both an event study analysis and an information content analysis on foreign companies withdrawing their stocks from the Tokyo Stock Exchange (TSE). Our objective is twofold: on one hand, we want to measure the impact of the delisting decision on the domestic stock price by observing the link between the pure value impact and both the reasons of delisting and the way it was publicly announced to the investors. On the other hand, we want to test the value adding power of qualitative research methods in finance. Our results show that companies delisting from the TSE market may loose value if the delisting decision is a consequence of the non fulfillment of the benefits of cross-listing or gain value if the delisting decision is presented as a part of a more general, coherent strategy of the company, e.g. focus on the domestic market only or wide cost cutting program. Hence, qualitative research methods provide useful guidance for a deeper understanding of the delisting process.

The paper is structured in the following way: firstly we present the literature review on the delisting phenomenon, secondly we present the multi-method research methodology and the sample, and thirdly we present the qualitative and quantitative analysis on the announcements and the stock value. The last part concludes and provides some indications on further possible developments.

2 Literature Review

Since this paper is concerned with the impact of information about foreign delistings on the domestic stock price, we limit our literature review to the empirical studies which address the question of the share reaction around a company's listing/delisting announcement, which also represent the majority of the literature in the field. However, four important studies that provide a more general treatment are worth mentioning. In [11], [5], [6] and [7] different ways in which cross-listing on multiple markets would lead to a higher equilibrium price and a lower expected return are presented.

The event studies measuring the impact of the listing decision on the market price basically fall in two categories: US companies listing abroad (see, among others, [12], [13], [14], [15] or [16]) and non-US firms listing on US exchanges (see for example [17], [18], [3], [19], [20], [21]). The results all point towards either slightly positive or neutral market reactions in the listing month following the event. Two other studies cannot be ignored in the field. For example in [4] a positive reaction for 183 Over-the-Counter listings, private placements and listings on the three US leading markets, i.e. NYSE, AMEX and Nasdaq, announced between 1985 and 1995, is reported. Moreover, the author also shows a significantly higher announcement-day price reaction for emerging market firms and a higher reaction for exchange listings. In [4], these results support the idea that the net benefits of cross-listings are essentially due to the decision to overcome investment barriers. The other leading paper in the field, [3], focuses on 183 ordinary and American Depositary Receipts listings. The results are slightly different from [4] in the sense that while they also find a significant positive listing week reaction they also record a pre-listing positive reaction and

a negative post-listing effect. These results were the same for developed and emerging markets. The provided explanation is not based on investment barriers arguments anymore but more on strategic timing decisions by managers and other diminished market incompleteness theories.

The literature on delistings is significantly scarcer. In [22] the first study on a sample of stocks delisted from the NYSE is conducted. The difference in price between the last quotation day on the NYSE and the first quotation day on the American Over-the-Counter market shows a significant decrease of 17%. In another article, [23], a significant drop of 8.5% upon announcement of the domestic delisting on a sample of 520 American stocks withdrawn involuntarily from 1962 to 1985 is recorded. Meanwhile, in [24] significant increases in both price and bid/ask spread for 47 companies moving voluntarily from the AMEX to the Nasdaq between 1992 and 1995 are found. A very complete paper - see [25] - examines the impact of involuntary exchange delistings on liquidity, trading costs and volatility of firms delisted from the NYSE which end up trading on the Over-the-Counter market. Finally, [9] and [26] include on firms that have decided to "go dark", i.e. deregister with the SEC and delist from the major exchanges despite having a large number of outside shareholders. The authors seek to understand the reasons behind this decision and to analyze the consequences for shareholders.

The paper about the effects of involuntarily foreign delistings from the US stock exchanges - see [27] - records a significant, permanent drop in prices of 4.5%. The sample consists of 103 foreign delistings from 20 different countries between 1990 and 2003. Finally, another interesting paper – see [28] - examines the impact of 45 US firms voluntary delisting from the TSE over a time window of 20 years, stretching from 1977 to 1997, and concludes that this leads to a significantly negative price movement both at the time of the delisting announcement and also around the date of the delisting. Hence, the authors suggest that the negative price response reflects both a temporary information effect and a more permanent valuation effect but which does not seem to be related to a decrease in liquidity. Our study goes along the line introduced by this paper. However, in our study we develop both the database, by including also non-US companies delisting from Tokyo over a more recent period going from 1994 to 2005, and the approach, taking into account, besides the announcement date, the exact content of this announcement in a multi-method research framework.

In a more general framework, some studies (see [29], [30], [31], [32] and [33]) show that the trading volume and the price reaction to an announcement are very complex. The various reactions will heavily depend upon the quality of the information contained in the announcement, namely its content, its global and public veracity, its timing, its acquisition costs and the degree of divergence in the opinion of the different investors.

In the financial markets microstructure literature, the information arrivals (e.g. announcements) lead agents to revise their expectations about the terminal value of a financial asset and then generate price fluctuations. However, traders interpret information in a heterogeneous way because they dispose of different mental models about economic fundamentals and private signals about market conditions or because they "agree to disagree" [34] about their interpretation. Consequently trading activities are tightly connected to the interpretation of financial news on a specific event.

Interpretation of information through mental models has been dealt with before in a variety of fields. A definition of mental models that connects information interpretation to financial markets is provided in [35]:

> *"Under conditions of uncertainty, individuals' interpretation of their environment will reflect the learning that they have undergone. Individuals with common cultural backgrounds and experiences will share reasonably convergent mental models, ideologies and institutions and individuals with different learning experiences (both cultural and environmental) will have different theories (models, ideologies) to interpret that environment. Moreover the information feedback from their choices is not sufficient to lead to convergence of competing interpretations of reality. In such cases, multiple equilibria will result." [35].*

While mental models emerge from experience and therefore should lead to better results, research shows that agents faced with situations slightly different from what they had previously experienced can fail to adapt and perform poorly. In studying traders' mental models the authors (see [36]) found that their capacity to interpret a situation with slightly different parameters from the ones they normally faced was very limited. Ultimately the brokers in Burns' study were outperformed by students that had similar technical skills but no previous experience to guide their actions. These findings are not dissimilar from the studies carried out by Weick on reaction to crisis situation [37] where in presence of a new situation, old mental models lead to negative (even catastrophic) rather than positive results.

In general traders may try to outperform the market trying to gather better information respect to the competition. Traders will also try to interpret information better than the consensus through better pattern recognition skills or more suitable mental models. Both in [36] and [37] it is shown that mental models are context dependent and therefore investors are likely to have developed them in the market reality in which they live. It is interesting to note that efficiency tends to prevail when agent errors are independent. In this framework, inefficiencies can therefore arise when errors are non-independent or in other words when people are all thinking the same way (leading to, for example, the presence of abnormal returns). When faced with critical changes in the market structure, shared mental models will direct behaviour towards consensus. Unconventional behaviour is possible, but it relies on interpreting information through modified mental models that are bound to emerge but with a certain delay respect to the change [37].

Besides being influenced by their mental models, traders will also filter information through them. Of the different kind of information that can be useful for trading there are some that are more easily found and some that are less. One key element of people's use of information comes from behavioural psychology that tells us that most people place the greatest weight on information that is accessible versus information that is most relevant [10]. Since business press articles, analyst reports, and investment banker pitches are ubiquitous, they are likely to be perceived to be the most valuable sources to make value estimations.

This means that it is primordial to include the study of the web of information surrounding and reaching the market operators before and after a critical juncture. This juncture was provided by the increasingly diminishing trading volumes that hit the foreign companies listed on the TSE in the years 1989 – 1994 where the volume decreased by a factor of 30. In a market that looked at delisting as "embarrassing" many companies saw delisting as the imperative to balance investment and returns from the foreign operation. Understanding the market reaction to this imperative and the impact of information on mental models becomes therefore the driver of this research.

3 Data and Methodology

The data and the methodology chosen depend on the phenomenon under investigation. In particular, our results are based on an event study where we will perform both quantitative and qualitative analysis, see [38]. To create the sample, the delisting announcement represents the event. In terms of analysis, the techniques used are content analysis for the announcements and cumulative abnormal returns for the effect of the announcements on prices. The sample and the methods are further explained in this section.

3.1 Sample Selection

The companies included in our sample are all foreign corporations listed on the Tokyo Stock Exchange (TSE) which decided to withdraw their stocks from the quote between 1994 and 2005. None of these companies were later relisted in Japan. The choice of the TSE is justified by the size of this market but also by the fact that it is the market that experienced the highest number of voluntary foreign delistings in the period considered, see [28].

Generally, one can distinguish four major categories of delistings: following a merger or acquisition, following changes in market standards or fall below continued listing standards, bankruptcy, and voluntary delisting reasons (strategic, focus on the domestic/other foreign market, etc.). As the price reaction following an involuntary delisting has become straightforward in the light of all the previous studies on the topic, in this paper we only focus on voluntary delistings.

The empirical material collected included both the announcement of a foreign delisting and the quantitative data about the stock price. The original list was provided by the TSE and consists of the effective companies delisting from this market in the period 1994-2005 and the exact announcement date. From this list we selected the foreign companies having delisted ordinary shares. Then we looked for the announcements communicated by the companies or appeared in the press and we scanned for announcements released up to 3 months before and 1 month after the delisting date to account for potential biases due to the presence of other relevant events in the event window. According to recent similar studies (e.g. [39]), information was obtained using multiple sources. Given the heterogeneity of the traders it was important to include equally heterogeneous sources of information. Specifically we scanned the companies' websites for press releases, and the databases Factiva and EBSCO Business Source Complete[1] The search line included only the company name in the specific time window. We were able to collect the exact content of the delisting announcement plus other announcements for 49 companies.

For those companies we collected daily prices for the stocks on the domestic market for a period comprised between [-198; +10] days around the event. We also collected daily prices for the domestic indices corresponding to each market. We use the DAX index for Germany, the ASX for Australia, the IBEX for Spain, the SMI for Switzerland, the TSX for Canada, the FTSE 100 for the UK, and the S&P500 for USA. All these data come from Datastream. By cross checking all the information we were able to gather our final complete sample (announcement + prices) of 49 companies

[1] Factiva and EBSCO index the vast majority of the business press including both generic publications and industry specific ones.

(4 German, 3 Australian, 1 Spanish, 3 Canadian, 5 British, 2 Swiss and 31 American). The detailed list of the companies appears in appendix A.

3.2 Methodology

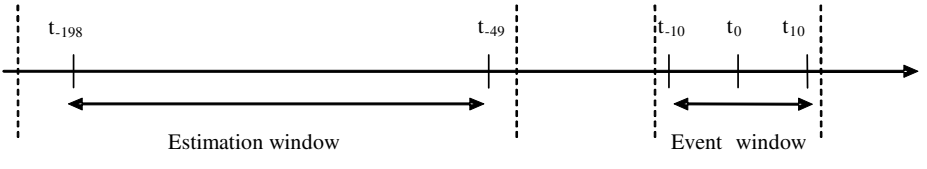
The research method included two parallel processes: the content analysis of the communications in the press and the quantitative analysis of the stock prices. The two processes were carried out separately not to influence the content analysis with the knowledge of the quantitative results.

Qualitative Part. We performed content analysis on the delisting announcements. Content analysis is a research method that is gaining increasing consensus in management research (e.g. [39]). Content analysis helps to classify and code the content of the selected document therefore allowing to creating clusters based on this content. We did not start from a pre-existing classification scheme but preferred to let the classification scheme emerge from the content collected. This would allow for the emergence of original categories. Therefore we first did the content analysis for voluntary delisting reasons, and then we re-coded the data for firm-specific communications.

We took several steps to prevent the data coding from being influenced by the researchers' biases. First, the coding was carried out by a researcher not used to work with financial topics. The particular statements that could catch the attention of the investors were also coded and classified. After this phase the coding scheme was developed. In a second step the announcements were re-coded by a second person, with the first coding scheme at hand, to test for stability and reliability. The independent coder had experience in investment management and traders' work. Following [39] we used Cohen's Kappa coefficient to test for stability and reliability. The Kappa coefficient was 0.84 allowing us to consider high agreement between coders. Examples of coding are presented in appendix B.

Quantitative Part. We measure the impact of a delisting from a foreign market by analysing the domestic market reaction at the public announcement of the decision to withdraw the stock. The abnormal returns with respect to the theoretical values which would have prevailed in the absence of the event illustrate the short-term gain/loss resulting from such a decision.

A very intuitive way to represent the main dates of an event study is as follows:



We run our estimations on a period of 150 days, ending 49 days prior to the announcement to avoid any interference between the estimation and the event window[2]. The event day $t_0$is the day the announcement of the delisting is made public. The delisting announcement is isolated from all other events that could also influence the market price (as dividends payment, profit warnings, other corporate actions).

[2] For more details see [40] among others.

Hence, the aim of an event study is to calculate abnormal returns (over the event window) of the different assets "i" concerned by the study and see if they are significantly different from zero or not. Abnormal returns for a stock "i" on day "t" are defined as follows:

$$AR_{i,t} = R_{i,t} - E(R_{i,t}) \tag{1}$$

Where
$AR_{i,t}$ = the abnormal return;
$E(R_{i,t})$ = the expected return of the asset "i" in t;
$R_{i,t}$ = the return of the asset "i" in t.

The daily returns on the domestic market are computed as logarithms of price changes, i.e. $R_{i,t} = \log\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$. The estimation of the theoretical expected returns are derived from the market model[3] (for a more detailed presentation of different approaches to perform an event study and their relevance for information systems and general management topics please refer to [41], [42], [43], [44], [45], [46], [47] among others). The market model supposes a linear relation between the return of an individual stock "i" and the whole market return:

$$E(R_{i,t}) = \alpha_i + \beta_i R_{M,t} + \varepsilon_{i,t} \tag{2}$$

With $R_{M,t}$ = the return of the domestic market index on the day t; $\varepsilon_{i,t}$ = the error term of the stock "i" on the day t, with zero mean and σ_ε^2 variance.

Hence, the abnormal return is the difference between the observed return on day t and the estimated return, over the estimation window, function of the parameters $\hat{\alpha}_i$ and $\hat{\beta}_i$.

The cumulative abnormal return corresponding to the event period takes into account the price changes during several consecutive days:

$$CAR_i(t_1;t_2) = \sum_{t=t_1}^{t_2} AR_{i,t} \tag{3}$$

The significance of the abnormal returns, i.e. $H_0 : CAR_t = 0$, is tested by using a parametric, Student test:

$$t - stat = \frac{CAR_i}{\sigma(CAR_i)} \tag{4}$$

For a T-day event window, $\sigma(CAR_i) = \left[T\sigma^2(AR_i)\right]^{1/2}$, where $\sigma^2(AR_i)$ is computed using information from the estimation window.

[3] A robustness check of the computations was performed by using the Capital Asset Pricing Model to derive theoretical expected returns. The results were significantly unchanged and are available upon request.

We also use the mean abnormal return on a day t, computed by using all the available returns on that day, i.e. I_t:

$$AAR_t = \frac{\sum_{i=1}^{I} AR_{i,t}}{I_t} \tag{5}$$

Its corresponding cumulative abnormal return over the event period will then be denoted $CAAR(t_1;t_2)$ and its significance is tested using exactly the same Student test.

4 Empirical Results

As stated in the methodology part, the delisting decision can be either dictated by the stock exchange itself and hence be involuntary or it can be voluntary initiated by the company itself. An involuntary delisting is expected to have a negative impact on the firm value whereas there is no ex ante expected effect in what concerns voluntary delistings. The only signal conveyed by the voluntary delisting decision is that the expected benefits of listing abroad were not fulfilled; hence, the stock price reaction to the announcement will depend on the exact content of the announcement.

4.1 Qualitative Results

The period 1994 – 2005 has been a very dynamic period for TSE with respect to foreign companies listed. The number of foreign companies went from 87 to 32 while the number of companies delisting peaked at 10 in 1995 (see Figure 1).

The content analysis of the delisting announcements shows the emergence of a rather wide variety of reasons for delisting. The following announcement from the

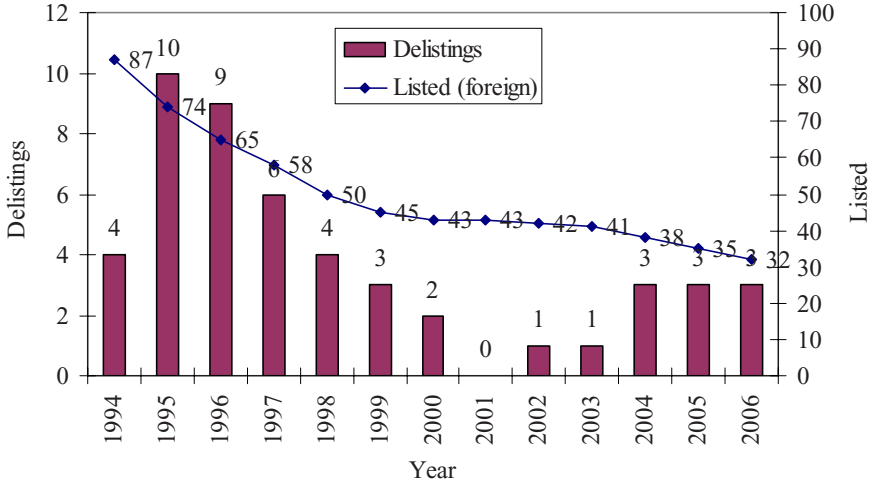


Fig. 1. Number of foreign companies delisting vs. remaining listed on TSE

firm Cordiant (owned by Saatchi and Saatchi Co.) captures very well the spirit of the announcements in the sample:

> *Tokyo, March 3 (Jiji Press)-Saatchi and Saatchi Co. said Friday it has decided to delist its shares from the foreign section of the Tokyo Stock Exchange. Saatchi and Saatchi saw no merit in keeping its shares on the Tokyo market because trading volume here has not met its expectations in terms of volume and value, and the firm has no plans to carry out equity finance programs in Japan in the future. The delisting from the Tokyo market is also part of the company's cost-cutting program, the company said. The pullout of one of the world's leading advertising agencies will leave 86 firms on the exchange's foreign section, which has seen an exodus of companies in recent years. Six companies have already applied for delisting from the TSE since the beginning of this year. (Jiji Press English News Service, March 3, 1995)*

The main reason that companies gave to abandon the Tokyo Stock Exchange was that the trading volumes were too low. Some companies limited the announcements to this brief statement while others, like Cordiant, added reasons (part of the company's cost-cutting program) that could influence positively the market. Cordiant, as well as other companies (e.g. Weyerhaeuser), stated that they did not have intentions to further carry out equity finance programs in Japan – something that instead could lead to a negative response. Many firms in the sample stated the cost of listing as a major reason for withdrawal. This is a statement from Commerzbank:

> *Commerzbank said the **high cost of preparing books and other statements in Japanese** was a major hurdle to maintaining the listing. Analysts said the average cost for foreign firms of listing in Japan is 14 million yen a year, 80 percent of which is used for translating financial statements into Japanese.*

This statement does not only show the specific reasons for delisting for Commerzbank but highlights a particular characteristic of the TSE and its openness/closeness towards foreign companies that appeared in the delisting-related news repetitively in the twelve years that we have observed.

The complete overview of the categories that emerged from the qualitative study is presented in Figure 2. Figure 2 shows the percentage of the firms that used a specific reason for delisting – the number near the label – as well as giving a graphical idea of the frequency with which a particular reason was given compared to the others. This is done because most companies gave more than one reason for delisting.

In particular we notice that the category of low trading volumes is used by almost 80% of the companies and it is therefore not a big differentiator for the stock performance. Other categories like *low number of shareholders in Japan* and *high cost of maintaining shares on TSE* are very common in the announcements – 35% and 47% respectively – and therefore can be used to create clusters for further study.

Further categories that emerged for creating clusters are for companies that announced *company wide cost cutting programs* (16%), companies that hypothesize that the *integration of financial markets* compensate for the lack of presence on different markets (12%), and companies that announced their wish to *delist from multiple markets* in a single move (10%).

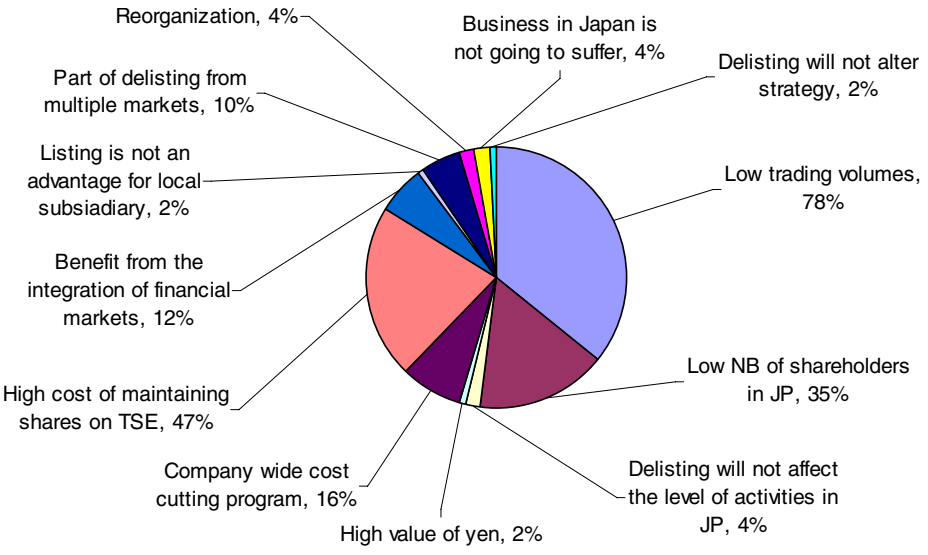


Fig. 2. Reasons for delisting

In particular, the integration of financial markets and the multiple delisting reasons were given by companies that delisted after 1997 and therefore are a recent phenomenon whose impact on the stock price is interesting to investigate.

The other categories emerged for a very low number of companies (less then 4%) and therefore are impossible to be treated from a statistical point of view. The qualitative study provided us therefore with 5 groupings of reasons:

- low number of shareholders in Japan,
- part of a company-wide cost cutting program,
- high costs of maintaining the listing on the TSE,
- multiple foreign delistings,
- financial markets integration.

The evolution of mental models and the particular situation of the TSE market, with more and more companies delisting because of the low level of trading volume (see figure 2), may have alerted investors that the benefits of listing on this market are not fully realized and this information may have been already incorporated into prices before the delisting announcement (this result is in line with [28]). If this statement is true and if the trading volume on the TSE is a good proxy for the benefits of foreign listing in this particular case, then the quantitative analysis would point towards a price response to the announcement neutral or even positive. If, on the contrary, investors maintain the working hypothesis that foreign listing is good and therefore discount the news that the expected benefits from the listing are not realized then we expect the price reaction to the announcement to be negative. Similarly, if the trading volume is not considered a good proxy of the benefits reported in the delisting announcement, than we expect the price reaction to the announcement to be negative or neutral.

The quantitative analysis following in the next section will show if these criteria, clusters and mental models, are significant in the response of the domestic stock market at the announcement of the delisting from the TSE.

4.2 Event Study Results

This section provides a summary of the quantitative analysis we performed on the domestic stock price of the companies that decided to delist from the TSE. We start by discussing the empirical results on the whole sample of 49 companies and then we analyse different clusters of companies created based on various reasons.

Empirical Evidence on the Entire Sample of Companies. The results of the event study analysis on the entire sample are provided in Table 1 and Table 2. One may notice that there is no clear, general response, either positive or negative, and the results are generally not statistically significant, which could mean that either the information was already incorporated in the prices or that the delisting announcement

Table 1. Average abnormal returns on the domestic markets around the announcement delisting dates

Day	AARs	T-stat	p-value signed rank test
-10	-0.1193%	-0.5789	0.6189
-9	0.0869%	0.4216	0.7654
-8	0.2226%	1.0801	0.6981
-7	0.3808%**	1.8475	0.2445
-6	0.2589%	1.2561	0.1668
-5	-0.0627%	-0.3044	0.1668
-4	-0.0850%	-0.4122	0.3706
-3	-0.3902%**	-1.8929	0.2287
-2	0.0741%	0.3594	0.6189
-1	-0.0920%	-0.4466	0.4862
0	0.2965%	1.4385	0.0967**
1	0.0316%	0.1536	0.7578
2	0.0184%	0.0892	0.9366
3	-0.3588%**	-1.7408	0.0750**
4	-0.3028%	-1.4689	0.9287
5	-0.0980%	-0.4757	0.4556
6	-0.0266%	-0.1289	0.6688
7	0.3824%**	1.8554	0.0350*
8	-0.4684%*	-2.2727	0.0456*
9	0.2749%	1.3337	0.2211
10	0.2474%	1.2005	0.1464

** denotes significance at the 10% level * denotes significance at the 5% level.

Table 2. Cumulative abnormal return on the whole sample of 49 firms around the announcement delisting dates

Horizon	CAARs	T-stat
[-10;+10]	0.2707%	0.2866
[-1;+1]	0.2361%	0.6613
[-5;+5]	-0.9689%	-1.4174
[-5;0]	-0.2593%	-0.5136
[0;+5]	-0.4131%	-0.8182

from the TSE market, given again the particular context of this market, is not considered as a major news by the traders on the domestic market. The signed rank test supports these conclusions. These results are in line with the ones presented in Table 2, which provides the cumulative abnormal return on the cross-section of our 49 companies, over the whole event window [-10; +10] and around the event date [-1; +1] and [-5; +5]. Again, no particular trend seems to emerge, even though over the entire event window there seems to be some positive effect on shareholder wealth consequent to the announcement, but again, not statistically significant.

In order to get some more highlights on this result, we plotted the average cumulative abnormal returns for the sample of delisting companies split into two subsets: one containing companies that announced their delisting before 1997 (included) and the other one including companies announcing their delisting decision starting with 1998. The choice of the partitioning year was dictated by the fact that the period 1994-1997

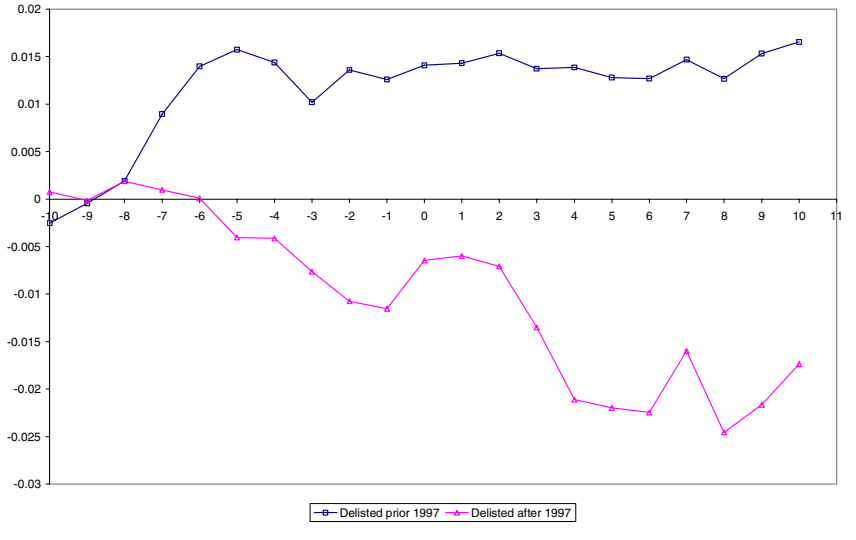


Fig. 3. Cumulative Abnormal Returns around announcement date for companies delisting before and after 1997

is reported in the literature as being the period with the largest number of delistings from the TSE market (see also [28]). Furthermore, the qualitative analysis also points out the appearance of some new delisting reasons quoted by companies in their delisting announcements after 1997. Figure 3 shows that while the average cumulative abnormal returns for firms announcing their decision to quit the TSE market before 1997 presents a gradual increase, the opposite applies for firms that publicly announced their delisting decision after 1997, i.e. gradual, highly non-linear decline. Potential explanations may be found on one hand in the effects of the Asian crisis that saw an important number of world financial markets plummeting. On the other hand, the TSE authority operated important changes in its trading rules starting with 1997, introducing more flexibility and significant trading cost reductions especially concerning the publication standards (e.g. translation of documents in Japanese). The reasons for voluntary delisting may have become less obvious in this new context.

To sum up, the quantitative evidence shows that firms delisting from Tokyo may gain or loose value on their home exchange market at the time of the delisting announcement which is not in line to what is previously reported, including [28]. These results are an indicator of the change in the mental models of the market operators that seem to have integrated the idea that delisting is not always a negative piece of news and its impact depends on the context. Since for market operators the context is created through the announcements, their exact content should provide an explanation concerning the change in stock value. Therefore, the next section will analyse the context starting with the clusters identified in the qualitative study and concluding with the traditional characteristics, i.e. home country and industry.

Empirical Evidence: Content of the Announcement Effect. In order to get a deeper understanding about the domestic price response to the delisting announcement, we develop our quantitative analysis on different sub-samples. Two categories of eventual groupings can be created. The first type of grouping is the result of the qualitative analysis on the exact content of the announcement as explained above, precisely regarding the reasons invoked for voluntary delisting from the TSE. These criteria are: low number of shareholders in Japan, part of a wide cost cutting program of the company, high costs of maintaining the listing on the TSE, multiple delisting (delisting from TSE occurs as the result of a delisting strategy from many markets), financial markets integration.

The second type of grouping includes two general characteristics: industry and home country[4]. Among the five criteria of classification identified through the qualitative analysis, three clusters show significant results.

Companies that justify their delisting decision by the low number of shareholders on the TSE market record significant, negative cumulative abnormal returns over the five days around the announcement and also over the five days after the announcement. Hence, the loss recorded due to the low liquidity of the Japanese market was not incorporated in the domestic stock price prior to the delisting announcement.

[4] Company's size is not a differentiating criterion as all the companies in our sample are big, international corporations.

Moreover, the number of shareholders seems to be a good measure for the benefits of cross-listing as its low level is penalized by domestic investors. The same applies for companies presenting the high cost of maintaining their stocks quoted in Japan as the reason for delisting; over the five days surrounding the announcement, these companies show significant, negative cumulative abnormal returns. Companies conclude that the benefits generated by being cross-listed in Tokyo do not justify the high costs related to the foreign listing and domestic investors penalize them for not fulfilling the cross-listing benefits. These results are in line with those reported by previous studies on delisting (e.g. [28]).

On the contrary, in the case of announcements focusing less on the context of the TSE market and more on the overall strategy of the company, the results concerning the impact of the delisting news are different. Investors on the domestic market seem to reward a strategy of wide cost cutting and even more, a strategy of concentration

Table 3. Cumulative abnormal return on firms clustered by the specific content of the announcement around the announcement delisting dates

	Horizon	CAARs	T-Stats
Low NB of shareholders	[-10;+10]	-2.1766%	-1.4805
	[-1;+1]	-0.6423%	-1.1558
	[-5;+5]	-2.4714%*	-2.3226
	[-5;0]	-1.0813%	-1.3759
	[0;+5]	-1.5803%*	-2.0110
Company wide cost Cutting program	[-10;+10]	2.1929%	0.8888
	[-1;+1]	0.5820%	0.6242
	[-5;+5]	0.8852%	0.4958
	[-5;0]	-0.6036%	-0.4577
	[0;+5]	1.6819%	1.2754
High cost of maintaining stocks	[-10;+10]	-1.6531%	-1.2941
	[-1;+1]	0.0947%	0.1961
	[-5;+5]	-1.8863%*	-2.0403
	[-5;0]	-0.2897%	-0.4243
	[0;+5]	-1.1015%	-1.6132
Financial markets integration	[-10;+10]	0.4202%	0.1405
	[-1;+1]	0.6557%	0.5803
	[-5;+5]	-1.9562%	-0.9041
	[-5;0]	-1.6749%	-1.0482
	[0;+5]	0.1648%	0.1031
Delisting from multiple markets	[-10;+10]	4.6942%**	1.8394
	[-1;+1]	1.5582%	1.6154
	[-5;+5]	3.1425%**	1.7014
	[-5;0]	0.8383%	0.6146
	[0;+5]	2.7373%*	2.0066

* denotes significance at the 5% level.

** denotes significance at the 10% level.

on the domestic market. For these companies we observe positive cumulative abnormal returns on the whole event window and posterior to the announcement, albeit not statistically significant. Companies that explicitly advertise that delisting from Tokyo is a part of a more general, coherent program of delisting simultaneously from multiple foreign markets record significant, positive abnormal returns on the whole event window, over the five days surrounding the event and the five days posterior to the announcement.

Finally, there seems to be no clear response of the domestic stock price to announcements that justify the delisting due to financial market integration, e.g. in a context of integrated markets, investment barriers are already removed and hence, there are no benefits of cross-listing to be expected. Companies using this argument in their delisting announcement experience both positive and negative abnormal returns over the periods surrounding the event date, but not statistically significant. This means that either the investors did not perceive this argument as relevant or they did not agree with it or that there is less interest in companies' decision to quit the Japanese market.

In conclusion, albeit with different outcomes, the exact content of the announcement plays an important role in the domestic stock price reaction to the delisting announcement.

Home Country and Industry Effects. When we take into account the nationality of the different companies, we notice a few interesting results. Basically, companies from countries with a very limited number of listings on the TSE market experience dramatic, statistically significant, price decreases. It is the case for BBV, the only Spanish company voluntary delisting from the TSE during the 1990s among the four Spanish firms that have been listed since the 1970s (two other firms delist as the result of mergers and one remains listed). It is also the case of the two Swiss companies that delist voluntarily out of the four listed; meanwhile, the two other Swiss firms disappear from the TSE market following mergers, leaving no other Swiss firm listed on the Japanese market. Moreover, the least developed market (Spanish) seems to penalize the most the delisted firm (BBV stocks record an abnormal loss of 10% during the event window and of 7% over the 10 days around the announcement date). A general negative impact, albeit less important and not statistically significant is also observed for Australian and German companies, for which the number of listings on the TSE was also rather limited during the last 30 years (6 and 10 respectively) but for which some delisted companies, following mergers for example, were relisted afterwards under the new name leaving this way Australian and German firms still quoted in Japan. Results are less clear for American, Canadian and British companies, more numerous on the Japanese market even after an important number of delistings, where the impact seems to be generally positive but not significant. One can then conjecture that listing in Tokyo still represents an important issue for investors on markets that record less cross-listings (e.g. Spain, Swiss) and are less important in terms of market capitalization, despite the decrease in the size and efficiency of this market, e.g. traders do carefully scrutinize the news related to listing abroad and consider delisting as a sign of failure in taking advantage of the benefits of cross-listing. On the contrary, on the biggest markets in terms of capitalization, investors are less interested in

following the benefits/loss of benefits of companies that decided and were able to be listed abroad.

Clustering by industry shows that the price response to the delisting announcement is generally positive for companies in the communication and chemical industries (with companies in the communication sector recording a significant gain of 17% over the event window). It is more controversial for beverage where abnormal returns, albeit non significant, are negative on the event window and the five days prior to and after the event date but positive during the +/-1 days around the announcement. Finally, the reaction is generally negative for the banking, automotive and electronic industries. The cumulative abnormal returns are negative and statistically significant for banks. A possible explanation may be that domestic investors fear that by delisting from Tokyo banks may loose the proximity with the Japanese financial market and hence, with potential customers. A deeper analysis would require more companies and information about the way traders in these different industries asses the delisting question.

Table 4. Cumulative abnormal return on firms clustered by home country

Country	Horizon	CAARs	T-Stats	Country	Horizon	CAARs	T-Stats
USA	[-10;+10]	1.1600%	0.9658	UK	[-10;+10]	3.4009%	0.9605
	[-1;+1]	0.3898%	0.8587		[-1;+1]	0.6472%	0.4836
	[-5;+5]	-0.5296%	-0.6092		[-5;+5]	-0.3205%	-0.1251
	[-5;0]	0.2304%	0.3588		[-5;0]	-0.3493%	-0.1845
	[0;+5]	-0.4567%	-0.7113		[0;+5]	1.0616%	0.5609
Germany	[-10;+10]	-0.3139%	-0.1141	Switzerland	[-10;+10]	-5.6939%	-1.3246
	[-1;+1]	0.2314%	0.2226		[-1;+1]	-3.2795%*	-2.0185
	[-5;+5]	-1.8543%	-0.9316		[-5;+5]	-7.2334%*	-2.3250
	[-5;0]	-0.1978%	-0.1346		[-5;0]	-7.3303%*	-3.1903
	[0;+5]	-1.1410%	-0.7761		[0;+5]	-0.6744%	-0.2935
Australia	[-10;+10]	-4.7132%	-1.3589	Spain	[-10;+10]	-10.2970%*	-2.2764
	[-1;+1]	0.4997%	0.3812		[-1;+1]	-1.1448%	-0.2593
	[-5;+5]	-1.6206%	-0.6456		[-5;+5]	-7.4202%*	-2.2666
	[-5;0]	0.7787%	0.4200		[-5;0]	-2.9162%	-0.6778
	[0;+5]	-2.3603%	-1.2731		[0;+5]	-5.1094%*	-2.1132
Canada	[-10;+10]	-0.8729%	-0.3283				
	[-1;+1]	0.5088%	0.5062				
	[-5;+5]	1.5698%	0.8157				
	[-5;0]	-0.6902%	-0.4856				
	[0;+5]	2.2366%	1.5736				

** denotes significance at the 5% level.*

Table 5. Cumulative abnormal return on firms clustered by industry

Sector	Horizon	CAARs	T-Stats	Sector	Horizon	CAARs	T-Stats
Banking	[-10;+10]	-4.5049%*	-2.4765	Communications	[-10;+10]	17.2217%*	2.9320
	[-1;+1]	-1.5787%*	-2.2962		[-1;+1]	3.1426%	1.4156
	[-5;+5]	-4.2176%*	-3.2035		[-5;+5]	8.3243%**	1.9582
	[-5;0]	-2.6672%*	-2.7431		[-5;0]	4.4852%	1.4286
	[0;+5]	-1.4899%	-1.5323		[0;+5]	3.9422%	1.2557
Beverages	[-10;+10]	-0.7055%	-0.2083	Electronics & Telecoms	[-10;+10]	-1.3413%	-0.5542
	[-1;+1]	0.8748%	0.6834		[-1;+1]	0.6466%	0.7069
	[-5;+5]	-1.0176%	-0.4151		[-5;+5]	-1.9160%	-1.0939
	[-5;0]	1.6619%	0.9180		[-5;0]	-0.8702%	-0.6727
	[0;+5]	-2.0471%	-1.1307		[0;+5]	-0.6346%	-0.4906
Chemical	[-10;+10]	2.2106%	0.9012	Automotive	[-10;+10]	1.6446%	0.5382
	[-1;+1]	0.9087%	0.9802		[-1;+1]	-0.0078%	-0.0068
	[-5;+5]	-0.0673%	-0.0379		[-5;+5]	-2.3019%	-1.0409
	[-5;0]	1.2163%	0.9277		[-5;0]	-0.3520%	-0.2155
	[0;+5]	-0.6422%	-0.4898		[0;+5]	-1.9284%	-1.1807

* *denotes significance at the 5% level.*
** *denotes significance at the 10% level.*

5 Conclusions

On the whole sample of companies, our results are consistent with the home stock market having already impounded the shortfall in listing benefits (e.g. low trading volume) or considering the delisting as stop loss decision due to the inefficiencies characterizing the market under study. Meanwhile, they may also suggest the lack of interest of domestic market operators regarding the foreign listing, either because the TSE market became less interesting or because delisting from TSE is not the major piece of information among all the information they receive and treat. In this respect, our results go in line with previous evidence on (foreign) delisting.

However, if we develop the analysis by classifying the delisted companies following different criteria, the results become less straightforward. First of all, the content of the announcement plays an important role in the domestic stock price reaction of the delisted firms. If the main reason invoked for quitting the Japanese market is linked to the inefficiencies characterizing the TSE, then investors on company's domestic market penalize the lack of benefit from listing abroad and the price reaction is significantly negative. If, on the contrary, the main reason for delisting is due to a more general, coherent strategy of the company, either of focusing on the domestic market or setting up a wide cost cutting program, then investors on the domestic market seem to value the decision of the company and the price reaction to the delisting announcement is positive. Second, our results also show that the nationality of the

company may also be important. Listing in Tokyo still represents an important issue for investors on smaller size markets, that also record less cross-listings. On these markets, investors consider delisting as a sign of failure in taking advantage of the benefits of going abroad, and the price response is therefore significantly negative. On the contrary, on big size markets, investors are dealing with a more important quantity of information, among which the delisting announcements of companies listed abroad are one piece of news among others and the price reaction remains insignificant and without particular pattern. Finally, when we take into account the industry, the results show that companies in some sectors, e.g. communication, experience important gains in value around the announcement date, whereas banks for example, record significant drops in stock prices around the same dates. The explanation we propose is based on the proximity banks need with a financial market for developing their business and hence, investors' fear that by quitting the Japanese market they might loose potential future business opportunities.

The results of the article, arguing against the basic conjecture that foreign delisting would always bring a decrease in the stock price, show the value of a multi-method research framework where the qualitative analysis provides both context specific clusters and behavioural explanations. Given the Japanese market characteristics as the major delisting market over the period under study, the TSE example indeed deserves generalization including analysis of delistings from other international financial markets.

One limit of this paper is the sample size and the dominance of American firms. Hence, one main further development of our work will be to extend the analysis on a sample of inter-European delistings, which would represent both the first study on this type of data and provide a unique opportunity to compare whether delistings affect shareholder value any differently depending not only on the reasons for delisting but also on the home/foreign market. The quantitative approach could also include the impact of delisting on the trading volume and the stock volatility on the domestic market.

From the qualitative analysis viewpoint we also intend to enrich the discussion by adding interviews with traders negotiating stocks issued by companies that decided to delist from foreign markets. This additional analysis will include information on the actual use and evolution of the mental models and also provide more precise information to set the event window for the calculation of the CARs. Moreover, distinguishing between cross-listed and cross-tradable stocks could also provide potential valuable insights on the delisting causes and consequences[5].

References

1. Karolyi, G.A.: Why do Companies List Shares Abroad? A Survey of the Evidence and Its Managerial Implications. Financial Markets, Institutions & Instruments 7, 1–60 (1998)
2. Karolyi, G.A.: The World of Cross-Listings and Cross-Listings of the World: Challenging Conventional Wisdom. Ohio State University (2004)
3. Foerster, S.R., Karolyi, G.A.: The Long-run Performance of Global Equity Offerings. Journal of Financial and Quantitative Analysis 35, 499–528 (2000)
4. Miller, D.P.: The market reaction to international cross-listing: Evidence from depositary receipts. Journal of Financial Economics 51, 103–123 (1999)

[5] We thank a discussant of the FinanceCom 2007 Workshop for this suggestion.

5. Alexander, G., Eun, C., Janakiramanan, S.: Asset Pricing and Dual Listing on Foreign Capital Markets: A Note. Journal of Finance 42, 151–158 (1987)
6. Eun, C., Janakiramanan, S.: A Model of International Asset Pricing with a Constraint on the Foreign Equity Ownership. Journal of Finance 41, 1015–1124 (1986)
7. Errunza, V., Losq, E.: International Asset Pricing Under Mild Segmentation: Theory and Test. Journal of Finance 40, 105–124 (1985)
8. Broberg, B.: Companies Claim There's Life after Delisting. Puget Sound Business Journal Seattle, September 21, 22 (20) (2001)
9. Marousi, A., Massoud, N.: Why do Firms go Dark. University of Alberta (2004)
10. Mauboussin, M.J.: The Invisible Lead Steer. Equity Research—Americas, Credit Suisse First Boston Corporation NI4450 (1998)
11. Stapleton, R.C., Subrahmanyam, M.G.: The Market Model and Capital Asset Pricing Theory: A Note. Journal of Finance 38, 1637–1642 (1983)
12. Lee, I.: The Impact of Overseas Listings on Shareholder Wealth: The Case of the London and Toronto Stock Exchanges. Journal of Business Finance and Accounting 18, 583–592 (1991)
13. Torabzadeh, K., Bertin, W., Zivney, T.: Valuation Effects of International Listings. Global Finance Journal 3, 159–170 (1992)
14. Varela, O., Lee, S.: International Listings, the Security Market Line and Capital Market Integration: The Case of U.S. Listings on the London Stock Exchange. Journal of Business Finance and Accounting 20, 843–863 (1993a)
15. Varela, O., Lee, S.: The Combined Effects of International Listing on the Security Market Line and Systematic Risk for US Listings on the London and Tokyo Stock Exchanges. In: Stansell, S. (ed.) International Financial Market Integration, pp. 367–388. Blackwell Publishers, Cambridge (1993b)
16. Lau, S., Diltz, D., Apilado, V.: Valuation Effects of International Stock Exchange Listings. Journal of Banking and Finance 18, 743–755 (1994)
17. Switzer, L.: The Benefits and Costs of Listing Canadian Stocks in U.S. Markets. In: Sarna, L. (ed.) Corporate Structure, Finance and Operations, pp. 241–254. Carswell Co., Toronto, (1986)
18. Alexander, G., Eun, C., Janakiramanan, S.: International Listings and Stock Returns: Some Empirical Evidence. Journal of Financial and Quantitative Analysis 23, 35–151 (1988)
19. Jayaraman, N., Shastri, K., Tandon, K.: The Impact of International Cross Listings on Risk and Return: Evidence from American Depositary Receipts. Journal of Banking and Finance 17, 91–103 (1993)
20. Viswanathan, K.G.: Listing in the U.S. Markets by Foreign Firms: Evidence on Return and Risk. Advances in International Banking and Finance 2, 99–113 (1996)
21. Ko, K., Lee, I., Yun, K.: Foreign Listings, Firm Value and Volatility: The Case of Japanese Firms Listing on the U.S. Stock Markets. Japan and the World Economy 9, 57–69 (1997)
22. Merjos, A.: Stricken Securities: Loss of an Exchange Usually Means Trouble for a Company and Its Shareholders. Barron's, 9 (March 4, 1963)
23. Sanger, G.C., Peterson, J.D.: An Empirical Analysis of Common Stock Delistings. Journal of Financial and Quantitative Analysis 25, 261–272 (1990)
24. Clyde, P., Schultz, P., Zaman, M.: Trading Costs and Exchange Delisting: The Case of Firms that Voluntarily Move from the American Stock Exchange to the Nasdaq. Journal of Finance 52, 2103–2112 (1997)
25. Macey, J., O'Hara, M., Pompilio, D.: Down and Out in the Stock Market: The Law and Economics of the Delisting Process. Financial Management, 2–52 (2004)

26. Leuz, C., Triantis, A., Wang, T.: Why Do Firms Go Dark? Causes and Economic Consequences of Voluntary SEC Deregistrations. Working Paper, University of Pennsylvania (2004)
27. Liu, S.: The Impacts of Involuntary Foreign Delistings: An Empirical Analysis (2004), http://ssrn.com/abstract=680641
28. Das, S., Saudagaran, S.M., Sinha, R.: An Empirical Examination of NYSE Stocks Voluntary De-listing from the Tokyo Stock Exchange. Review of Accounting and Finance 3, 47–63 (2004)
29. Kim, O., Verrecchia, R.E.: Market Reaction to Anticipated Announcements. Journal of Financial Economics 30, 273–309 (1991a)
30. Kim, O., Verrecchia, R.E.: Trading Volume and Price Reactions to Public Announcements. Journal of Accounting Research 29, 302–321 (1991b)
31. Kim, O., Verrecchia, R.E.: Market Liquidity and Volume around Earnings Announcements. Journal of Accounting and Economics 17, 41–67 (1994)
32. Harris, M., Raviv, A.: Differences of opinion make a horse race. Review of Financial Studies 6(3), 473–506 (1993)
33. Kandel, E., Pearson, N.D.: Differential Interpretation of Public Signals and Trade in Speculative Markets. Journal of Political Economy 103, 831–872 (1995)
34. Aumann, R.J.: Agreeing to Disagree. The Annals of Statistics 4, 1236–1239 (1976)
35. Denzau, A.T., North, D.C.: Shared Mental Models: Ideologies and Institutions, 47th edn., pp. 3–31. Blackwell Publishing, Kyklos (1994)
36. Burns, P.: Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction. Research in Experimental Economics 3, 139–157 (1985)
37. Weick, K.: The Collapse of Sensemaking in Organizations: The Man Gulch Disaster. Administrative Science Quarterly 38, 628–652 (1993)
38. Cooper, M., Dimitrov, O., Rau, P.R.: A Rose.Com by Any Other Name. Journal of Finance 56, 2371–2388 (2001)
39. Agrawal, M., Kishore, R., Rao, H.R.: Market Reactions to E-business Outsourcing Announcements: An Event Study. Information & Management 43, 861–873 (2006)
40. Peterson, P.: Event Studies: A Review of Issues and Methodology. Quarterly Journal of Business & Economics 28, 36–67 (1988)
41. Brown, S.J., Warner, J.B.: Using Daily Stock Returns: The Case of Event Studies. Journal of Financial Economics 14, 3–31 (1985)
42. Chatterjee, D., Pacini, C., Sambamurthy, V.: The Shareholder Wealth and Trading-Volume Effects of Information Technology Infrastructure Investments. Journal of Management Information Systems 19, 7–42 (2002)
43. Chatterjee, D., Richardson, V.J., Zmud, R.W.: Examining the Shareholder Wealth Effects of Announcements of Newly Created CIO Positions. MIS Quarterly 25, 43–70 (2001)
44. Dehning, B., Richardson, V.J., Zmud, R.W.: The Value Relevance of Announcements of Transformational Information Technology Investments. MIS Quarterly 27, 637–656 (2003)
45. Im, K.S., Dow, K.E., Grover, V.: Research Report: a Re-examination of IT Investment and the Market Value of the Firm - An Event Study Methodology. Information Systems Research 12, 103–117 (2001)
46. McWilliams, A., Siegel, D.: Event Studies in Management Research: Theoretical and Empirical Issues. Academy of Management Journal 40, 626–657 (1997)
47. Subramani, M.R., Walden, E.: The Impact of E-commerce Announcements on the Market Value of Firms. Information Systems Research 12, 135–154 (2001)

Appendix A. Companies and Dates of Announcements (by Home Country)

Company	Sector	Country	Date*
Pacific Dunlop	Chemical	Australia	21/09/1995
Fosters Group	Beverages	Australia	18/12/1998
National Australian Bank Group	Banking	Australia	15/08/2006
CIBC	Banking	Canada	24/10/1994
BCE Inc.	Telecom Service Provider	Canada	21/03/1996
Alcan Inc.	Metal	Canada	05/09/1997
BASF	Chemical	Germany	30/07/1999
Commerzbank	Banking	Germany	21/04/2004
DeimlerChrysler	Automotive	Germany	20/06/2006
Deutsche Bank	Banking	Germany	21/08/2006
BBV	Banking	Spain	18/02/1998
Nestle	Food	Switzerland	30/11/2000
Credit Swiss	Banking	Switzerland	21/08/2002
Standard Chartered PLC	Banking	UK	24/10/1994
Cordiant	Communication	UK	03/03/1995
BOC Group	Gas	UK	11/07/1996
GKN	Automotive	UK	27/02/1997
Glaxo Welcome	Pharmaceutical	UK	07/07/1999
PPG Industries	Chemical	USA	21/11/1994
Hewlett-Packard	Electronics	USA	30/12/1994
Amex	Banking	USA	24/03/1995
Ford	Automotive	USA	24/03/1995
General Electric Capital	Banking	USA	24/03/1995
Rockwell	Electronics	USA	21/07/1995
Knight-Ridder	Publishing	USA	25/08/1995
Walt Disney	Communication	USA	14/11/1995
Bankamerica	Banking	USA	24/11/1995
Transamerica	Investment fund	USA	21/03/1996
Dun and Bradstreet	Communication	USA	21/03/1996
Weyerhaeuser	Paper and Lumber	USA	22/05/1996
Exxon	Chemical	USA	26/09/1996
Brunswick Corporation	Marine	USA	24/10/1996
AT&T	Telecom Service Provider	USA	21/03/1997
Texas Instruments	Electronics	USA	20/06/1997
Anheuser-Busch Cos.	Beverages	USA	22/09/1997
Chrysler Corporation	Automotive	USA	22/09/1997
Occidental Petroleum	Chemical	USA	22/01/1998

3M	Chemical	USA	24/08/1998
McDonald	Food	USA	10/02/1999
ADM	Chemical	USA	07/09/2000
Eli Lilly and Company	Pharmaceutical	USA	13/03/2003
Procter&Gamble	Consumer goods	USA	30/03/2004
Apple Computer Inc.	Electronics	USA	17/11/2004
IBM	Electronics	USA	29/03/2005
Pepsico	Beverages	USA	11/04/2005
Motorola	Telecom Service Provider	USA	17/06/2005
Allied Signal	Automotive	USA	24/03/1995
Du Pont	Chemical	USA	27/10/1996
Northern Tel	Electronics	USA	27/10/1996

*Date indicates the date of the announcement.

Appendix B. Coding Examples

Supporting Information for delisting	**Example**
Low trading volumes	**AT&T:** *AT&T said the trading volume of its shares remains thin. Only 88,400 AT&T shares changed hands in 1996 in Tokyo, compared with 818.96 million shares on Wall Street.*
	BCE: *Spokesperson Don Doucette said only 250 BCE shares a day were traded on average in Tokyo last year, compared with 174,000 on the New York Stock Exchange, 137,000 on the Montreal Exchange and 599,000 in Toronto.*
Low number of Japanese investors	**BankAmerica:** *BankAmerica said a decline in the number of shareholders in the Japanese market boosted the costs of maintaining shares on the bourse.*
	McDonald's: *McDonald's decided to leave the Tokyo bourse since the costs required do not justify the continued listing, with the number of shareholders in Japan steadily declining to only 933 as of March 1998.*
No plans to carry out equity finance programs in Japan	**Weyerhaeuser:** *Weyerhaeuser said the company has decided to leave the exchange after it saw the numbers of listed shares, shareholders and trading volume decrease year after year. The company made the decision also as part of its cost-cutting efforts. Weyerhaeuser added it has no plan to raise funds through new share placements in Japan.*

Excessive listing costs	**BOC Group:** *BOC Group said the move was due to a decline in the number of Japanese shareholders since the listing was obtained in July 1986. "The number of Japanese shareholders has declined to an extremely low level," it said in a statement, adding this "has resulted in disproportionate costs in maintaining the listing and arising from the continuing administrative and compliance requirements".*
	Commerzbank: *Commerzbank said the high cost of preparing books and other statements in Japanese was a major hurdle to maintaining the listing. Analysts said the average cost for foreign firms of listing in Japan is 14 million yen a year, 80 percent of which is used for translating financial statements into Japanese.*
Market integration	**BASF:** *Shareholders won't be handicapped by the delisting because the shares will continue to trade on the major international exchanges in London, Frankfurt and Zurich, BASF said.*
	Brunswick: *Brunswick does not need a Tokyo listing because investors in Japan can trade its shares in the New York market.*
Company-wide cost-cutting efforts	**Alcan:** *As part of the company's overall cost-cutting efforts, Alcan decided to cut the cost of having its shares listed on the Tokyo bourse and European stock exchanges.*
	Occidental Petrolem: *Occidental Petrolem also plans to delist its shares on the Toronto, Pacific and Swiss exchanges as soon as possible, and from other foreign stock exchanges thereafter.It has decided to delist its shares from all exchanges except the New York Stock Exchange because the cost of maintaining other listings is not justified.*

Instruments for an Integrated Business Network Redesign in the Financial Industry

Falk Kohlmann[1], Stefan Reitbauer[2], Clemens Eckert[2], and Rainer Alt[1]

[1] University of Leipzig, Marschnerstrasse 31, 04109 Leipzig, Germany
{kohlmann,alt}@wifa.uni-leipzig.de
[2] University of St. Gallen, Mueller-Friedberg Str. 8, 9000 St.Gallen, Switzerland
{stefan.reitbauer,clemens.eckert}@unisg.ch

Abstract. Driven by factors like globalization, increased competition and declining customer loyalty the financial industry is facing a structural transformation. To focus on core competencies banks adjust their business models and reduce their degree of vertical integration by sourcing complementary activities. Operating in a changing market with regards to sourcing strategies, the (re)design of business model and thus also the business network is a main challenge for the financial institutions. This research paper introduces an integrated approach of business network redesign (BNR) for the financial industry that not only covers all three layers of business engineering – business model, process model and information systems – but also provides extensive methodological support. Present approaches either only partially cover these layers and thus show a restricted usability or lack methodological support so that they are merely applicable in practice. Since the research is still in progress, this paper concentrates on the core instruments used in the process of BNR. Thereby the redesign is discussed for the example of the investment process and thus presents the following fundamentals: a reference network, a reference process and a reference service map for investments.

Keywords: Value chain redesign, instruments for business network redesign, investment process, financial networks, reference process model, service map.

1 Introduction

1.1 Motivation

The main drivers of change in the banking industry are globalization, the ongoing innovation and the increase in market competition [1]. Due to the information technology (IT) today's customers are well-informed and able to access their financial asset management without the need of a nearby branch or consultancy [2]. This considerably lessens the customer loyalty – which has been one of the barriers of competition on the banking market for a long time. Although banking still is a highly profitable industry, the exemplified recent market developments have urged banks to abandon their "never change a running system" tactics. Whereas banks are usually not early adopters of the latest business-, process- or technology-trends [3], changing business environments

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 136–150, 2008.

have drawn the banks out of their comfort zone. Often compared with the industrialization process in the automotive sector in the early 1980s the majority of European banks are still highly vertically integrated. In Germany for example 96.7 percent of the 2.355 banks listed in 2003 were universal banks covering a full spectrum of banking services [4]. According to Lamberti [5] the bank of the future will not only differentiate by the scope of its service offering, but additionally by its (low) vertical integration. In order to become such a bank and to react appropriately to the market evolution, banks are intensely (re)defining their core competencies [6] and thus enhancing new business models [7] [8]. Now, focusing on differentiating activities requires increased networking, such as in the cases of the expert-to-expert banking of Credit Suisse or the service offerings of the BPO provider B-Source in Switzerland. Consequently market researchers (e.g. [7], [1]) predict the redesign of the whole banking value chain, in the shape of increased sourcing [9] activities as a valid approach of simplifying the banking business resulting in small, networked institutes concentrating on specific business functions.

To implement such business models and to re-engineer their business accordingly the companies have to ensure the alignment of their business strategy, processes and information systems being a key requirement for effective innovations [6]. In order to reach the promises of business network redesign (BNR) in terms of efficiency and business flexibility and thus enhancing the agility and performance of banks an alignment of network, process and service design is necessary [10]. Present approaches to BNR show significant shortcomings – especially concerning a broad coverage of redesign layers and methodological elements (combined as methodological support) [10]. Thus our research seeks to develop an appropriate method (cf. fig. 1) incorporating instruments, guidelines and techniques for networks, processes and service clusters. Within this research program this paper focuses on instruments for BNR exemplified for the investment process.

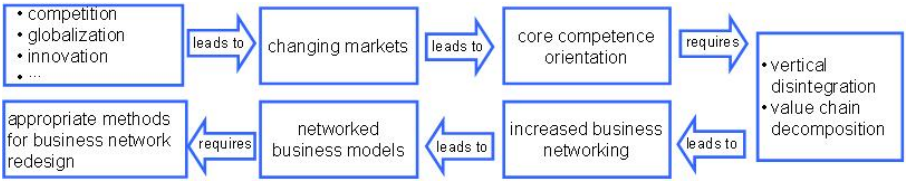


Fig. 1. Causal Chain of the need for a new approach for business network redesign

1.2 Methodology and Structure

This paper belongs to a multilateral, two-year research program that started in summer 2006 and investigates the management of flexible business networks in the banking industry succeeding a completed two-year research program about bilateral sourcing. Our 18 research partners cover various institutional sizes and roles in the banking value chain (e.g. regional retail bank, international private bank, outsourcing provider, software provider). Beside specific bilateral projects, the partners contribute to the research in biannual steering committee meetings and quarterly workshops supplemented by case studies and interviews taking place throughout the research program substantiating the applicability of the envisioned approach.

Within the research program the paradigm of "emphasizing collaboration between researchers and practitioners" [11, p.95] is followed to enhance the contribution magnitude to the practical concerns of financial institutions "and to the goals of social science by joint collaboration within a mutually acceptable ethical framework" [12, p.499]. Furthermore our research applies principles such as the Design Science guidelines proposed by Hevner et al. [13] to prevent critics by suspecting excessive practical influence. Following Design Science [13] the focus of our research is a methodological approach for BNR representing the artifact and implying financial networks, business processes and service clusters. Core is the alignment of different instruments and procedure models resulting in a vertical consolidation of the design elements, which can enhance the ongoing structural transformation and the emergence of new business models. A unified methodological approach for BNR on all layers has not been reached yet, even though BNR has been in debate for several years. The proposed approach entwines current models by incorporating the concept of service-oriented architecture as instrument of fostering the required flexibility, alignment and standardization of IT landscapes as well as by implying different sourcing strategies and models. By combining and extending existing methods from the organizational, strategic management and information systems literature as well as from the practice research rigor is followed. In this paper the proposed instruments are exemplified by using the investment scenario providing financial network, business process and service clusters for an in-depth analysis.

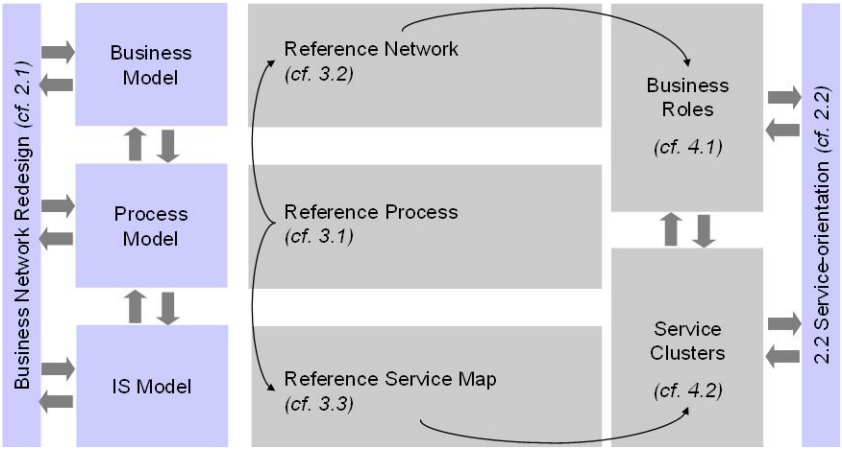


Fig. 2. Allocation of instruments and layers (following [6, p. 16]

The structure of the paper (cf. figure 2) reflects its goals:

- Developing instruments for BNR addressing network, process and service layer following the theoretical foundation. For all three layers reference models have been deducted from desk research and validated in interviews with practitioners and workshop sessions with the projects research partners.
- Examining the derived approach by applying business roles and service clusters to the case of Vontobel.

2 Foundation

This section primarily discusses the current state of research in the field of business network redesign and thereby highlights distinct approaches and concludes by a list of requirements to an appropriate BNR method. The second subsection deals with a specific aspect of BNR namely the role of service oriented architectures as enabler of the ongoing value chain redesign.

2.1 Business Network Redesign

Already Venkatraman [14] conceived the redesign of (external) business networks (BNR) as logically next step after the redesign of cross-functional processes inside an organization in the diction of business process redesign (BPR). Information technology and systems are important drivers towards increased networking and a vertical dissection of a value chain. The development of the Internet, powerful standard software packages (as ERP or core banking products) and emerging technologies (e.g. Web Services) facilitate inter-organizational collaboration [10]. Malone et al. [15] describe the impact of IT on the institutional form by the electronic communication, brokerage and integration effect.

Still, BNR is not primarily a technological issue, as the transformation towards more networked structures in the course of value chain redesign requires an alignment of these technologies with strategies and processes [10]. According to Alt [10] a complete BNR methodology should meet the following requirements:

- The methodology should be relevant to practice.
- It should cover redesign of strategy and processes as well information systems.
- It should provide detailed methodological support.

Models are important instruments for reducing complexity and distinguishing various elements on several interconnected layers. The enterprise modeling approaches follow this principle. For example Multi-Perspective Enterprise Modeling (MEMO) emphasizes a technique that includes internal and external actors as well as the processes and the involved IS [16]. All layers aim at a high level of formalization and show a close link to software engineering methodologies. A similar but more object-oriented idea is behind the Semantic Object Model (SOM) [17]. It aims at building IS by covering aspects of an inside, an outside, and a resource-view. Relationship patterns explicitly define the interrelationships between the model layers. As MEMO and SOM, the Architecture for Integrated Information Systems (ARIS) (e.g. [18]) focuses on the business process layer and its translation into IS. Business Engineering (BE) is a less formalized technique which as well recognizes the business process as main lever of change and therefore key element in shaping future business solutions and the underlying IS [19]. Similar to other approaches of enterprise modelling (cf. [20]) BE mainly distinguishes a strategic, process and systems layer. The key idea of BE is to enable the systematic development of future business solutions. By an alignment of strategy and systems to the process, the BE approach assures consistency across the three layers. Therefore the latter has been used in this research to structure the different instruments along the three layers: strategy, process and (information) systems. Services and service clusters are regarded as an intermediate layer between the process and systems layers, providing the link between business processes and

applications. Figure 2 indicates the allocation of the three instruments as well as the roles and service clusters referred to in the case study.

2.2 Service Modeling and Service Oriented Architecture

Following the tradition of object- and component-oriented architecture models, the service-oriented architecture (SOA) concept promises to improve the integration of heterogeneous application environments as well as the sourcing of entire or fractional business processes in a business network by combining individual application components. In many contributions and discussions SOA is attributed a 'silver bullet' status to reach these goals. However, SOA originally received as a technological concept needs adequate integration into the business world.

Though SOA can foster a transformation process within an enterprise or a network, SOA has like many 'magic words' numerous different definitions. For example, SOA is conceived in a technical view as a "paradigm that supports modularized exposure of existing application functionality to other applications as services" [21, p.41]. However service-orientation from a business view denotes the ability of reusing tasks and processes by solving them at one location. The business value of SOA in addition to the technical value concerning interoperability improvement, platform independence and reduced technical dependency can be ascribed to flexibility and profitability drivers. Both are the basis for numerous benefits frameworks [22]. In order to reduce suspected complexity a classification framework for SOA and services should be provided. Services in the context of BNR are understood in this paper as "independent usable and extensive specified functional components, which support the value performance of process activities" [23].

Core element of any SOA is specified services, which can be identified in general by two approaches: technical-driven service modelling (bottom-up) and business-driven service modelling (top-down). Service identification and design is therefore considered as a key step of establishing any SOA. Services identified in a top-down approach are mainly used when understanding SOA as a concept of connecting business and technology. Based upon the analysis and decomposition of business processes or business events [24], service candidates are identified by applying widespread design principles of SOA (e.g. [25], [26]) such as loose coupling, modularity, business orientation and interface orientation. Core element and basis for the bottom-up approach of service modelling is existing applications. A key step within provided process models is the analysis of currently existing applications and their IS functionality [21] as foundation for systems reengineering [27]. Researchers of bottom-up service modelling are focusing on consolidating and rationalizing access to IS functionality by using services. For a combination of bottom-up and top-down the term hybrid can be used [24]. A comparison of different identification strategies has indicated a lack of methodologies especially in combining business-driven service identification and clustering with network design within an engineering framework [23].

By applying the service orientation paradigm to BNR the restricted networkability of software communities [28] such as Avaloq or Finnova in Switzerland can be exceeded and cross-software community networking can be enhanced. Furthermore by applying the service design principles, mentioned above, on the strategy layer, business models of different business roles such as the integrator can be fostered. An example would be the DZ bank as integrator for the cooperative pillar in Germany.

3 Instruments for Business Network Redesign

The discussion about value chain redesign in the banking sector until now remains rather fuzzy. This section now outlines the situation and clarifies the networking / souring potentials in the investment process

- by presenting a reference process model as base for understanding and discussing the flow of a securities transaction,
- by deducting a reference network model (strategic layer) that enables the analysis of distinct networks as well as the determination of a company's position and
- by outlining a reference service map containing all services enabling / supporting the execution of a securities transaction.

According to the approach of BE the process layer is the linking element of business and IS / service layer. Thus the first subsection presents the reference process.

3.1 Process Model: Reference Process for Investments

A literature review for a description of an investment process results in a variety of proposals, but none is detailed enough to highlight the important aspects of sourcing decisions (e.g. the responsibility of the administration of pending orders) a company has to face when redesigning their business network. Thus the authors developed a reference process (cf. fig. 3), following the approach of design science, in cooperation with the business partners (cf. section 1.2) and in comparison to alternatives in theory (e.g. [29], [5]) and practice (e.g. [30]). This approach ensures not only the integrity of the resulting reference process but also its relevance for practical use. In advance the research team formulated several requirements for the reference process as e.g.:

- The reference process has to cover the complete procedure of an investment transaction independent of the investment instrument (e.g. share, structured product). Thus, available channels and instruments have to be considered.
- The level of detail in the reference process has to enable sourcing discussions, i.e. show a clear separation of the utilization of a service and its delivery.
- Dependencies to other processes have to be educible. In a bigger research context that is achieved by relating the process to a comprehensive banking process model.

The reference process as shown in figure 3 enables the design of an economically reasonable networking between players that execute subsequent sub processes towards a common service delivery. As the aim is to incorporate the channels and instruments as origin for the enabling of accurate sourcing decisions, the modeled level is quite high. For example the sourcing of process steps *A* (Oder initiation) and *B* (Order entry) for the paper-based order channel is rather complex and requires a high setup effort. For the electronic order channel however it is highly automated and rather simple (e.g. no signature checks). Nevertheless the process can be detailed in activity diagrams implied in the design tool associated with our research program.

The process contains seven macro process steps (cf. *A* till *G* in figure 3). These steps represent the execution of an order from its entry until its processing. Each step is detailed into sub processes whilst staying generic above all security instruments. The cross-order process steps *H* to *U* are divided into transaction related processes

and transaction spanning processes. The transaction-related steps *H* to *K* are connected to the execution or result of a single transaction whereas the transaction-spanning steps base upon an aggregated view.

The reference process is intended to be used as methodological support to business network redesign and will enable a company to map their process steps to the reference process in order to discuss process gaps and possible business cuts for sourcing options. Furthermore it provides a consistent way to exemplify an investment process supporting network partners to discuss and analyze their network architectures by standardizing their language to a common and reusable model.

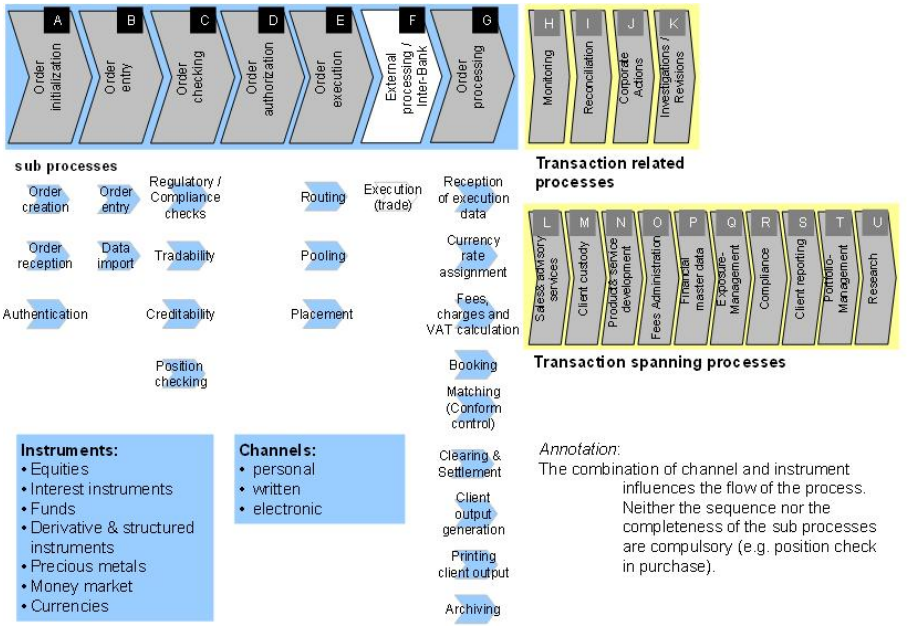


Fig. 3. Reference Model for the Investment Process

Combining the reference process with the reference network, dedicated process steps can be assigned to each player/role of the reference network. Using our BNR approach a company will therefore be able to identify their actual position in their own business network as well as their desired role by comparing their own processes with those of the according roles of the reference network.

3.2 Business Model: Reference Network for Investments

The next step after designing the reference process is to define a generic role model. This artifact aims to support the design of flexible business models based on the individual sourcing of services. Thus it dissects the process into the smallest functionally reasonable roles. Literature provides several approaches for restructuring a value chain (e.g. [31]). According to Alt [10] a segmentation of competencies

should consider operation (back-office and infrastructure), product development and customer relationship management as three distinct areas. The role model in figure 4 elaborates these ideas and dissects the reference process in a way that each role represents an isolate, reasonable business model. It therefore considers specialties of the investment process (e.g. differentiation of banking customer and market / exchange side) as well as theoretical network models (e.g. [32]).

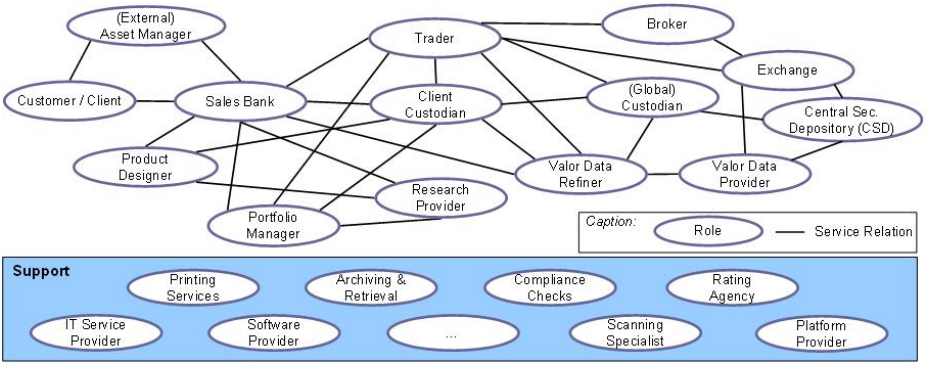


Fig. 4. Reference Network for the Investment Process[1]

Table 1 contains a description of the activities of each role in figure 4 as well as a mapping of the according steps in the reference process to the network roles. The column "Examples"[2] cites examples on the Swiss market for each role.

3.3 Information System Model: Reference Service Clusters and Service Map

Being able to identify the actual position and the desired role within a network via the assignment of dedicated process steps fosters the strategic flexibility of businesses. However the business adjustments need to be supported by flexible information systems architecture in order to originate maximal benefits. As alteration is initiated on the process and strategy layer, service-orientation used in a BNR methodology should be business-driven. Furthermore as numerous services increase complexity, service classification and structuring methods need to be applied. Besides a service classification scheme is needed. Following prior research a layer model based (see [23]) comprises (1) process services which support activities of the core processes of a company and include some references to at least one activity of a business process, (2) rule services which encapsulate business and validation rules used by process services, and (3) entity services which encapsulate core entities and business objects, such as contract, partner or order, the pattern approach can be applied to enhance structure.

[1] The roles marked as "Support" at the bottom of figure 4 represent functions that are important for the value creation in the investment process, but are not specific for this business area. They are considered as enablers and are therefore not part of the role description in table 1.

[2] The companies cover the relevant processes but are not limited to a single role. The Swiss market is currently not fragmented to an extent that allows pure examples for each role.

Table 1. Roles in the Value Chain of the Investment Process

Role	Description	Examples
Sales Bank (Process steps: A, B, D, H, K, L, N, O, Q, R, S)	Focuses on the customer relation, namely sales and advisory services. The sales bank acquires products and processing services from specialists in the network. Differentiators are price, product and service range as well as quality of consultancy.	Bank Linth, Clientis Banken, Bank Reichmuth
Asset Manager (A, B, D, H, L, Q, S)	External asset managers purely concentrate on the customer relationship and only administer client data. As asset managers have no banking license, the customer additionally has to be client of a (sales) bank providing account and custody keeping.	MLP, AWD
Portfolio Manager (B, D, H, Q, T)	The management of client portfolios does not require direct contact to the customer. Essential are knowledge of market evolution and availability of good research data. Differentiator is performance.	Investment Center SGKB, MBC
Product Designer (N)	Provides innovative products, often white-labeled. Differentiators are speed, performance and creativity.	Bank Wegelin, Vontobel
Research (U)	Offers market insights concerning single companies, commodities, financial markets and national economies.	ZKB, UBS, Credit Suisse
Valor Data Provider (P)	Provides access to comprehensive data for securities administration, risk management, portfolio management and trading departments.	Telekurs, Reuters
Valor Data Refiner (J, P)	Valor data have to be filtered and enriched in order to meet individual requirements. As these are similar for groups of banks a concentration can create synergies (fostered by the use of common IT systems).	BEKB, Comit with Fin-Log, Accenture
Client Custodian (Back-Office Specialist) (C, G, H, I, J, K, M, R, S)	Covers all administrative tasks of order processing as well as custody of client accounts including corporate actions, settlement of transactions and investigations (all except for client contact). The client custodian is in most cases responsible for the operative link to the trader and the (global) custodians. The question of who has the contract with the latter depends on the business model.	B-Source, Vontobel, RBA Service, Maerki Baumann, Sourcag
Trader (E, F, H, I, J, O)	Receives approved orders from the sales bank or the client custodian and is then responsible for routing and placing the order. The role represents the central counterparty (CCP) for sales banks and client custodians and establishes the link to several exchanges either direct or via brokers. The trader provides reconciliation with custodians.	Maerki Baumann, Vontobel, Lombard Odier, Credit Suisse, UBS
Broker (F)	Provides access to an (non electronic) exchange and offers know-how about local specialties.	Swissquote
Central Securities Depository (F, I, J)	Administers all securities traded on (a) distinct exchange(s). The Central Securities Depository (CSD) just keeps relationships to institutes that are under regulatory control– no direct customer link.	SIS
(Global) Custodian (F, I, J)	Holds an account at one CSD at least and offers financial institutions to store securities. The service range also includes administration (e.g. corporate actions, legal reports) and complementary services (e.g. money market). The custodian is responsible for the reconciliation of its clients' transactions with the CSDs.	UBS, Credit Suisse, SIS, BNP Paribas, Citibank

Pattern such as design pattern or architectural pattern are broadly used to structure e.g. communication elements or software systems in object-orientation. Service clusters and service maps ensue this pattern paradigm by structuring services. Classifying and consolidating services within a service cluster reduce the complexity arising from SOA implementations and provide new potentials for designing business models by using specified interfaces and a common methodology. Applying the process model for business-driven service identification and clustering as discussed in [23] to the field of investment and the above presented business process result in 64 services, which can be composed to 19 service clusters shown in the service map in figure 6. The obtained service clusters provide a direct link to processes, roles and sourcing strategies through abstraction and composition of services with high functional and semantic proximity. They can be used to describe and exemplify the scope of business of each business partner in a network with an altering environment. Through the incorporation of business-oriented services the adjustment of the current position in a network becomes more effective as IT-business alignment and application landscape has gained flexibility.

As the service cut is based upon sourcing models, business processes and business roles the incorporation of legal requirements such as customer data access increases further the reusability of the specified services by enhancing the ability to support diversified business strategies (scope and scale) to the same extent. The analysis of business networks is enriched by the embodiment of used services and service clusters. The instrument of the service map incorporates the services, the composed service clusters and its relationships. Simultaneously it affiliates the approach of reducing complexity by structuring the services clusters in three domains: execution competence, transaction spanning competence and support.

4 Application of the Instruments in a Case Study

This section focuses on the instantiation of the generic instruments presented in section 3. Hereby the authors chose the case of Vontobel, a Swiss private bank, which nowadays offers comprehensive banking services, also. The role model as well as the simplified service map is based upon the reference process as indicated in figure 3. As therefore both instruments incorporate implicitly the business process, the case study will focus on the network layer in subsection 4.1 outlining the strategic position of the company in the investment process and the service layer in subsection 4.2 describing which services VONSYS is providing to its clients.

4.1 Business Model / Network Layer

In the last two years the private bank Vontobel intensively extended its existing collaboration in investment fund products with the Swiss association of Raiffeisenbanken (SVRB). Now Vontobel provides a broad range of services in the investment process and thereby enhances the market presence of SVRB by professional services (e.g. products, sales training). For example Vontobel acts as official portfolio manager for the Raiffeisen Classic Portfolio and thus creates direct customer value at SVRB. In the sense of reciprocity Vontobel profits from the exclusive access to the broad sales channel (SVRB has the tightest network of branches in Switzerland) and economies of scale e.g. in client custody. The collaboration was enabled by the complementary business

models: SVRB focuses on sales in the retail segment while Vontobel is specialized on wealth management and private banking. Thus, the collaboration fosters the specialization of SVRB on sales activities and enhances the core business of Vontobel.

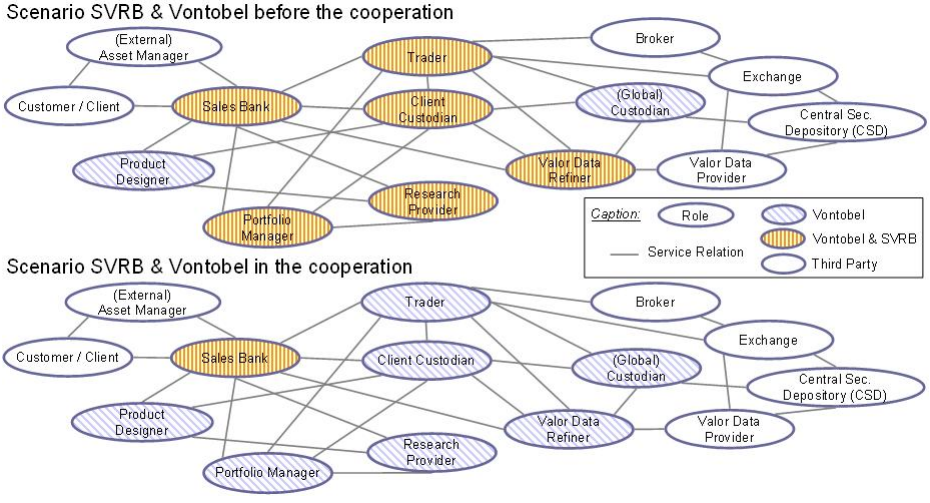


Fig. 5. Customized Role Model for the Cooperation Vontobel – SVRB

Figure 5 clarifies that the cooperation enabled SVRB to focus on its core competencies around the customer relation. The comparison of the two snapshots highlights that both partners could reduce significantly the redundancies in their investment processes. Since Vontobel as a private bank has to cope with more sophisticated processes and products in the securities business, the requirements of SVRB could be rather easily covered. The trader is modeled in blue because Vontobel executes all the trades for SVRB. Even though, SVRB still performs proprietary trading via the Vontobel network. Roles modeled in white are performed by third parties only; some of the blue roles (assigned to Vontobel) are shared with third party providers (e.g. research provider and product designer).

4.2 IT / Service Layer

As its host system is not multi-client capable, Vontobel introduced the third-party platform Avaloq for the client-custody services. Because their own customers are still administrated on a host system, Vontobel currently runs the client-custody processes on two separate systems. The planned consolidation of the platforms is expected to create significant synergies. Trading and global custody are already executed on a single platform (OTMS of IBM) for both partners. The consolidation of the two brokerage and custody networks has created significant cost savings, even though a few duplicate broker relations are retained due to reciprocal business deals (e.g. research for brokerage purchase).

Vontobel intends to acquire further customers for their back-office and investment service offering. The multi-client capability of the introduced Avaloq system is an

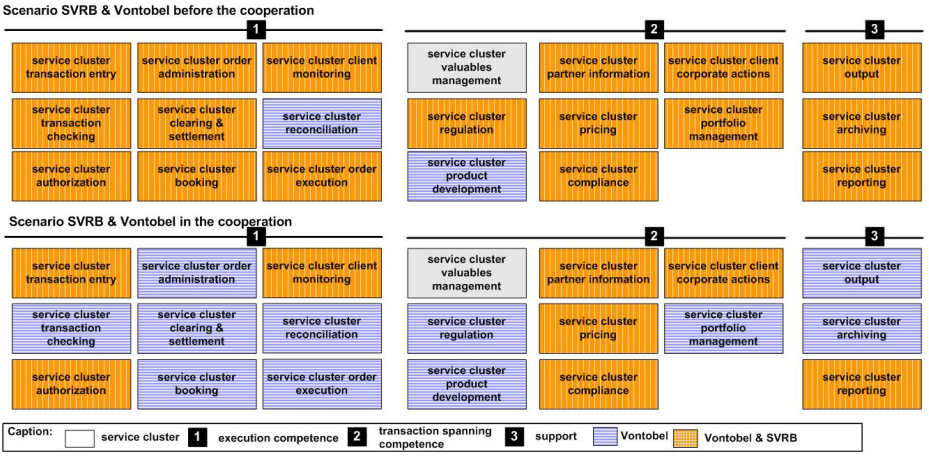


Fig. 6. Service map for the Cooperation Vontobel – SVRB

example how a standard software package can enable new business models. Furthermore, it can be a competitive advantage to have the service offering already implemented on this system, as Avaloq is a popular software platform in Switzerland, Since the IT services and the maintenance of systems like Avaloq are not part of the core business of Vontobel, the private bank decided to establish a partnership for the securities processing called VONSYS with the IT service provider T-Systems. Thereby Vontobel only cooperates with T-Systems in the IS maintenance but not in the execution because the processing of sophisticated products demands for excellent banking know-how. The snapshots of the service map with its service clusters shown in figure 6 indicate the competence orientation already mentioned in chapter 4.1. The 19 presented service clusters imply 64 specified services on the three service layers referred to in chapter 3.3 and were identified according to the procedure model in [23]. The service cluster execution contains e.g. the process services: execution service, placement service and pooling service, underlined by the routing and pooling rule services as well as by the order data service, providing the transaction data.

The instrument of the service map provides a more detailed outline of the service offerings than the roles and processes. It allows to analyze the as-is or to-be positioning of Vontobel within the network by linking strategy towards IS. The methodology developed in the research program implies furthermore techniques for mapping the service clusters towards implemented application or IT services on a finer granularity e.g. within the Avaloq system. The service map provides therefore simultaneously the interface between the roles with its contained business functionalities and the encapsulated IT functionalities of the Avaloq platform.

5 Conclusions

Swiss banks are currently facing a fundamental transformation towards more networked structures (cf. section 1). In order to reach high efficiency and flexibility the redesign of networks should be supported by an integrated methodology implying

procedure model, guidelines and instruments. Though BNR is in debate for several years a holistic methodology has not been reached yet (cf. section 2.1). This paper introduced different instruments on the three layers of Business Engineering: strategy, process and systems, as part of such a holistic methodology (cf. section 3). Coinstantaneous the service-oriented architecture concept has been integrated as it promises to enhance the required flexibility of the application landscape. The application of the service orientation paradigm fosters the core competence orientation in the field of the loans or investment process, besides highly standardized processes like payments. Therefore the instrument of the service map has been deduced (cf. section 3.3). As the network model and service map is based upon business processes, a reference process for the investment scenario has been differentiated (cf. section 3.1). The case study VONSYS has been used to apply network model and service map (cf. section 4) in a before and afterwards scenario.

The instruments provide a common understanding on the different layers concerning the positioning within a network as well as the service offerings exemplified in the service map. Using the reference network model, an enterprise is able to identify its actual position in a business network as well as its desired role by comparing it to the according roles in the reference network and their characteristics. The value creation of each role can be further analysed by using an assessment model, which differentiates qualitative and quantitative assessment [33]. Meanwhile the assessment model, presented in [33], is enhanced by allowing the network oriented assessment including the service oriented dimension. A consistent integration of design and analysis on one side and assessment on the other side is to a specific extent already reached within the methodology and will be deepened in future research.

The service map allows mirroring of the network to the IS layer and thus realize the consequences of the redesign on the IT landscape. A more detailed link to underlying systems and data structure can be reached by applying a mapping technique affiliating services provided in the service map and services on a finer granularity representing implemented functionality.

Further research should focus simultaneously on the application of the proposed instruments to further business processes in the finance industry and on the detailed formulation of the mentioned process models, guidelines and techniques. The holistic methodology will be based upon a meta model, which is in progress and will be presented in one of the next papers. Moreover we will dare to apply the service-orientation paradigm towards the strategy layer by converging business roles and service clusters as already indicated in figure 2 aiming at a consistently enterprise architecture. By the time we will apply the proposed instruments to further industries in order to provide a generalized methodology.

References

1. Geiger, H., Hürzeler, H.: The Transformation of the Swiss Private Banking Sector. Journal of Financial Transformation 9, 93–103 (2003)
2. Homann, U., Rill, M., Wimmer, A.: Flexible Value Structures in Banking. Communications of the ACM 47(5), 34–36 (2004)

3. Ferguson, T.: Banks ignore tech at their peril, silicon.com Financial Services (02.03.2007) (2006), http://www.silicon.com/financialservices/0,3800010322,39162817,00.htm
4. Lammers, M., Loehndorf, N., Weitzel, T.: Strategic Sourcing in Banking – A Framework. In: Proceedings 12th European Conference on Information Systems, Turku (2004)
5. Lamberti, H.-J., Poehler, P.: Die Industrialisierung des Backoffice am Beispiel der etb. In: Lamberti, H.-J., Marlière, A., Poehler, P. (eds.) Management von Transaktionsbanken, Springer, Berlin (2004)
6. Oesterle, H.: Business in the Information Age, Berlin etc., pp. 13–18. Springer, Heidelberg (1995)
7. Geiger, H.: The future of Switzerland as a banking centre, realtime. The international news magazine of Telekurs Financial 3, Zuerich (2004)
8. Accenture: Industrialization: The pathway to high performance banking, Accenture, Frankfurt (2006)
9. Hirschheim, R., Dibbern, J.: Information Systems Outsourcing in the New Economy. In: Hirschheim, R., Heinzl, A., Dibbern, J. (eds.) Information Systems Out-sourcing: Enduring Themes, Emergent Patterns and Future Directions, pp. 3–23. Springer, Berlin (2002)
10. Alt, R.: Business Network Redesign – Overview of Methodologies and Example of Process Portals. In: Markus, L.M., Grover, V. (eds.) Business Process Transformation. Advances in Management Information Systems, M.E. Sharpe (2006)
11. Avison, D.E., Lau, F., Myers, M., Nielsen, P.A.: Action Research. Communications of the ACM 47(1), 94–97 (1999)
12. Rapoport, R.N.: The Three Dilemmas of Action Research. Human Relations. 23(4), S.499–513 (1970)
13. Hevner, A., March, S., Park, J., Ram, S.: Design Science in Information System Research. MIS Quarterly 28(1), 75–105 (2004)
14. Venkatraman, N.: IT-Enabled Business Transformation: From Automation to Business Scope Redefinition. MIT Sloan Management Review 35(2), 73–87 (1994)
15. Malone, T., Yates, J., Benjamin, R.: Electronic Markets and Electronic Hierarchies. Communications of the ACM 30(6), 484–497 (1987)
16. Frank, U.: Multi-Perspective Enterprise Model-ing (MEMO). In: Proceedings 35th Hawaii International Conference on System Sciences (HICSS), IEEE, Los Alamitos (2002)
17. Ferstl, O.K., Sinz, E.J.: Modeling of Business Systems Using the Semantic Object Model (SOM), Bamberger Beiträge zur Wirtschaftsinformatik, 43 (1997)
18. Scheer, A.-W.: ARIS – Vom Geschäftsprozess zum Anwendungssystem, 4th edn. Springer, Heidelberg (2002)
19. Oesterle, H.: Enterprise in the Information Age. In: Oesterle, H., Fleisch, E., Alt, R. (eds.) Business Networking: Shaping Collaboration Between Companies, Berlin etc., pp. 17–53. Springer, Heidelberg (2001)
20. Janssen, W., Steen, M.W.A., Franken, H.: Business Process Engineering versus E-Business Engineering. In: HICSS 2003: Proceedings of the 36th Hawaii International Conference on System Sciences, p. 135c (2002)
21. Nadham, E.G.: Seven Steps To a Service-Oriented Evolution. Business Integration Journal, 41–44 (2004)
22. IBM Global Business Services (eds.): Service-oriented architecture: A Practical Guide to Measuring Return on that Investment, New York (2006)
23. Kohlmann, F., Alt, R.: Business-Driven Service Modeling - A Methodological Approach from the Finance Industry. In: BPSC 2007: Proceedings of the 1st International Working Conference on Business Process and Services Computing, Leipzig, pp. 180–193 (2007)

24. Klose, K., Knackstedt, R., Beverungen, D.: Identification of Services – A Stakeholder-based Approach to SOA Development and its Application in the Area of Production Planning. In: ECIS 2007: Proceedings of the 15th European Conference on Information Systems, St. Gallen, pp. 1802–1814 (2007)
25. Baskerville, R., et al.: Extensible Architectures: The Strategic Value of Service-Oriented Architecture in Banking. In: ECIS 2005: Proceedings of the 13th European Conference on Information Systems, Regensburg, Regensburg, pp. 761–772 (2005)
26. Papazoglou, M.P., Georgakopoulos, D.: Service-oriented Computing. Communications of the ACM 46(10), 25–28 (2003)
27. Zhang, Z., Liu, R., Yang, H.: Service Identification and Packaging in Service Oriented Reengineering. In: SEKE 2005: Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering (2005)
28. Alt, R., Smits, M.: Networkability of Organizations and Business Networks. In: ECIS 2007 Proceedings of the 15th European Conference on Information Systems, St. Gallen, pp. 119–130 (2007)
29. Schrauth, K.-D.: Die Erwartungen einer Landesbank an die TxB. In: Lamberti, H.-J., Marlière, A., Poehler, P. (eds.) Management von Transaktionsbanken, Springer, Berlin (2004)
30. Middendorf, J., Goettlicher, C.: Visionen für den Transaction Banking-Markt, BearingPoint, Frankfurt (2004)
31. Hagel, I.J., Singer, M.: Unbundling the Corporation. HBR 77(2), 133–141 (1999)
32. Miles, R.E., Coleman, H.J., Snow, C.C.: Organizational Dynamics, 20 (3) (1992)
33. Alt, R., Reitbauer, S.: Towards an Integrated Architecture and Assessment Model for Financial Sourcing. In: Proceedings of the 2nd International Workshop on Enterprise, Applications and Services in the Finance Industry, Regensburg (2005), Regensburg, 25.05.2005, pp. 67–74. IEEE, Los Alamitos (2005)

Explaining the Adoption of Value Metrics in Retail Banks' Customer Management

Sven C. Berger and Sebastian F. Martin

E-Finance Lab Frankfurt and University of Frankfurt, Mertonstraße 17
60325 Frankfurt, Germany
{sberger,smartin}@efinancelab.de

Abstract. A competitive customer management system requires the use of adequate metrics for performance controlling and management. This is of particular relevance for industries focusing their value creation on marketing and sales as it is the case in retail banking. Customer Lifetime Value (CLV) and related metrics are well-accepted solutions in scientific literature, but the diffusion into bank practice has not fully taken place yet. Adoption modeling can help explaining the use of CLV and identify major factors of influence for its adoption. In analogy to the Technology-Organization-Environment (TOE) framework which structures influential factors for the adoption of technological innovations, we developed an Innovation-Organization-Environment model, transferring the TOE approach to the adoption of a generic innovation. The competitive pressure is identified as a dominating driver, but perceived ease of use is also of high relevance, whereas a negative link from firms' profitability to the adoption of CLV is revealed.

Keywords: Adoption Research, Customer Lifetime Value, Customer Management, Retail Banking.

1 Introduction

The past decade has seen rapid and substantive changes in the way the customer relationship is dealt with in financial services [1]. Customer Relationship Management (CRM) is widely in use nowadays and the relevance of value-based customer management has increased significantly (see [2] for literature review). As the controlling and managing of any activity requires specific and adequate metrics, there is a strong need for metrics related to the value of the customer base.

At the individual level, Customer Lifetime Value (CLV) is considered being the focal construct of customer management [3]. The understanding of the CLV concept is a prerequisite for managing customer management processes [4].

Still, these metrics are not widely in use at the moment and financial services providers are reported not to have an accurate sense of customer value [5]. Existing studies investigating the usage of these customer value metrics mostly do not analyze causal effects; they just focus on the description of usage frequencies [6, 7] or analyze the adoption of CRM as a whole [8]. Thus, the factors influencing the CLV adoption

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 151–164, 2008.

in retail banking should be analyzed, in order to provide a basis for further research aiming to improve the way such metrics are introduced, implemented and communicated to facilitate a more comprehensive adoption within retail banking. Therefore, our research question is:

What are the major drivers for CLV adoption in banks?

To approach this question, after a short description of the object of research (section 2), we review current literature related to organizational adoption of innovations (section 3), develop a causal model and derive hypotheses based on prior literature on innovation adoption (section 4). We describe our methodology (section 5), then analyze the model using Partial Least Squares (PLS) method and data from European retail banks (section 6) and conclude with a discussion and the limitations of our research (section 7).

At the bottom line, we can reveal a significant impact of competitive pressure on the adoption of CLV in retail banks. Moreover, quite surprisingly, we find that a bank's profitability negatively impacts the adoption of the CLV.

2 Customer Lifetime Value (CLV) Concept

In this section we introduce the concept of Customer Lifetime Value and give a brief overview on related research. The ongoing trend towards value orientation in companies is reflected in every aspect of companies, as it is the case in customer relationship management [e. g. 9, 10-12]. Investments into such processes should be evaluated according to its value contribution [e. g. 13, 14], involving strategic decisions on business process modeling and IT investments [e. g. 15], but also tactical decisions of the day-to-day business [16].

CLV has been introduced as the value of the customer relationship to the firm in monetary terms. The CLV integrates assumptions about the future stream of income from a customer, the appropriate allocation of costs to customers, the discount factor and the expected retention of a customer [17]. It is computed as a customer's sum of current and future cash flows [18] or profits and discounted using a specific discount rate (e. g. the weighted average cost of capital [WACC]). The knowledge of the individual customer's CLV enables solving typical business decision problems regarding targeting and acquisition, segmentation and resource allocation. On an aggregate level, the total of all individual customers' CLV can be aggregated to the customer equity which provides the management with a sound understanding of the long-term value of the customer base [19, 20]. The CLV is therefore considered the "focal construct" of customer management on the individual level [3].

Nevertheless, little has been published on the adoption of CLV, and especially not in retail banking, which is one of the industries predestined for a valuation of customer relationships, due to its contractual customer relationships [21].

The vast amount of research regarding CLV focused mainly on methods for computation and evaluation on the one hand and on the impact of strategies and tactics on CLV, but not on the adoption and usage of the metric itself. CRM concepts in financial services were studied by Karakostas et al. [8] or Ryals et al. [22], without specific consideration of value metrics. The dissemination of customer metrics is considered

by Reinecke [6] or Reichold et al. [7], not focusing on adoption issues. Wiesel et al. [23] studied the reporting of customer value metrics in investor relations communication in Europe. They find a low usage in financial services also indicating, that there is not an industry-wide consensus of the importance of these metrics.

Even though this field is of certain theoretical and managerial significance, there is not a clear understanding of factors which determine banks to adopt CLV. This is why we address this specific issue by using an "innovation adoption lens". For the adopting bank, introducing CLV is associated with important organizational challenges [24]. This implies not only changing existent and introducing new reporting and management processes, but also adopting the necessary technology, like the software required for the computation of CLV, integration with the existing data warehouse infrastructure, implementation into controlling tools and employee training. Thus, while studying CLV adoption from an innovation perspective is an appropriate approach for coping with the above challenges, this article focuses – as a first step – on the determinants of CLV adoption. The adoption process itself is not covered by our study.

As CLV incorporates not only a performance figure, but also allows a paradigm shift in the way the customer is dealt with, and therefore has a major impact on management practice, marketing strategy and organizational behavior.

3 Related Literature

To analyze the adoption of CLV, one can draw from the vast literature on innovation adoption especially in the field of information systems. A majority of adoption research literature copes with the acceptance of innovations by individual persons, whereas the organizational adoption is less common [25]. Nevertheless, theoretical aspects from the individual perspective can be, and have been, transferred to decision-makers or managers of small businesses or specific departments [26, 27]. The Technology Acceptance Model [28] is a well-known model for the individual perspective based on the Theory of Reasoned Action [29], explaining the individual adoption by two constructs, the Perceived Ease of Use and the Perceived Usefulness. Both constructs have been used to study adoption of innovations by both individuals [30] as well as organizations [31]. In order to legitimately use innovation characteristics like Perceived Usefulness and Ease of Use in an empirical study of the organizational context, one has to make sure that the people questioned are the ones responsible for making the original and any subsequent decisions to adopt the innovation [31].

While the individual adoption research is dominated by innovation properties and the environment, the organizational adoption research also needs to take properties of the adopting organization into consideration [32]. In their thorough literature review, Jeyaraj et al. [33] have found that organizational factors like organization size and top management support range among the best predictors of organizational innovation adoption.

Another perspective has been added to the innovation and the organization perspectives by Tornatzky et al. [34], who added environmental perspective and established the so-called technology-organization-environment (TOE) framework, which serves as an important framework for predictive studies of organizational innovation

adoption. Thus, the framework identifies three categories of factors which have a major influence on the adoption of innovations by organizations. First, characteristics of the innovation, second, aspects describing the organizational context of the adopter and third, factors describing the environmental context as the external setting consisting of customer, competitors or regulators [34]. The framework has been tested broadly in information technology adoption [see 35 for literature review]. For example, Kuan and Chau [36] implemented this to explain the adoption of EDI technology. The usefulness of the framework for understanding the diffusion of complex IS innovation was shown [37] and the transfer of the TOE framework to other innovation domains was suggested as 'one future line of research' [38, p. 17]. A thorough review of factors belonging to one of the categories above, which have been used in IT innovation adoption research, is given by Jeyaraj et al. [33].

As this framework covers aspects similarly relevant for the adoption of value metrics, we make use of it to identify the specific contextual variables for CLV adoption, thereby transferring it from the technology background to a generic innovation adoption.

4 Research Model and Hypotheses

In this section, we develop our research model, covering aspects of the innovation itself, the organization and the environment.

4.1 Innovation

The introduction of the CLV concept is a major milestone in the development of marketing metrics [39-41]. The CLV allows the transition from the traditional product-centric to a customer-centric approach, enabling a new paradigm for the way the customer relationship is dealt with, respectively the mere enabler of understanding a customer relationship in a quantitative way [3]. This is especially the case for retail banking [21, 42].

The CLV was considered necessary for the implementation of competitive customer management processes [4] and was attributed the ability to generate competitive advantage [43]. It therefore qualifies as an innovation according to Rogers [25, p. 8], who states that innovation is *"[…] an idea, practice, or object that is perceived as new by an individual or other unit of adoption"*.

In a technological context, the vast quantity of studies following Davis' technology adoption model [for a recent review see 44], identified relationships between the constructs of *Perceived Usefulness (PU)* and *Perceived Ease of Use (PEU).* As a metric like CLV is typically applied on different managerial levels, business units or individual levels [11], its application follows a similar logic as the introduction and use of a technological innovation.

In contrast to technological innovations, metrics do not just serve to support a specific task, but are embedded into a broader system of metrics to enable a multidimensional performance measurement aligned to superordinate aims [45]. Thus, there is a need to test in how far the CLV fits into this system and is relevant for the management of the business unit *(Perceived Importance).* This Perceived Importance is obviously influenced by the Perceived Usefulness. We therefore hypothesize:

H1a: Perceived Usefulness of CLV has a positive impact on Perceived Importance of CLV.
H1b: Perceived Importance of CLV has a positive impact on adoption.
H1c: Perceived Ease of Use of CLV has a positive impact on adoption.

4.2 Organization

As Gopalakrishnan et al. [46] show, the impact of firm size on innovation adoption led to contradictory results in research. Whereas one argumentation proposes that larger companies are more successful in adopting innovations due to superior access to resources [47], others argue that flexibility is beneficial for smaller companies [48, 49]. We follow the first argumentation, as size is important because larger organizations tend to be more functionally differentiated, with a greater variety of specialized tasks, influencing the adoption of innovation [50]. Even more important, the benefit of advanced customer management systems incorporating CLV measurement grows with the number of customers [4]. Furthermore, for the specific context of CLV adoption, flexibility plays only a minor role, as performance management systems and metrics are usually not changed within a short time.

H2a: The size of the bank has a positive impact on CLV adoption.

There is a similar argumentation for the relationship between firm profitability and (technology) adoption. As more profitable firms can allocate more available resources to the adoption, the adoption process will speed up and be more successful [51]. Hence,

H2b: Profitability has a positive impact on CLV adoption

4.3 Environment

The environmental context of retail banks typically consists of customers, (product) suppliers, sales partners, competitors, regulators and investors. Customers are not involved in bank management at all (they are just the object of controlling activities) and, although customer value based metrics have a high explanatory power, regulators as well as investors do not show any interest at the moment [23] and can therefore be left out. As the vertical integration in banks is typically high [52], products are mostly produced in-house, and the customer relationship is being managed by the bank itself, which results in negligible demand for customer metrics from supplier or partner side at the moment, which is also induced by data protection regulations. The main environmental influence originates from competitors. The competitive pressure refers to pressure on using a technology or practice that has already been successfully adopted by competitors [53]. It tends to press companies to seek for competitive edge [54] and thereby makes the firms more innovative overall [55]. The competitive pressure in retail banking is high [56] and the customer lifetime value as a metric has the ability to generate competitive advantage [43]. Hinshaw [57] sees the implementation of customer valuation as a major success factor and expects the financial providers to experience competitive pressure.

In analogy to [37] we define the competitive pressure as the percentage of companies in the peer group who have already adopted the innovation. The competitive pressure does not only affect the adoption itself, but also the attitude against the innovation. We therefore hypothesize:

H3a: *Competitive pressure has a positive impact on the perceived importance of CLV.*
H3b: *Competitive pressure has a positive impact on adoption of CLV.*

Figure 1 depicts our research model.

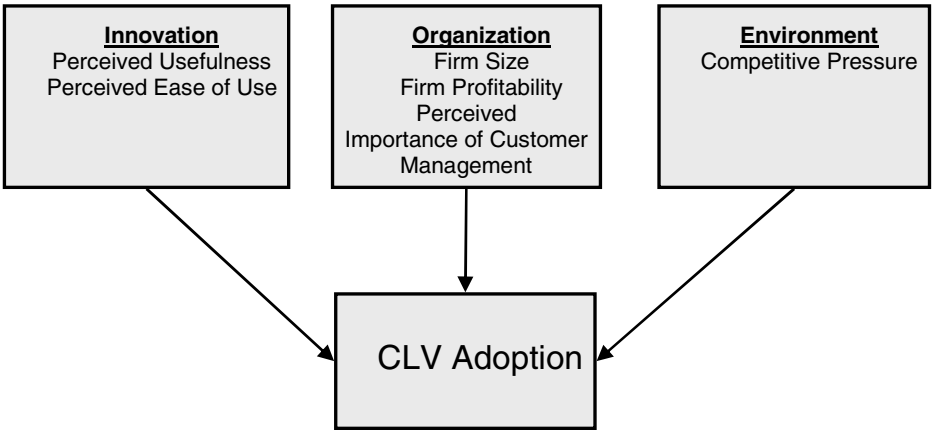


Fig. 1. Conceptual model for the adoption of CLV

5 Methodology

The constructs were operationalized using established indicators whenever possible. The constructs and the relevant sources are depicted in Table 1.

The constructs were used in the survey in a German version using the back-translation method [60]. When applying a positivist approach and testing a set of hypotheses in a structural equation model, a strong underlying theory is recommended [61]. Therefore, all hypotheses are well-grounded in the existing literature, most of them being grounded in the TOE framework.

To determine the adoption behavior by testing those hypotheses, an empirical study among retail banks in German-speaking countries has been conducted. The survey was conducted using a written questionnaire, distributed online as well as in paper form from May till June 2006.

The research model was operationalized as a structural equation model and estimated using the Partial Least Squares (PLS) approach [62, 63] with the software implementation SmartPLS [64]. In contrast to covariance-based approaches (e.g. LISREL), PLS has minimal demands on measurement scales, sample sizes and residual distribution [62], which makes it suitable for the analysis of a small sample of banks.

Table 1. Constructs and related indicators

Construct	Indicator	Loading	t-statistic
PU [28]	PU-1: Using the Customer Lifetime Value would enable to accomplish the company's aims more quickly.	0,868383	42,589495
	PU-2: Using the Customer Lifetime Value would improve firm performance.	0,860614	23,821281
	PU-3: Using the Customer Lifetime Value would increase sales productivity.	0,846621	14,147528
	PU-4: Using the Customer Lifetime Value would improve the ratio between resources and benefits.	0,798549	15,425134
	PU-5: Using the Customer Lifetime Value would make it easier to accomplish the company's aims.	0,878610	34,255960
	PU-6: I would find the Customer Lifetime Value useful for the company.	0,882714	32,309709
PEU [28]	PEU-1: The introduction of the Customer Lifetime Value would be easy for me.	0,890282	27,199328
	PEU-2: I would find it easy, to achieve the aims by using the Customer Lifetime Value.	0,831791	5,065857
	PEU-3: Using the Customer Lifetime Value would be clear and understandable.	0,811479	11,088420
	PEU-4: I would find the usage of Customer Lifetime Value flexible according to our application.	0,783673	14,528682
	PEU-5: It would be easy for me to become skillful at using the Customer Lifetime Value.	0,759518	10,381079
	PEU-6: I would find the Customer Lifetime Value easy to use.	0,730636	9,124180
Perceived Importance CLV	PC-1: Rank the metric Customer Lifetime Value according to its importance for your decision making in customer management at the moment.	0,897596	27,889340
	PC-2: Rank the metric Customer Lifetime Value according to its importance for your decision making in customer management in future.	0,886374	48,327931
Profitability [58]	PR-1: Return on Assets (RoA) [balance sheet data]	0,991	5,056
	PR-2: Revenues (per Employee) [balance sheet data]	0,758	3,365
Firm Size [59]	FS: Total Sum of Assets [balance sheet data]	*1,00*	—

Each behavioral construct in the research model is represented by a set of indicators being measured on a fully anchored 5-point Likert scale. PEU/PU constructs were measured using scales from "strongly disagree" to "strongly agree", relevance-related constructs ranged from "very low relevance" to "very high relevance". The questionnaire was pre-tested independently with three managers from different banks which have not been included in the final sample. Based on the insights acquired in these pre-tests, the questionnaire was modified and finalized.

The banking sector in German speaking Europe has been chosen as a research domain because of the substantial importance customer management plays in this over-banked and over-branched market [65].

6 Analysis

6.1 Demographics

In 2006, the questionnaire was sent out to the 430 largest retail banks selling to German speaking customers. 68 responding banks lead to a response rate of 18%, which consisted of 53 banks from Germany, 7 each from Swiss and Austria and 1 bank from South Tyrol/Italy (with German as the language of business). The distribution of responses amongst the banking groups (private banks, savings bank, cooperative banks, and other banks) matches the distribution in the sample, as does the size of the banks.

6.2 Model Validation

Measurement Model Specification
The complex behavioral constructs used in the model are derived from other studies (Perceived Ease of Use and Perceived Usefulness from Davis [28]) and adapted to the specific research domain. All constructs are measured in reflective mode as the indicators meet the criteria postulated in Jarvis et al. [66] for reflective measurement models. Indicators for the relevance constructs were adapted and validated accordingly. The adoption of CLV itself was measured as the mere existence of the metric in the organization and the diffusion across organizational levels and departments. Firm size was operationalized by the total sum of assets in the balance sheet, as one of the two proposed measures by Tschoegl [59] and the only one available for non-listed companies, which is reflected in many other studies [e. g. 67].

Firm profitability is measured by Return on Assets (RoA) and Revenues (per employee) as these measures are typically used in literature, [e. g. 58, 68, 69, 70]. To avoid any common method variance, data for these indicators was obtained from the individual published balance sheet and P&L to obtain measures of the predictor and criterion variables from different sources [71].

Reflective Measurement Model
The quality of the reflective measurement model is determined by (1) convergent validity, (2) construct reliability and (3) discriminant validity [72].

Convergent validity is analyzed by indicator reliability and construct reliability. In the model tested, all loadings are significant at least at the 0.1 level and above the recommended 0.707 parameter [62].

Construct reliability was tested using the composite reliability (CR), estimated indices were above the recommended thresholds of 0.6 [72].

Discriminant validity of the construct items was analyzed by examination of the cross-loading. The loadings of the indicators resp. Pearson's correlation of the specific construct are always higher with this construct than with others [73]. Furthermore, the Average Variance Extracted (AVE) is always higher than the recommended threshold of 0.5 [74].

Structural Model

In addition to the review of the measurement model, the explanatory power of the structural model is evaluated. The squared multiple correlation (R^2) for the dependent variable adoption of 0.261 indicates explanatory power, although it is rated "low" [62].

The Stone-Geisser test (Q^2), measuring the quality of each structural equation by the communal validity redundancy (cv-redundancy index), is positive and therefore the model has predictive relevance [75, 76].

Note that all paths' significance is above the criterion of 0.1 [77], the relevant paths are depicted in figure 2. The analysis of overall effect size (f^2) reveals that all constructs on significant paths have moderate or weak effects [62].

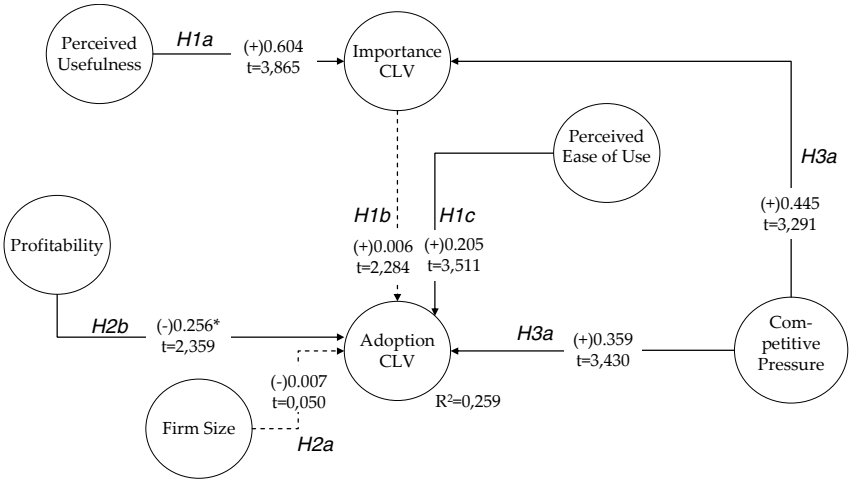


Fig. 2. Path model with significant relations

6.3 Results

We estimate the results using PLS, the results are shown in figure 2 to save space. We find that all aspects of our model have a certain influence on the adoption, although not all hypotheses on positive directions are supported.

Among the characteristics of the innovation, the relevance of CLV for customer management is explained by the Perceived Usefulness the CLV (H1a; path coefficient of 0.604). The adoption of CLV is explained by the Ease of Use (H1c; path coefficient of 0.205), there is not a significant effect of the CLV Importance (H1b). This may result in the interpretation that the participating marketing manager's view on the CLV Importance was not sufficiently taken into consideration when the metric was implemented or the view changed after the implementation.

Among the aspects of organization, the negative causal effect of profitability on the adoption of CLV shows that banks realizing a high profit do not value the CLV metric, which is considered useful for supporting management geared to long-term profitability. As the profitability measures, which are fully accepted in literature, are not explicitly long-term oriented (and in case of the revenues per employee rather

short-term), one may assume that profitable banks strongly prefer short term metrics, avoiding a long-term orientation in terms of performance measurement.

Firm size does not have any effect on CLV Adoption. This might be, on the one hand, a result of the nature of the business models implemented – as CLV is especially beneficial for the standardized high-volume business there is not a big functional differentiation between smaller or bigger banks in this business. Products as well as sales channels are rather similar. On the other hand, a possible explanation for the missing influence of firm size lies in the accessibility of the innovation. Implementing CLV requires certain effort and knowledge, but in contrast to some technological innovations, the capital investments and the necessary technical knowledge for implementation are limited. Therefore, there is not an extreme difference in the innovation accessibility for smaller and bigger banks. All in all, both hypotheses H2a and H2b can be rejected.

Regarding the environmental influence, we find a very high impact of competitive pressure on the adoption of CLV with a coefficient of 0.359, which makes it the most important factor influencing the adoption. It has an even higher influence on the perceived CLV Importance with a coefficient of 0.445. Thus, the hypotheses H3a and H3b are supported.

As all aspects related to innovation, organization and environment have significant (albeit not always positive) impact on the adoption of CLV, we regard this model as useful for explaining the adoption of this innovation.

7 Discussion / Limitations and Conclusions

Using data from a survey in German-speaking countries, our analysis reveals a significant explanatory power of the aspects innovation, organization, and environment to explain the adoption of the CLV metric in retail banking. The study clearly shows that Competitive Pressure is the factor with the highest influence, followed by Perceived Ease of Use and Profitability (with a negative effect).

This supports the introduction of value-based customer metrics in so far, as companies as well as applied researchers should concentrate on the Ease of Use of their customer value models and software implementations at first to gain acceptance in practice. The hesitant overall use of CLV and the high influence of Competitive Pressure reveal that where adoption has taken place, a competitive advantage is possibly implied.

The study is limited so far, as it includes only retail banks operating business in German language at the moment. Therefore the results cannot be generalized, as the adoption behavior may differ in other environments. For example, it has already been shown that cultural affinities have a higher impact on adoption behavior [78]. The model's robustness would be increased by also considering the effect of perceived pressure on CLV adoption as a mediating factor between actual peer pressure and adoption. This could be covered in a future survey.

The study contributes to the existing research by providing an explanation of retail bank's adoption behavior of value based customer metrics. This helps bridging the gap between the high acceptance of CLV in scientific literature and the less comprehensive acceptance in retail banking practice. The study is also of high relevance for industry practitioners as it supports decisions on how to communicate and develop customer value models and software implementations.

References

1. Peppard, J.: Customer Relationship Management (CRM) in Financial Services. European Management Journal 18, 312–327 (2000)
2. Payne, A., Frow, P.: A Strategic Framework for Customer Relationship Management. Journal of Marketing 69, 167–176 (2005)
3. Kumar, V., Lemon, K.N., Parasuraman, A.: Managing Customers for Value: An Overview and Research Agenda. Journal of Service Research 9, 87–94 (2006)
4. Jain, D., Singh, S.S.: Customer Lifetime Value Research in Marketing: a Review and Future Directions. Journal of Interactive Marketing 16, 34–46 (2002)
5. GartnerG2: GartnerG2 Says Financial Services Providers Must Calculate Customer Lifetime Value to Improve Profitability. Stamford, Conn (2003)
6. Reinecke, S.: Marketing Performance Management: empirisches Fundament und Konzeption für ein integriertes Marketingkennzahlensystem. Dt. Univ.-Verl., Wiesbaden (2004)
7. Reichold, A., Kolbe, L., Brenner, W.: Performance Measurement of CRM in Financial Services. In: Radnor, Z., Kennerley, M. (eds.) Performance Measurement and Management:Public and Private, pp. 859–866. Emerald Group Publishing, Edinburgh (2004)
8. Karakostas, B., Kardaras, D., Papathanassiou, E.: The state of CRM adoption by the financial services in the UK: an empirical investigation. Information & Management 42, 853–863 (2005)
9. Bowman, D., Narayandas, D.: Linking Customer Management Effort to Customer Profitability in Business Markets. Journal of Marketing Research (JMR) 41, 433–447 (2004)
10. Doyle, P.: Valuing Marketing's Contribution. European Management Journal 18, 233–245 (2000)
11. Gupta, S., Zeithaml, V.: Customer Metrics and Their Impact on Financial Performance. Marketing Science 25, 718–739 (2006)
12. Venkatesan, R., Kumar, V.: A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. Journal of Marketing 68, 106–125 (2004)
13. Bayón, T., Gutsche, J., Bauer, H.H.: Customer Equity Marketing: Touching the Intangible. European Management Journal 20, 213–222 (2002)
14. Eberling, G.: Kundenwertmanagement: Konzept zur wertorientierten Analyse und Gestaltung von Kundenbeziehungen. Dt. Univ.-Verl., Wiesbaden (2002)
15. Berger, S.C., Gensler, S.: Evaluating customer management implementations. In: Proceedings of the IIMA Conference on Research in Marketing 2007, Indian Institute of Management, Ahmedabad (2007)
16. Hennig-Thurau, T.: Relationship Marketing Success Through Investments in Customers. In: Hennig-Thurau, T. (ed.) Relationship marketing: gaining competitive advantage through customer satisfaction and customer retention, pp. 127–146. Springer, Heidelberg (2000)
17. Bell, D., Deighton, J., Reinartz, W.J., Rust, R.T., Swartz, G.: Seven Barriers to Customer Equity Management. Journal of Service Research 5, 77–85 (2002)
18. Berger, P.D., Nasr, N.I.: Customer Lifetime Value: Marketing Models and Applications. Journal of Interactive Marketing 12, 17–30 (1998)
19. Blattberg, R.C., Deighton, J.: Manage Marketing by the Customer Equity Test. Harvard Business Review 74, 136–144 (1996)
20. Rust, R.T., Lemon, K.N., Zeithaml, V.A.: Return on Marketing: Using Customer Equity to Focus Marketing Strategy. Journal of Marketing 68, 109–127 (2004)
21. Wiesel, T.: Kundenkenngrößen und Unternehmenswert. University of Frankfurt, Frankfurt/Main (2006)

22. Ryals, L., Payne, A.: Customer relationship management in financial services: towards information-enabled relationship marketing. Journal of Strategic Marketing 9, 3–27 (2001)
23. Wiesel, T., Skiera, B., Villanueva, J.: Customer Equity – An Integral Part of Financial Reporting. Journal of Marketing 72 (in press, 2008)
24. Kumar, V.: CLV: The Databased Approach. Journal of Relationship Marketing 5, 7–35 (2006)
25. Rogers, E.M.: Diffusion of innovations. Free Press, New York (2003)
26. Riemenschneider, C.K., Harrison, D.A., Mykytyn Jr., P.P.: Understanding it adoption decisions in small business: integrating current theories. Information & Management 40, 269–285 (2003)
27. Harrison, D.A., Mykytyn Jr, P.P., Riemenschneider, C.K.: Executive Decisions About Adoption of Information Technology in Small Business: Theory and Empirical Tests. Information Systems Research 8, 171–195 (1997)
28. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13, 319–340 (1989)
29. Ajzen, I.: Understanding attitudes and predicting social behavior. Prentice-Hall, Englewood Cliffs (1980)
30. Lucas Jr., H.C., Swanson, E.B., Zmud, R.W.: Implementation, Innovation, and Related Themes Over The Years In Information Systems Research. Journal of the Association for Information Systems 8, 206–210 (2007)
31. Plouffe, C.R., Hulland, J.S., Vandenbosch, M.: Research Report: Richness Versus Parsimony in Modeling Technology Adoption Decisions: Understanding Merchant Adoption of a Smart Card-Based Payment System. Information Systems Research 12, 208–223 (2001)
32. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: toward a Unified View. MIS Quarterly 27, 425–478 (2003)
33. Jeyaraj, A., Rottman, J.W., Lacity, M.C.: A review of the predictors, linkages, and biases in IT innovation adoption research. Journal of Information Technology 21, 1–23 (2006)
34. Tornatzky, L.G., Fleischer, M., Chakrabarti, A.K.: The processes of technological innovation, Lexington, Mass. Lexington Books (1990)
35. Zhu, K., Kraemer, K., Sean Xu, K.: Electronic business adoption by European firms: a cross-country assessment of the facilitators and inhibitors. European Journal of Information Systems 12, 251–268 (2003)
36. Kuan, K.K.J., Chau, P.Y.K.: A perception-based model for EDI adoption in small businesses using a technology-organization-environment framework. Information & Management 38, 507–521 (2001)
37. Zhu, K., Dong, S., Xu, S.X., Kraemer, K.L.: Innovation diffusion in global contexts: determinants of post-adoption digital transformation of European companies. European journal of information systems 15, 601–616 (2006)
38. Chau, P.Y.K., Tam, K.Y.: Factors Affecting the Adoption of Open Systems: An Exploratory Study. MIS Quarterly 21, 1–24 (1997)
39. Rust, R.T., Zeithaml, V.A., Lemon, K.N.: Driving customer equity: how customer lifetime value is reshaping corporate strategy. Free Press, New York (2000)
40. Keiningham, T.L., Aksoy, L., Bejou, D.: How Customer Lifetime Value Is Changing How Business Is Managed. Journal of Relationship Marketing 5, 1–6 (2006)
41. Hoekstra, J.C., Leeflang, P.S.H., Wittink, D.R.: The Customer Concept: The Basis for a New Marketing Paradigm. Journal of market-focused management 4, 43–76 (1999)
42. Haenlein, M., Kaplan, A.M., Beeser, A.J.: A Model to Determine Customer Lifetime Value in a Retail Banking Context. European Management Journal 25, 221–234 (2007)

43. Xu, M., Walton, J.: Gaining customer knowledge through analytical CRM. Industrial Management & Data Systems 105, 955–971 (2005)
44. Benbasat, I., Barki, H.: Quo vadis, TAM? Journal of the Association for Information Systems 8, 212–218 (2007)
45. Tuzovic, S., Bruhn, M.: Integrating customer orientation, employee compensation and performance management: a conceptual framework. International Journal of Business Performance Management 7, 255–274 (2005)
46. Gopalakrishnan, S., Damanpour, F.: The Impact of Organizational Context on Innovation Adoption in Commercial Banks. IEEE Transactions on Engineering Management 47, 14–25 (2000)
47. Nord, W.R., Tucker, S.: Implementing routine and radical innovations, Lexington, MA. Lexington Books (1987)
48. Damanpour, F.: Organizational Size and Innovation. Organization Studies 13, 375–402 (1992)
49. Utterback, J.M.: Innovation in Industry and the Diffusion of Technology. Science 183, 620–626 (1974)
50. Swanson, E.B.: Information Systems Innovation Among Organizations. Management Science 40, 1069–1092 (1994)
51. Stoneman, P., Myung Joong, K.: Technology Adoption and Firm Profitability. Economic Journal 106, 952–962 (1996)
52. Wahrenburg, M., Hackethal, A., Friedrich, L., Gellrich, T.: Strategic decisions regarding the vertical integration of human resource organizations: evidence for an integrated HR model for the financial services and non-financial services industry in Germany, Austria and Switzerland. International Journal of Human Resource Management 17, 1726–1771 (2006)
53. Teo, H.H., Wei, K.K., Benbasat, I.: Predicting intention to adopt interorganizational linkages: an institutional perspective. MIS Quarterly 27, 19–49 (2003)
54. Gatignon, H., Robertson, T.S.: Technology Diffusion: An Empirical Test of Competitive Effects. Journal of Marketing 53, 35–49 (1989)
55. To, L.M.: Determinants affecting organizations' intent to adopt e-tailing: a study based on innovation theories. SIGCSE Bull. 36, 243–243 (2004)
56. Beccalli, E., Casu, B., Girardone, C.: Efficiency and Stock Performance in European Banking. Journal of Business Finance & Accounting 33, 245–262 (2006)
57. Hinshaw, M.: A survey of key success factors in financial services marketing and brand management. Journal of Financial Services Marketing 10, 37–48 (2005)
58. Ramaswami, S.N., Bhargava, M., Srivastava, R.: Market-based Assets and Capabilities, Business Processes, and Financial Performance. Working Paper. Marketing Science Institute (2004)
59. Tschoegl, A.E.: Size, Growth, and Transnationality among the World's Largest Banks. Journal of Business 56, 187–201 (1983)
60. Brislin, R.W.: Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology 1, 185–216 (1970)
61. Dibbern, J.: The sourcing of application software services: empirical evidence of cultural, industry and functional differences. Physica, Heidelberg (2004)
62. Chin, W.W.: The Partial Least Square Approach to Structural Equation Modeling. In: Marcoulides, G.A. (ed.) Modern Methods for Business Research, pp. 295–336. Lawrence Erlbaum Associates, Mahwah (1998)
63. Wold, H.: Systems under Indirect Observation Using PLS. Praeger, New York (1982)
64. Ringle, C.M., Wende, S., Will, A.: SmartPLS. University of Hamburg (2005)

65. Korner, V., Zimmermann, H.-D.: Management of Customer Relationship in Business Media. Electronic Markets 10, 162–168 (2000)
66. Jarvis, C.B., Mackenzie, S.B., Podsakoff, P.M.: A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. Journal of Consumer Research 30, 199–218 (2003)
67. Mian, A.: Distance Constraints: The Limits of Foreign Lending in Poor Economies. Journal of Finance 61, 1465–1505 (2006)
68. Ryals, L.: Making Customer Relationship Management Work: The Measurement and Profitable Management of Customer Relationships. Journal of Marketing 69, 252–261 (2005)
69. Cao, Y., Gruca, T.S.: Reducing Adverse Selection Through Customer Relationship Management. Journal of Marketing 69, 219–229 (2005)
70. O'Sullivan, D., Abela, A.V.: Marketing Performance Measurement Ability and Firm Performance. Journal of Marketing 71, 79–93 (2007)
71. Podsakoff, P.M., MacKenzie, S.B., Jeong-Yeon, L., Podsakoff, N.P.: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. Journal of Applied Psychology 88, 879–903 (2003)
72. Bagozzi, R.P., Youjae, Y.: On the Evaluation of Structural Equation Models. Journal of the Academy of Marketing Science 16, 74–94 (1988)
73. Bollen, K., Lennox, R.: Conventional wisdom on measurement: A structural equation perspective. Psychological Bulletin 110, 305–314 (1991)
74. Homburg, C., Baumgartner, H.: Beurteilung von Kausalmodellen: Bestandsaufnahme und Anwendungsempfehlungen. In: Hildebrandt, L., Homburg, C. (eds.) Die Kausalanalyse: ein Instrument der empirischen betriebswirtschaftlichen Forschung, pp. 343–369. Schäffer-Poeschel, Stuttgart (1998)
75. Tenenhaus, M., Vinzi, V.E., Chatelin, Y.-M., Lauro, C.: PLS path modeling. Computational statistics & data analysis 48, 159–206 (2005)
76. Götz, O., Liehr-Gobbers, K.: Analyse von Strukturgleichungsmodellen mit Hilfe der Partial-Least-Squares(PLS)-Methode. Die Betriebswirtschaft 64, 714–738 (2004)
77. Sellin, N., Keeves, J.P.: Path analysis with latent variables. Elsevier Publishers, London (1994)
78. Phillips, L.A., Calantone, R.: International technology adoption: Behavior structure, demand certainty and culture. Journal of Business & Industrial Marketing 9, 16–28 (1994)

Information Risk in Financial Institutions: Field Study and Research Roadmap*

Sara Sinclair[1], Sean W. Smith[1], Stephanie Trudeau[1],
M. Eric Johnson[2], and Anthony Portera[2]

[1] Department of Computer Science, Dartmouth College
[2] Center for Digital Strategies, Tuck School of Business at Dartmouth
{sara.sinclair,sean.w.smith,stephanie.trudeau,
m.eric.johnson}@dartmouth.edu

Abstract. Large financial firms with thousands of employees face many challenges ensuring workers have access to the right information, yet controlling access to unneeded data. We examine the problems of role lifecycle management and entitlement review processes in the context of large financial institutions. We describe observations from field study research in both retail and investment banks. We examine technologies to enable role and entitlement management and present a roadmap for future research.

Keywords: Access control, information risk, entitlement, provisioning, matrixed organizations, organizational complexity.

1 Introduction

Mergers in many industries have created very large global enterprises with thousands of employees, contractors, and partners scattered around the world. These massive collections of people lead to greater anonymity of the employees as they fade into the masses and can conceal actions that challenge modern security and controls [1]. Financial institutions provide a vivid example of this complexity. There are two major kinds of banks in the United States: retail banks and investment banks. Retail banks are the institutions that most of us are used to seeing on street corners, and advertise services like checking accounts and mortgages. Washington Mutual, Citizens Bank, and Sun Trust are examples of large retail banks. In contrast, most Americans will never directly interact with an investment bank. Investment banks serve companies and governments, raising capital by helping them issue stocks, bonds, and other securities. Investment banks often also engage in security trading services, hedge fund support, asset management for high net worth individuals, and analyst coverage of

* This research was supported through the Institute for Security Technology Studies at Dartmouth College, under awards 60NANB6D6130 from the U.S. Department of Commerce and U.S. Department of Homeland Security under Grant Award Number 2006-CS-001-000001. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Institute of Standards and Technology (NIST), the U.S. Department of Commerce, or U.S. Department of Homeland Security.

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 165–180, 2008.

securities traded in the public markets. Consolidation in the banking industry has created large investment bank firms including Goldman, Sachs & Co. with $37.7B in revenues and 26K employees and Lehman Brothers with $17.6B in revenues and 26K employees. There are also banks that have both retail and investment banking operations, such as Bank of America with revenues of $85B and 176K employees, and CitiGroup with $120B in revenue and 307K employees.

As these organizations grow and consolodate, they face new technological challenges. In this paper, we explore a set of such issues that several financial firms have encountered in recent years, as researched during a field study encompassing a number of retail and investment banks. This field study, while encompassing general issues of technology as it relates to business requirements and constraints, focused particularly on the sharing and control of digital information.

Table 1. Summary of visits and individuals interviewed during field study in retail and investment banks

Institutions	3
Individuals Interviewed	31
On-site (many multiples)	24
Remote location	7
Team Days On-Site	16

The primary field research of the study was conducted at three institutions: two investment banks and one retail bank. (Table 1 provides details on the number of institutions and employees interviewed.) All interviews were conducted in person, either on-site or at remote locations. In nearly every case, there were at least two interviewers; in some case, an additional 1-2 researchers joined the interviews via teleconference. Interviews lasted at least one hour, and many individuals were interviewed multiple times over a period of a year. In addition to one-day visits and interviews held off-site, a team of two researchers spent ten working days on-site in one institution, attending security group meetings and interacting with key individuals multiple times. Interviews were captured by both researchers, with one entering responses on a computer and the other taking hand-written notes. The resulting records consisted of over a hundred pages of notes, as well as sanitized documents provided by the financial partners.

As we will outline in the rest of the paper, we found that the large financial institutions who participated in our study must balance on a tightrope when it comes to data access: information security compromise can lead to significant financial loss [2], but so can overly stringent security measures that prevent employees from getting their jobs done in a timely fashion. This paper focuses primarily on internal information access controls and the risks of overaccess. We present our field study results along with a roadmap for future research in the following order. We first consider high-level organizational and business factors that help us understand the challenges the financial sector faces in this domain; in particular, we examine organizational complexity and security technology strategies of our partner firms.

2 Organizational Complexity Feeds Security Complexity

Banks are well-known for the intense expectations they place on their employees to meet the demands of clients. This intensity often results in significant organizational change and employee turnover. As professionals move within the organization, their information needs change, requiring rapid, precise computer systems that allow a great deal of flexibility, but without comprising security. It is no surprise that fraud prevention and confidentiality are major concerns of banks inside and outside of the firm. Confidentiality concerns focus on keeping client identities, strategies, intentions, and just about everything else held in confidence. Fraud spans issues from theft and misapplication to inappropriate use of confidential information.

Historically, opportunities for fraud may have been a greater concern in small firms where a handful of individuals administered the back office, each wearing many functional hats, and modern separation of duties and limited-access protocols were not feasible. As companies get larger, the intricacies of controls and needed permissions become more complex as both the number of systems and the granularity of access rights within enterprise-spanning infrastructures grow. In the large systems used by financial services, firms struggle to enable the right levels of access while restricting privileged information. Many users may not be aware of their access capabilities or the possible conflicts that their access could create.

A *toxic combination* is a conflict of system access permissions that allows a user to break the law, violate rules of ethics, damage customers' trust, or even create the appearance of impropriety. There are many ways for toxic combinations to occur. Sometimes it is a mistake of not terminating access following a promotion or transfer; other times it is a fault of entitlement[1] design. An example of toxic combinations occurring from promotion could be as seemingly innocuous as an accounts payable clerk retaining the access to write checks once he has been promoted so he can assist with that task at busy times. If his new job allows him to go back to edit and even delete check writing records, he has the opportunity to steal money while circumventing traditional checks and balances. A design flaw example would be a trader in a commercial bank having access to see holdings of the accounts for clients she manages, as well as those of other trader's clients. The trader's access could be used to counter the aggressive positions of her non-direct client to the enrichment of herself and others, which is not only unethical, but also highly illegal.

Over-entitlement is also risky. It may not seem problematic for employees to have access to systems they never use or are unaware of. However, such access introduces risk. The root of the problem is that unnecessary or uncontrolled access can lead to unintended data editing, accidental disclosure, or internal misuse. That is why Sarbanes-Oxley (or SOX)[2] auditors will flag unnecessary access as a weakness. The large investment banks have thousands of information systems and millions of different entailments resulting in extreme complexity for new or transfer employees to get the permissions they need. A common solution to this problem is that a new employee or

[1] An *entitlement* is a resource that a person can be authorized to access in a certain way; for example, "opening case files" might be an entitlement for application X (we could also call it a "privilege" or "permission").

[2] Also known as the "Public Company Accounting Reform and Investor Protection Act of 2002.

a transferred employee will generate a request to simply copy the entitlements of the most well-endowed person in the department to the new individual, frequently leading over-entitlement for the new individual's needs.

From an employee's point of view, increased portability and accessibility of information facilitates productivity. Employee turnover has always been a concern for information leakage, but, as information becomes more fluid and more easily accessed over public networks, control of voicemail, remote email, PDA email, and home use of corporate files becomes more important. For example, a few years ago, one large bank's IT security group came to the realization that much of their important data lived in Excel spreadsheets, rather than the large, secure proprietary systems one might have expected (and hoped). This realization led them to prioritize the purchase and implementation of additional software and controls to further limit the movement and access of these files.

In addition to the large, multi-function corporate systems, companies are also shifting management strategies and human resources faster than ever before. In the investment banking space it is quite common for people to move between internal organizations and be transferred across information boundaries. The frequent shifting of staff may result in information users collecting system entitlements over time if the system access is not actively managed, resulting in a toxic combination of privileges.

Innovations in organizational structure also make security more difficult [3] as far as approving, monitoring and terminating information access are concerned. Most people are familiar with static hierarchical organizations in which everyone has a boss, who reports to a higher-level boss, and organizations have defined lines of responsibility as depicted in Figure 1. Team-based and matrix-structured organizations [4] are becoming more and more prevalent in professional society and especially in the financial services industry. In a matrix organization [5] employees may report to many "bosses," each for a set of different projects or activities, as also depicted in Figure 1. Some entitlement technologists have called these intertwining multiple hierarchies *polyarchies*. The polyarchy is simply a group of non-linear reporting relationships that come together to make an organizational matrix.

As more organizations take on a matrix structure, it becomes less evident who reports to whom and who is responsible for permitting and terminating data access. Employees may no longer have direct managers, as roles such as functional manager, group manager, engagement manager, review manager, and co-manager have become prevalent in financial institutions, professional service firms, and corporations. Additionally, the need for easy transfer from one role to another within a company is part of the organizational design that accentuates the complexity of securing information. Questions like, "Who owns this data or file?" and, "Who should approve access?" become common as data stores accumulate. The natural trap is for IT to give people access to whatever they ask for without an appropriate approval process; this often happens in cases in which there is no credible manager and IT is not able to track when the entitlement is no longer needed, or identify when it conflicts with new responsibilities. We have heard IT professionals jest that one can track a person's career path by examining their system entitlements, which seem to rarely be adjusted down in advancement or transfer. We have been cited examples in which 50-90% of the individuals with access to particular data store also have legacy access to information that they no longer need.

One critical goal of information security in the financial industry is to get the right information to the right people with maximum efficiency. Often, this involves the laborious task of individually granting access rights and entitlements of specific program applications to each user. A growing desire from security professionals within the industry is to create automatic rules and roles that can identify user attributes and access needs so as to automatically perform entitlement and potentially anticipate future permission needs. This goal is not dissimilar to that of sophisticated on-line retailers and content providers. The recent buzz among the management for websites like Yahoo.com and Facebook.com is that they aspire to create a "segment of one." Essentially, they would like to perfectly tailor the content a user receives to precisely meet the needs of that user based on historical and current activities. Given rapidly evolving content availability and the vast amount of traceable data a user can generate, perfect individual tailoring is nearly impossible without dedicating vast amounts of human resources to constantly code new rules or artificial intelligence that can learn and predict trends and personalities. While the "segment of one" may be far off, the institution of automated rules to guide information access may not be far from reality for investment banks. These corporations' desire for automatic permission tailoring is similar to customized content delivery, but the risks of getting it wrong are higher both in productivity downtime and unintended access. We will further discuss this issue after we examine some key security strategies employed by several investment banks.

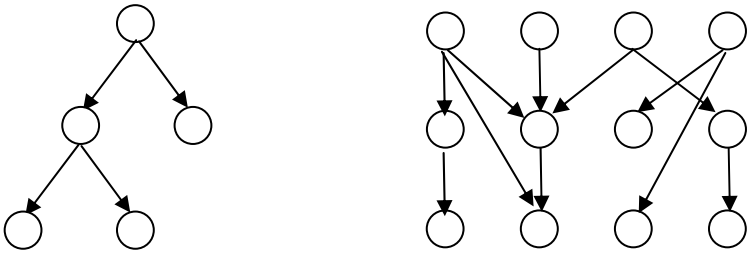


Fig. 2. Privileging in traditional hierarchical corporate structures (left) vs. in dynamically, "matrixed" organizations (right). An arrow represents a supervising relationship (directed graph). We note that on the left, each person has exactly one direct supervisor, whereas on the right, each may have two or more.

3 Security Strategies

As part of our research, we interviewed IT professionals in a collection of large Wall Street financial institutions. We found one firm to be highly progressive, whose security professionals see the ambiguity of their matrix organizational design as an opportunity to "get into the business" and build collaborative bridges between IT and revenue-generating operations, both adding value to solutions to improve operations and generating a more secure environment through understanding the issues surrounding the business. This practice of getting "into the business" requires a highly competent IT staff, a self-starter culture that fosters collaboration and self-sacrifice for the

good of the organization, and the tolerance for the additional investment in relationships. While expensive, it seems to result in bleeding edge innovation and preventative results that seem to pay big dividends in the complex, fast paced world of financial markets.

In this particular financial organization, security strategy is tied to its culture. This culture is strongly oriented toward self-starters who work in teams with the goal of enriching the firm immediately through profitable activities, and over the long term by building the firm's reputation. This view that reputation is paramount in importance to the long-term fortune of the firm and individuals motivates the firm leaders to do whatever is conceivably possible to protect the IT resources of the firm within the bounds of productivity requirements. The urgency to protect the firm results in a few key strategic positions.

Willingness to Buy or Build New Applications: It sounds simple to say, "If you cannot buy something, you can just build your own." However, this is not a common approach outside the upper echelon of the financial services industry, and even there, building custom solutions can be seen as too laborious or expensive to make it worth the firm's while. This particular progressive firm is truly agnostic on the buy-versus-build decision in its opinions on headache and firm resources. It feels that the security of its specialized systems and data is too important to risk not having the right system attributes. To this end, it has an army of employee programmers and contractors at the ready to build what it needs for both security and business applications. One staffer notes that they are "totally fearless" when it comes to building a solution if they cannot find a commercial one that meets their needs. They are also not afraid to ask a vendor to sell them part of a program (and then build it into existing or new applications currently used by the firm), or to request that a vendor heavily modify an application. One employee stated that this determination to find the right solution, even if it doesn't currently exist on the marketplace, "teaches them a lot" about what they need and what to look for in future solutions.

At this firm, an initiative to build a new program for the firm is not taken lightly. Employees are highly intolerant of half-baked applications (whether created internally or by vendors) that unduly slow or block the flow of business for a moment. The decision to build an application is a commitment to see it through to its operational completeness.

To contrast this firm's perspective, another firm we spoke with had the capability to and does build its own systems for productivity and security functions, but we detected its bias to purchase software or work with an outside vendor who will develop a commercial application that includes their needs. Their perspective is based on the benefits of financial efficiency of using outside vendors and appreciates the support that comes with purchasing software.

Financial Freedom to Explore Options: The delegation culture that results from management's expectation that all employees of this particular firm are self-starters can allow for great amounts of freedom to explore new concepts in information security without burdensome oversight. This freedom is provided via financial funding and an absence of day-to-day accountability for personal actions within the project management level of the security group. The security group is *expected* to keep the firm safe. As long as security breaches do not occur or, more importantly, that nothing

shows up in the *Wall Street Journal* indicating a confidentiality or fraud event, the group does not fall under scrutiny and feels free to work on whatever initiative seems important to them, even if the results are not immediately visible.

The Chief Information Security Officer expressed his constraints well: "Funding and resources have never been much of an issue. It is a question of control and flexibility. The hardest part is to achieve balance from the client side." The control of the information resource and flexibility for the business to get the job done are of paramount importance. Each one must be satisfied while not overpowering the other.

This freedom allowed this firm to develop innovations in information rights auditing for key applications years before Sarbanes-Oxley came into existence and regulators began penalizing other investment firms for their lack of foresight. This firm's institution of the SOX-type tools was not motivated out of the expectation that Congress would institute new rules, but that the firm and the industry needed this type of controls.

Open Vendor Relations: Despite its willingness to build whatever it needs, this firm is a big buyer of vendor-created applications. However, its size and clout within the industry leads to relationships that are generally more than the standard arms-length buyer-seller relationship. These relationship differences are derived from this firm's objective in getting precisely the right program attributes and willingness to collaborate. It is not uncommon for this firm to ask technology vendors to customize their products and make them go through months of, and sometimes years of, proof-of-concept exercises.

This firm also sees itself as an innovator within the industry. Management often shares the program attributes that it has created generally within its internal systems with vendors who approach them with "competing" products. They see this as both an opportunity to help the vendor create a product that may be useful to them in the future and an opportunity to give vendors insights that may help others in the industry.

Innovative Security Group Structure: Many investment banks separate the responsibility of monitoring security needs and generating security policies to a business risk organization staffed with business analysts who frequently work with a separate IT security organization that is responsible for the development and management of security applications. At this firm, however, there is no separation of the business risk organization and IT risk organization. The security innovators are the security creators. Many of the professionals who join the team come in with backgrounds in various technical and non-technical fields, but very few had previous experience as information security specialists. They learn security on the job and are expected to be conversant in both the IT needs and the business activities of the firm. When they discover a need to mitigate risk they do not give the task to another organization they find their own solution or directly manage its development with the programming resources within the firm. They have created an environment where policy creation is not the end of one person's job that is handed off to another, but the beginning of a development cycle where the policy developer is also the solution creator who must prove hypotheses and get firm buy-in for the changes he spurs in the organization.

As a security application or tool is created, the team makes sure it can be "commoditized." They use the term commodity in this sense to indicate that they do not want the usage of the tool to be so complex that it cannot be handed to someone else.

In fact, the security team does not want to build a security empire. They want to develop tools and spin them out to fall under the responsibility of other existing or new organizations that report outside the information security group; one employee stated that they see themselves as a "factory solution" that creates and distributes technology and processes to the rest of the firm. The security team retains some control or input into the strategy, architecture, and budget of the spin-out, but does not want to be part of the day-to-day administration of security, just the innovation.

This structure may not be fully feasible to other companies, but the practices of eliminating boundaries by combining functions and raising the bar on security professionals to see themselves as part of the business could be helpful to many organizations.

4 Managing Complexity in Security

Thus far we have explored business and organizational characteristics that affect information security in financial institutions; we have seen that organizational complexity introduces tremendous challenges, and that business strategies on technology can greatly impact a firm's effectiveness in data management. While conducting interviews on these aspects of the problem, we also used the field study to focus on a number of key technological issues; of particular interest was one firm's experiences conducting a SOX-mandated review of entitlements, as well as the challenges one firm faces in its efforts to deploy role-based access control (RBAC). Examining these problems and the technology firms used (or could not find or build to use) helps us better understand the need for new tools and research in this domain.

4.1 Entitlement Review

During our research visit, a team within the technology group was in the process of finishing a comprehensive entitlement review of all SOX-relevant applications (on the order of 200 different programs). The purpose of the review was to verify that each employee of this firm who has access to these applications has an appropriate set of entitlements, neither so few that they can't complete their jobs, nor so many that they pose an information risk. In many lights the review was a success: most employees' active entitlement set was reduced 30-50%. Yet, although the participation was high enough to satisfy audit requirements, the review also saw a low level of buy-in from certain business units (even when we account for the business units that were exempt because they already had an entitlement review process in place); given these two statistics, it is difficult for this firm to estimate the impact that the review had on the risk posed to the company. As one staffer told us, they cannot know if reducing Alice's entitlements twice as much as Bob's indicates that Alice's risk was reduced twice as much as her counterpart, nor can we accurately estimate how business unit A's 100% participation compares to business unit B's 50%. The technology risk team deals with this mix of hard statistics and nebulous interpretation all the time, which makes measuring the success of their efforts (and thus the value to the firm) very difficult.

Employees were asked during the entitlement review process to undergo a self-review, during which they were provided with a list of their active entitlements and asked to release those they did not need. This phase of the review was more productive

than the staff expected; employees collectively reduced their entitlement to applications (not just individual entitlements within an application) by 15%. Technologists said that employees "just didn't want to worry about" having access to applications they didn't need. This could be in part motivated by the fact that employees' managers were to perform a review of each employee directly following the self-review; an employee knew that they could only benefit in their manager's eyes by giving up entitlements. Given other observations, we feel that this voluntary de-privileging was perhaps more subtly motivated by the larger culture toward doing what's best for the firm. If employees genuinely have the best interests of the company at heart, and understand the risk over-entitlement poses, they would choose to give up as much as they could.

Unfortunately, this enthusiasm for de-provisioning during the review resulted in under-provisioning for many employees. (Unfortunately, we do not have data on the amount of resultant under-entitlement, but anecdotal evidence suggests that this surprising phenomenon was widespread.) In some cases, this was because employees could not correlate the entitlement description in the review to an entitlement they actually use. This was often because the human-readable entitlement description required of each application owner was not sufficiently clear, or was missing, but sometimes attributable to the users simply not knowing that clicking on the green button on their computer was really launching application ABC. In other cases, employees became under-provisioned because there were hidden entitlement dependencies, i.e., it was not clear that access to application ABC was essential to important work in application XYZ.

Why was the entitlement review process difficult? Shouldn't it be straightforward to list the number of applications, list the different privileges within them, and evaluate the mapping of those privileges to users? Shouldn't it be clear which entitlements users need by reading their job description and project assignments? This is, in fact, an area in which the size and complexity of the company inherently complicate the task at hand. The vast majority of applications are used and maintained by specific business units; there are thousands of applications spread throughout the firm, and new ones are constantly being deployed. Each application has its own notion of entitlements, its own entitlement descriptions, and required its own mental model of how to map human privileges to data access (and there are tens of thousands of humans who might be users for each system). This results in each user having hundreds of individual privileges, and supervisors being forced to review potentially thousands of individual entitlements (which, again, are not necessarily represented in a very human-understandable way).

The biggest challenge isn't the massive number of entitlements or users, however, but the highly dynamic nature of employees and organizational structure within the firm. The matrixed environment is hard to evaluate, and even harder when an individual's manager and entitlement needs are in constant flux (even though this dynamicism is a source of strength in the business world). The only way staffers were able to complete the entitlement review is because they took a "snapshot" of the entitlement systems, and only re-evaluated this snapshot a few times during the review process. During a few months of the review, one business group of 3,000 people witnessed 1,000 changes to organizational structure; in the space of a few weeks, 158 users in another group had changed job positions. If the process took into account all of these changes as they were happening, employees conducting the review would be so busy

updating the picture of organizational structure that they would have no time to actually review entitlements. Conducting an entitlement review annually would be much easier if there were a persistent and up-to-date picture of the organizational structure and entitlements throughout the year.

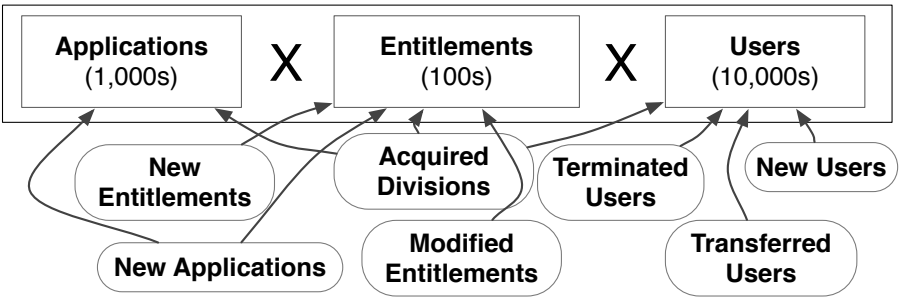


Fig. 3. Complexity and dynamics in entitlement systems. The number of applications, entitlements, and users make entitlement a large-scale problem, and the number of daily modifications to each of these sets makes it a fast-moving target.

4.2 Using Roles for Structure and Entitlements

Many corporations are looking for the best way to manage entitlements in this kind of large, complex, and dynamic environment. Employees of some companies we have interviewed doubt the feasibility of any kind of centralized management of entitlements; they feel that the space and diversity of data is simply too massive. However, we found in the firm we observed not only a willingness to consider such technologies, but deployment of one already in progress. The team of technologists behind this implementation is driven by a vision of a *role-based* system that maps entitlements to classes of users via *rules*.

Organizations such as NIST have been championing Role-Based Access Control (RBAC) schemes [6, 7] since the mid-nineties, and there exist a number of frameworks to implement access control using this model [8]. RBAC traditionally affiliates business-level functions, or roles, with sets of permissions on a given system; this grouping of entitlements facilitates provisioning (e.g., Alice is a desk manager, so we give her access to the same kinds of resources as other desk managers), as well as entitlement review (e.g., all desk managers should have the set "S" of entitlements) and modification to existing entitlement groups (e.g., we want to give all desk managers access to system XYZ when it comes online). In theory, RBAC schemes allow us to segregate the massive numbers of employees and entitlements into distinct groups that are easier to manage. However, the size and complexity of large banks make role-based systems challenging. At one very large retail bank that we interviewed, the CISO had recently completed an RBAC project creating 11,000 roles across the firm to control access to its nearly 22,000 applications. Developing the roles took a team two years and the ongoing review process was expected to be significant.

To study the feasibility of deploying a role-based system, one team at an investment firm we observed set out to estimate the number of roles that would naturally

fall out of existing applications; they found that the number of simple roles within the company greatly outnumbered employees. This is a reasonable result given the number of applications and entitlements; while some users in the firm had entitlements in one or two applications and could generally be covered by one role each, others had hundreds of entitlements in dozens of applications, and the privilege set overlapped with colleagues in varying ways. (The task of choosing the minimum number of sufficient roles in this situation is reminiscent to the set cover problem, which is known in computer science to be computationally difficult.) Of course, for roles to bring value to the company, the system must group users into more manageable units, not just add a layer of complexity and abstraction. A role-based system can clearly mitigate some forms of complexity, but offers also its own set of management challenges in defining appropriate roles. Furthermore, we can only assume that the dynamicism inherent to organizational structure within an investment firm will lead to highly dynamic roles as well; as jobs and applications evolve, so will the logical groupings of employees doing those jobs and using those applications. (Generally, it seems that retail banks are more naturally suited to simple role schemes, given the more static nature of the corporate structure.)

In response to these observations, the team who is integrating role-based systems into the corporate infrastructure is focusing on role management, and on the *life-cycle events* that define a role and determine which users are assigned to it. For example, a new employee joining the firm is an event that triggers assignment of appropriate role(s) to her: when that employee changes divisions, she may lose some roles and gain others; when she leaves the firm, she should be removed from membership of all roles. The series of events starting with an employee's hiring and ending with proper de-provisioning upon her leaving reflects the grand vision of role management within this firm; through constant event-based roles updating, it should be possible to get a clear picture of the entitlements currently assigned, as well as easily change entitlements for classes of users when necessary.

Thus far, staffers have identified some key characteristics of a system that would meet this grand vision. (They have also contracted to use a commercial product to help with role management, which we discuss later in this document.) Most importantly, they believe that role management will consist less of technological solutions, and more of business processes. This is in line with prior work that suggests many security problems are due to a lack of understanding of human systems [9, 10], or to technology that does not appropriately model the needs of those human systems [11, 12]. Parallel mechanisms already exist in entitlement and provisioning systems; there are well-defined ways in which business units communicate with human resources managers, or in which application users can request new access privileges from a particular application owner. Integrating roles and role management into their infrastructure will require discovering the correct business processes to tap into, and inventing new business processes where appropriate ones do not exist. Security technologists at the firm will be heavily involved in the initial deployment of roles, but once they are established, business-focused employees should run these systems; to ensure that they maintain up-to-date in the dynamic environment, they must be managed by those already effecting dynamicism within each business unit.

This process-based vision of role management is different than the traditional research-based idea of roles. Indeed, this firm's definition of a "role" as a persistent group of entitlements with a name and correspondence to business role is different from most instantiations of the RBAC model; in the latter, an individual user often has many roles, and only takes on one at a time. The traditional RBAC model strives to make sure that a user only has access to the set of entitlements with the least privilege to accomplish a given task; in this way, if a user's access is compromised, the number of resources affected is minimized. The firm's view, in which users always employ the roles assigned to them at all times, is much more similar to "group"-based access control schemes. However, as one staffer noted, the types of groups they need are much more sophisticated than current implementations (such as Active Directory). This is an area of authentication theory we hope to explore more with this firm, and share with the research community at large.

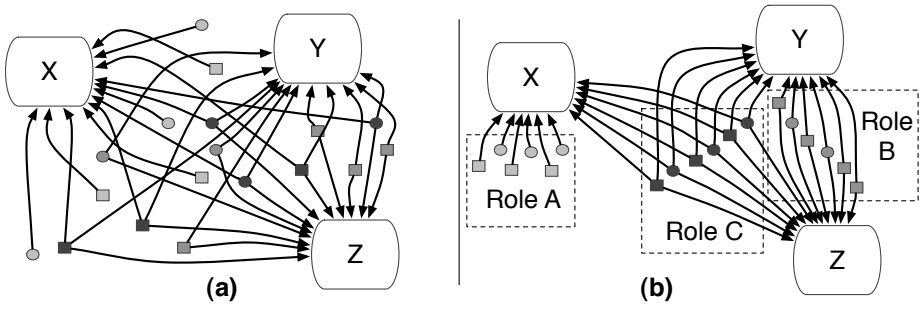


Fig. 4. A set of users may have entitlements for resources X, Y, and/or Z (b); here small shapes represent users, and shade is determined by the set of entitlements the user has. A "blind" clustering algorithm (a) groups users into roles based on the sets of entitlements they have. Here, users with entitlement X (light grey) are grouped into Role A; those with Y and Z (dark grey) are in Role B, and those with all three entitlements (black) are assigned to Role C.

Although this firm recognizes the need for effective role management mechanisms, they also understand that they must establish roles before they can manage them. However, because they hope to use roles to get a better grasp on the different types of resources in the firm, we find a bit of a chicken-and-the-egg problem: how do we define roles as a function of resource entitlements, when we're planning on using roles to help understand what our resource entitlements are? Given the scale of the problem, the CISO (Chief Information Security Officer) scoffed at the idea of manually grouping people into roles.

The commercial product that the firm has licensed features a "role discovery" function, which allows system owners to mine entitlement information for pre-existing clusters of privileges that might be incorporated into roles. This approach is very attractive, especially in a large matrixed organization; it promises to automate what would be a very tedious manual process. However, we joined our collaborators at the firm in questioning the scope and the value of role discovery, and uncovered a line of inquiry on the nature of roles and their evolution. Because role discovery (as implemented in this case)

operates by blindly clustering groups of people (Figure 3) with similar entitlements, how can we be sure that the resulting roles can map meaningfully to existing business roles (Figure 4)? If there is no mapping to business roles, what happens when we decide to add or remove entitlements within specific roles? A role-based system's ability to apply entitlement changes to large groups of people is an attractive feature; if our clusters of users are poorly defined, we might find the need to split, merge, or otherwise redefine role membership from the very beginning.

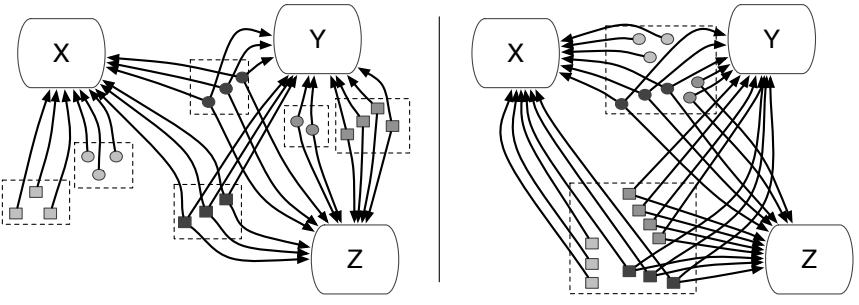


Fig. 5. Alternate clustering algorithms from that presented in Figure 3. Depending on the nature of additional information, such as job status (temporary or permanent, represented here by shape), better definitions of roles may be dependent on more than just entitlement data.

Of course, as the firm evolves, we expect there to be some redefinition of roles. For example, one business unit may divide its accounts into two sets, and determine that there must be a Chinese wall between them. Before the division, perhaps there was a Role A that encompassed the entire unit. After the division, should the group be split into multiple roles: X for those who have worked with one set of accounts, Y for those who have seen the other set, and Z for those who have not yet stepped off the Chinese wall? Or should the role acquire an attribute that allows the system to identify subgroups, i.e., attributes A_X and A_Y, which can overlap with other characteristics in the same role? In subtler cases, might we see gradual "drift" in roles, in which a group or subgroup diverges over time from the original role definition?

These possibilities indicate that role discovery might be a useful tool in first implementing a role-based system, but that there may be elements of the nature of roles that prevent it from being useful in already running systems. (Then again, if the role discovery algorithm were modified to not just blindly cluster, but also take in other information that can influence the grouping of people into roles, it could prove to be more useful in the long term.) These questions also reflect a certain level of ignorance on our part about the nature of roles; currently, we cannot predict how they will evolve once implemented. In a larger context, this problem represents a larger inability to predict the effects technical changes in a system that involves both humans and computers will have; we can simulate changes in networks or data centers, but human systems add new and puzzling factors. We know that the firm we observed is considering these kinds of questions in its plan for role deployment, and hope that through deployment they will be able find some answers questions; we hope to extend our collaboration to include research in this area.

4.3 Role-Based System Technology and Deployment

Different parties in research and industry have developed different notions of "roles," depending on their needs. We have discussed the importance of role life-cycle management in deployment of a role-based system, but we must also consider the technical elements that will make this deployment possible. Perhaps even more relevant to our goals of understanding the security technology deployment in the financial industry, we examined the process by which the technology was chosen, and how it is being deployed.

One product we studied was SmartRoles by Bridgestream. The software is "an enterprise-class application to capture, model, and update relationships between people, processes, projects, documents, locations, and business resources" [13]. The system seemed most focused on understanding "roles and organizational hierarchies." As a repository for relationship information and workflow interface for updating those relationships, one might expect a human resources team to be deploying it instead of a team of risk focused technologists. However, the application's sophisticated use of rules, which translate business policies into automated decision-making on entitlements, and the fact that this system will tie into existing technical entitlement solutions, clearly establish the need for technology professionals to evaluate and motivate its use.

The group of technologists who are behind the SmartRoles deployment first interacted with Bridgestream in the product's early stages. This recognition of the scope and importance of the problem in question has also been mirrored by professionals at other firms in the industry, although at that time there was no product available to even come close to solving it [14]. Once establishing contact, the firm worked with Bridgestream over the course of more than a year to figure out what the product would need to include before being deployed in an operating environment. This delay, and corresponding willingness of both to negotiate features and interfaces over a surprisingly long period of time, reflects both Bridgestream's eagerness to have an important financial customer, and the firm's particularly open yet stringent philosophy regarding vendors, which we discussed previously in this document.

Once IT professionals decided to push a deployment of SmartRoles to the corporation, they had to establish sufficient buy-in among various business units. This emphasizes the cultural expectation that business groups are not forced to conform to IT initiatives; if the IT team felt that the application was good for the firm, they would have to make the case and convince non-technical employees, too. Because the IT team must get significant input and acceptance from the business group (who will essentially be the users of the application), the chance of deploying an unusable or imperfect product is even further reduced than if the IT team were evaluating it alone. This stands in stark contrast to the policies and customs of other companies and industries.

The technical team solicited buy-in from business groups using a well-rehearsed "road show," in which they collaborated with Bridgestream to demonstrate the product and the ways in which it would be useful to various groups within the company. Generally, it seems that IT personnel at the firm we were observing are unusually in touch with the needs and attitudes of their business counterparts, which makes the process of selling a product much easier. However, they felt that there was a distinct "terminology barrier" in discussing the roles rules at the heart of the product. In many

cases, business and IT staffers have developed a common language to discuss existing notions and technology, but our collaborators found in this instance that the concepts necessary were both radically new, yet similar enough to existing business ideas to be confusing. For example, when discussing a role, does one mean a job title, the subtler part that an employee plays within the business group, or something else entirely? More generally, there seemed to be a difficulty in understanding the fundamental need for such a product, the ways in which this firm's matrixed structure and demand an intricate and flexible system. We feel that this reflects a prescience of their technical staff: they are ahead of the curve in understanding the needs of their business counterparts, as well as in solving problems that other firms in their industry are still grasping to formulate.

The firm is now in the process of rolling out SmartRoles in a select number of systems within the firm. This initial deployment is critical to the success of the application, as it will determine whether or not other groups decide to adopt it. They have initially targeted systems that already have a strong notion and need for with entitlement, that will not require a lot of maintenance by the business teams but that will demonstrate the benefits and efficiency that can be gained.

5 Conclusions and Future Work

The results of our field research provide insight into the challenges facing data-driven financial firms today. The increasingly distributed availability of data (and the potential for improved efficiency and profit this availability presents) pushes developers and managers to provide better ways to share information faster. At the same time, increasingly strict regulations require new standards of auditing and risk management; firms thus face both the need for increased functionality surrounding their data and more rigorous control over who can access it and when. Our field study also highlighted how environments of rapid organizational change (such as highly dynamic matrixed organizations) reqire new degrees of scalability, manageability, and usability from their data access control solutions.

We hope to use our future research in this space to help financial firms meet these challenges posed by the access control problem. To this end, we are developing models of both the organizational and system application structure to allow us to simulate the effectiveness of potential technical and access policy changes. For example, a model of an organization that allows the simulation of employee hiring, termination, promotion, and supervisory relationship changes could enable us to predict how auto-provisioning users with a certain role at a certain lifecycle event would affect the overall system. (At this time we are considering a highly parallelized discrete event simulator, as it has the potential to scale to very large and complex organizations. We are also exploring the eventual use of agent-based simulation in this space.) As a continuation the research effort with our partners in the financial industry, we plan to pilot our model with them to solicit their feedback, and hopefully improve its quality by acquiring sanitized firm data and structural information.

Thus far, our collaboration with the field study firms has focused on role management. As we move forward in exploring the deployment of role technology, we also hope to examine the nature and usefulness of *rules* in entitlement systems. The firms

we observed clearly recognize the benefits of rules in role-based systems, yet do not (nor does the research community) have a firm grasp on how to measure the effectiveness of a rule system, how to gauge the usability of a role-crafting interface, or how to integrate rule design with facilities for rule assignment, administration, and use monitoring. In particular, we have noted significant concern among all our financial research partners with "toxic combinations" of data, yet none has found a technical solution that allows them to deploy rules to prevent them. We hope that further work in this space will allow us to generate tools that enable companies to both identify and reduce the occurrence of these high-risk combinations of entitlements.

References

1. Johnson, M.E.: A Broader Context for Information Security. Financial Times 4 (September 16, 2005)
2. Scott, D., Andrijcic, E., Johnson, M.E.: Costs to the U.S. Economy of Information Infrastructure Failures: Estimates from Field Studies and Economic Data. In: Proceedings of the Fifth Workshop on the Economics of Information Security, Cambridge University (June 2006)
3. Johnson, M.E., Goetz, E.: Embedding Information Security Risk Management into the Extended Enterprise. IEEE Security and Privacy, pp. 16–24 (May–June, 2007)
4. Anderson, R.E.: Matrix Redux. Business Horizons, pp. 6–10 (November–December, 1994)
5. Burns, L.R., Wholey, D.R.: Adoption and Abandonment of Matrix Management Programs: Effects on Organizational Characteristics and Interorganizational Networks. Academy of Management Journal 36(1), 106–139
6. Ferraiolo, D., Kuhn, R.: An Introduction to Robe-Based Access Control. NIST/ITL Bulletin (December 1995), http://csrc.nist.gov/rbac/NIST-ITL-RBAC-bulletin.html
7. Sandhu, R., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-Based Access Control Models. IEEE Computer 29(2), 38–47 (1996)
8. Li, N., Mitchell, J.C.: RT: A Role-Based Trust-management Framework. In: Proceedings of The Third DARPA Information Survivability Conference and Exposition (DISCEX III) (April 2003)
9. Kunreuther, H., Heal, G.: Interdependent Security. The Journal of Risk and Uncertainty 26(2), 231–249 (2003)
10. Kunreuther, H.: Risk Analysis and Risk Management in an Uncertain World. Risk Analysis 22(4), 655–664 (2002)
11. Smith, S.W.: Humans in the Loop: Human-Computer Interaction and Security. IEEE Security and Privacy 1(3), 75–79 (2003)
12. Smith, S.W., Masone, C., Sinclair, S.: Expressing Trust in Distributed Systems: the Mismatch Between Tools and Reality. In: Forty-Second Annual Allerton Conference on Communication, Control, and Computing (September 2004)
13. Bridgestream, Products (2006), http://www.bridgestream.com/products.php
14. Donner, M., Nochin, D., Shasha, D., Walasek, W.: Algorithms and Experience in Increasing the Intelligibility and Hygiene of Access Control in Large Organizations. In: Proceedings of the IFIP TC11/ WG11.3 Fourteenth Annual Working Conference on Database Security, Kluwer, Dordrecht (2001)

Technology for Trading: What Works and What Fails

Bruce W. Weber

London Business School, London, UK
bweber@london.edu

Abstract. The computerization of financial trading has evolved differently than experts have forecast. Tremendous progress in order processing speeds and new algorithmic tools have driven market volumes well beyond projections. At the same time, expensive, high-profile technology developments have failed, and legacy practices based on intermediated floor and telephone trading have proven resilient. Based on 20 years of research on trading and market systems, this article identifies the regularities among the IT developments that have succeeded, and contrasts them with those trading technologies that have failed. We find market participants adopt new market systems when they offer advantages: (1) in openness and transparency, (2) in meeting a real need of traders, (3) in lowering cost, and (4) in retaining simplicity and ease-of-use. Examples from a number of equity and financial market technology initiatives are described. We conclude with a three-part framework for analyzing the prospects of new market systems.

1 Early Ideas

Financial markets consolidate supply and demand for securities, currencies, and derivatives contracts. By providing price discovery and ownership transfer of the most standardized and commoditized of goods: stocks, bonds, foreign currency, and futures contracts play an important role in free market economies. Beginning with the introduction of the telegraph by Samuel Morse in 1838 and the electromechanical stock ticker by Edward Calahan in 1867, information technology has been used to distribute prices from financial markets. Before the 1970s, however, determining current quotes and trading at attractive prices required a physical presence in the market. In this environment, a single market tended to dominate trading in any particular issue because communications technology was too primitive to allow transactions to occur without face-to-face contact among traders.

Since the early 1970s, trading technologies have played increasingly important roles in distributing market data, communicating buy and sell orders, establishing prices, and transferring ownership in the post-trade settlement process. The fungible, non-physical nature of financial assets has enabled I.T. to play innovative roles in establishing and refining market structures. Manual procedures have been replaced with computerization, and the value chain of trading is being transformed with straight-through-processing (STP) initiatives.

Trading in financial markets has become steadily more computerized in the past 20 years. However, the pace of change and the outcomes of IT developments have not

D.J. Veit et al. (Eds.): FinanceCom 2007, LNBIP 4, pp. 181–189, 2008.

conformed to the predictions made in the early years of market automation. Among the earliest proposals for electronic trading were Fischer Black's "automated specialist", described in 1971, and Morris Mendelson's Automated Trading System (Black, 1971), (Mendelson, 1972). Black suggested *"… that the need for floor specialists and market makers can be eliminated almost completely by an efficiently operated computer exchange."* The attractions of electronic trading were reduced errors, greater speed and capacity, but mainly the increased efficiency that market transparency would bring.

2 Incentive *Non*-alignment

Not surprisingly, the dominant exchanges and firms in the securities industry were reluctant to abandon opaque and heavily intermediated forms of trading. In an illuminating article, Stanley Ross (1990), former board member and Head of Global Trading for Deutsche Bank explained why trading margins were so large in the 1960s and 1970s before prices were electronically distributed:

> *"The client sent us a letter outlining a long-term buying programme for Peru National 4s (Eurobonds with 4% coupons) asking that certain blocks of securities be offered at the end of every month and told us the price that they would pay. We had the client who was invisible to the market, which in turn was invisible to him."*

Ross' article contrasted the prior environment with the losses suffered by London trading firms after the 1986 Big Bang reforms and introduction of the SEAQ market systems. Ross continued:

> *"Our objective of course was to make as big a 'turn' as having this contact would allow. In other words, we took what we thought the market would bear. The less the general mass of investors see of transactions being done, the greater the possibility of profit being made."*

Early attempts to computerize trading took place in an environment of profitable market making by banks and securities firms, and member-owned, not-for-profit stock exchanges. Economics teaches us that industries will seek to mute competition and other threats to rents. Since technology generally increases the bargaining power of customers over their suppliers, Wall Street and City of London securities firms were generally uninterested in and unsupportive of trading automation projects. Without 'incentive alignment' for the established players in the industry to drive the development of trading technology, market computerization progressed slowly. A number of early failures were the Ariel trading system in London (1976), the Cincinnati Stock Exchange's National Securities Trading Systems "NSTS" (1978), and the Intex futures exchange in Bermuda (1984).

3 The New Era Finally Arrives

The London Stock Exchange signed an agreement in 1983, which prevented a restrictive business practices lawsuit with the Thatcher government, and began the reforms

of its market rules. The changes culminated in 1986 with the Big Bang and the introduction of SEAQ, a screen-based trading system. After SEAQ trading launched, Clemons and Weber (1990) surveyed LSE traders and documented the benefits of the new market technology, and its competitive impact on other European exchanges.

Liquidity shifted from Stockholm, Paris, and Frankfurt to London's transparent market, and the London Stock Exchange became a leading cross-border trading center. While trading volumes increased in London, new entrants and competitive intensity meant Stock Exchange member firms' profitability decreased. In the period 1987-1990, the aggregate Return on Equity Capital for LSE members was –2%, which contrasted with a low, but still positive +6% for New York Stock Exchange members in the same period.

The impact of SEAQ's success showed that trading technology was now a competitive necessity for financial markets. Exchanges and firms could disagree on the timing of the arrival of open, transparent markets, but they could no longer safely neglect planning and investing for screen-based markets. In the late 1980s, market-providing entrants such as Instinet and Posit were beginning to attract significant order flow from securities firms and their institutional customers. Now, markets had to computerize to cut costs, enhancing trading efficiency, and retain order flow. Frits Bolkestein, European Union Commissioner of Financial Services made this point: *"Technology-driven trends have transformed the financial trading landscape: the era of utility-run stock exchanges acting as a single, uncontested national liquidity point is gone."*

4 What Have We Learned?

Several conclusions arise from the study of market computerization over the past 20 years. First, there is now a "market for markets", with technology and market models as key differentiators. Exchanges, broker-dealer firms, and market providers and data vendors are competing to offer trade execution services that will attract customers and trading volumes. Established players have reduced their resistance to potentially disruptive market IT.

The growth of technology-driven firms providing market services such as Instinet, Bridge, Bloomberg, Reuters, and ITG-Posit also stimulated new entrants to develop market systems, and greater management expertise and funding emerged for these "off-exchange" players. Wide dissemination of market quotes and information reduces the need for trader contact and lessens the pressure for trading to consolidate in an individual market. With fast, reliable linkages among liquidity pools, what might appear to be market fragmentation can actually be a virtual, consolidated market.

Second, one size does not fit all. Investors do not want to send their orders to transparent, limit order book markets that disclose price and size information. In fact, the "free option" value of submitted limit orders leads to "shading" or orders priced greater than or less than reservation prices (Schwartz, Francioni, and Weber, 2006). Market models of trading have proliferated, and trading mechanisms for converting submitted orders into trades that are favored by some investors, and ignored by others.

Small order-placers and traders needing rapid trade execution prefer the strict price and first in-first out time priorities in exchange order matching systems. Traders with

larger orders that can potentially move prices, are gravitating toward new P2P markets such as Liquidnet and Pipeline. These "block boards" facilitate bilateral "size discovery." Alternative market structures enable investors to direct their orders to the trading mechanism(s) that best suits their portfolio and investment strategy. The heterogeneity of markets models will continue to reflect the different investor segments. Roles for IT will arise in linking and integrating the various sources of liquidity today with the goal of matching trading tactics with fund management goals while minimizing all-in trading costs.

A third learning is that many of the "rules" of strategic management are difficult to apply in the financial markets industry. These complications make for challenging management and technology decisions. Three *paradoxes* of managing in an increasingly IT-intensive financial market sector are

i) *Who is a Competitor? Who is a Customer?*

A perplexing question for executives responsible for trading technology is to serve customers, while remaining aware that they will sometimes be competitors. A vivid example is provided in this paragraph from a recent Financial Times article[1]:

> *"Instinet Europe yesterday announced that it had sold a minority stake in its Chi-X equities trading platform to a group of its largest customers for an undisclosed price. Among the new shareholders are some of the largest U.S. and European securities dealers including Goldman Sachs, Morgan Stanley, Lehman Brothers, and UBS and liquidity providers Citadel and Getco. Most of these investors also own stakes in other US and European alternative trading platforms including BATS Trading, BIDS Trading and Europe-based Project Turquoise. Several, such as Credit Suisse, are also backers of a new platform which aims to challenge the CME Group in the trading of interest rate futures contracts."*

For the London Stock Exchange, Euronext, Deutsche Börse, CME, and other exchanges, these leading financial names are among their largest suppliers of order flow and liquidity. How should an exchange react when its best customers are also investing in significant rivals to the exchange? Does the exchange *put the client first*, and share information about its technology developments and strategic plans with key customers? What position does the exchange take on volume pricing and intermarket linkage when these decisions could drive more order flow to the competing trading venues.

ii) What is a Complement? What is a Substitute?

Data vendors such as Quotron, Reuters, Telerate, and Bloomberg benefited the exchanges by distributing their prices and trading data. The vendors provided value-adding analytics and maintained historic databases that aided investor decision-making and stimulated trading. Exchanges worked with the data vendor to establish data formats and

[1] Cohen , N. "Instinet brings new investors to Chi-X" FT.com, January 11, 2008.

consistent reporting. Today, many data vendors provide execution services with their data services. For instance, the Bloomberg Tradebook is a leading ECN in the U.S. providing order matching in competition with exchanges. Exchanges and market providers also co-operate with technology standards groups such as the Financial Information eXchange (FIX). From a strategy perspective, FIX trade message formats "level the playing field" and make different trading venues less differentiated. Market providers appear to have decided that proprietary trading information format are not in the best interests of customers, and focus on other sources of competitive advantage.

iii) Increasingly difficult to measure trading costs

In recent years, "dark pools" for block trading have attracted trading volumes. ITG-Posit's periodic crossing sessions introduced in 1987 were one of the first systems to given investor the ability to trade without publicly disclosing their orders. Since the failure of the Optimark system in 2000, two other non-transparent block trading systems have emerged, Liquidnet and Pipeline. By not appearing in the displayed bid-ask quote or showing in the publicly displayed liquidity, market users need to judge the value of participating in these systems subjectively for themselves and their circumstances.

Even statistical studies of trading costs that include dark pool are subject to interpretation. While the executed orders in these dark pools appear to trade cheap, there is a "negative selection" problem due to the far more costly delayed or unexecuted orders that do not match in the dark. The result for exchanges is to explain that what may appear as "high cost" trading according to transactions cost analysis (TCA) may in fact be the untraded residuals from dark pools. Any innovative market model today will be held up to TCA scrutiny, and its ability to attract participants will depend on the potentially misleading cost data.

5 Three Success Factors

Examining markets and trading practices, three dimensions that characterize successful market trading systems. There are numerous challenges for any technology innovation to have a successful commercial impact, but market trading innovations are particularly difficult. Good ideas that fail to attract a critical mass of liquidity will not succeed as trading systems. The three vital ingredients to success in new market systems developments are:

Success Driver	Description
1. Technology and Architecture	A new trading system must offer: • Speed, latency, reliability • Intuitive operation • Connectivity and openness • Audit trail and ex-post analyzability

Success Driver	Description
	These are necessary capabilities in today's environment of 40% of trade orders being generated from software algorithms. Openness means APIs are available to developers to provide direct connection into the matching engine without translations or mappings. Front-ends must be flexible and intuitive.
2. Business and Process	A well-rounded business model and economics basis for new market developments are needed. Without: • Collaboration and partnerships with important order flow providers • Pricing arrangement and fee schedules that draw in users • Well-crafted cooperation and competition plans a new market system will fail to launch successfully. An overly aggressive, "closed" approach can hinder a system finding a niche in the market as others counter the threat.
3. Community and Social Network	A market is ultimately a community of users that apply rules to their activities. • Anonymity offered – identities should only be revealed when it generates economic advantages in the market models • Membership categories and privileges – some markets benefit from restrictive entry criteria so that participants trust others in the trading system • Fairness and interactions rules must be tailored to the market model and balance rewards to both the liquidity providers, passives and the aggressive, liquidity removers

Examples of the importance of these three are detailed below:

1. Technology and Architecture

The much-heralded $350 million Optimark block trading system opened January 29, 1999. Its black box model used a rapidly sequenced auctions and a patented matching engine to generate trades from submitted buy and sell profiles or price/quantity schedules. The promise of anonymity, non-disclosure or trading submission, and sophisticated matching/pricing algorithms were seen by many as a serious threat to the established exchanges and Wall Street block trading desks. The 1998 drop in the price of a seat on the NYSE before OptiMark's debut as evidence that investors saw it as a legitimate threat to the NYSE's dominance. Seats on the NYSE had been selling for an average of $2 million in the mid-1990s, but in September 1998 a seat sold for $1.18 million. Within a year of OptiMark's debut, the price of a seat would return to about $2.5 million.

Optimark was regulated as a "Trading Facility" of the Pacific Exchange (PCX), and affiliated with Nasdaq. Its trading engine was integrated with that Intermarket Trading System (ITS) and other US exchanges to interact with resting orders in those markets. In September 1999, Optimark had 50 users and 600,000 shares a day on average with a peak day volume of 1.5 mil shares. Ultimately Optimark's computerized approach to large block order matching failed to develop a critical mass of users and liquidity, and it closed September 2000. The Optimark responded to early traders, who claimed that the system was too complex and difficult to use, by creating the OptiMark Institute, a course designed to teach traders how to use the technology. About 3,000 traders took the course between 1997 and 2000. Optimark also added three features to its user input screens, MarketBeater, Super Mid-Point, and eMarketMaker. These were intended to streamline its operation, attract additional profiles entries and enhance liquidity. OptiMark was spending about $8 million per month, but attracting little trading and fee income. Despite its efforts, in December 1999, Optimark laid off 54 staff members or 14 percent of its 389-employee workforce.

A March 24, 2002 article in the Durango Herald: "Missing the Mark: Optimark's Rise and Decline" quoted Peter Jenkins, the managing director of global equity trading at Scudder Investments, who said that while *"the OptiMark system was good in principle, it was clunky in its execution."* OptiMark may have expected too much of the traders who used it. Trading desks are inherently busy places, with people doing a lot of different things, and the OptiMark system required attention that was in short supply.

The technology and architecture of a trading system need to provide an intuitive operation for users, and the capability for users to review their activities ex-post in an audit trail. There needs to be an opportunity for learning and ex-post analyzability. Otherwise traders have no way to adjust their order placement strategies for best results the next time. Unfortunately, Optimark was complex and its "black box" opaqueness did not offer feedback for traders to adjust and perhaps enter more profiles and generate more trading volume.

2. Business and Process

A new market technology should have a business model and business processes that are sufficiently robust that they cannot be undermined by competitive rivals. Consider the CapitaLink corporate debt issuance system. It was developed in 1988 to enable issuance of $100 million to $300 million in "plain vanilla" investment grade corporate bonds. The start-up was 50% owned by J.P. Morgan. In CapitaLink, buy-side investment managers bid directly, avoiding underwriters and syndicates. It offered a 0.20% commission to undercut the standard 0.75% underwriting spread at the time.

In September 1990, John Deere & Co. was prepared to issue bonds via CapitaLink, and a polling of fund managers indicated the issue would yield 80 basis points above the comparable Treasury yield. Merrill Lynch, defending its 17% market share, bid the entire issue "net" at 70 bps before the formal CapitaLink auction. Deere accepted the highly competitive yield and Merrill "bought" the issue, but ultimately lost 15 bps on the deal. Shortly after, CapitaLink suffered a similar outcome on a US West debt issue, and JP Morgan abandoned its sponsorship. Without bank sponsorship and capital, the venture was closed. In retrospect, CapitaLink had a market model that was susceptible to 11^{th} hour scooping by investment bank underwriters, and it did not have negotiating leverage with issuers to prevent them from accepting the better Wall Street offers that materialized.

3. Community and Social Network
A new front-end trading technology will develop a community of its participants and users in much the same way that floor markets fostered camaraderie. Members share certain privileges in trading and in lower fees, and have an incentive to route orders to exchanges they are in which they are market makers or members. A new, fully electronic options exchange, the International Securities Exchange (ISE) launched in 2000, and was followed in 2004 by Boston Options Exchange (BOX).

The ISE provides membership categories and market-maker designations. Market-making firms purchase exchange memberships, and have greater trading rights and responsibilities than nonmember firms that may send orders to the market. The ISE's fixed number of memberships are either purchased or leased. Securities firms initially paid $1.5 million for an ISE competing market maker (CMM) membership and $7.5 million for a primary market maker (PMM) membership. The ISE has a total of 160 CMM memberships, owned by 22 firms, and 10 PMM memberships, owned by 8 firms. Having committed to the membership, firms appear to use the ISE more actively. BOX, on the other hand, is based on a non-exclusive, member-less market structure, and all BOX participants have identical access and trading privileges in its market structure.

Order-routing disclosures by brokerage firms to the SEC provide insights into how the major firms distribute their customers' orders across the six exchanges (Weber, 2006). Of BOX's major backers, only Citigroup Global Markets (CGM) reported more than 5 percent of its customer options order being routed to BOX in 2Q 2004. According to its Rule 11Ac1-6 report, CGM sent 7 percent of its customer option orders to BOX, while 44 percent went to the American Stock Exchange, 39 percent to CBOE and 6 percent to the Pacific Exchange. Unless the BOX can attract a similarly committed group of order-flow providers from among its investors – CSFB, J.P. Morgan Chase, Salomon Smith Barney and UBS (USA), along with one of the three founding organizations, Interactive Brokers Group – it will remain one of the smaller options exchanges.

In contrast to BOX, the ISE's members are among its leading users. Many of ISE's leading users, including Morgan Stanley and Goldman Sachs, also serve as PMMs and CMMs. The scarcity and exclusivity of their market memberships may be providing a further incentive for routing orders to the ISE market. For all the theoretical advantages of an inclusive approach to memberships, new trading systems that do not create a committed network of users will find the going difficult.

6 Ripe for Research

The growth and importance of financial markets make the issues around optimal trading technology valuable topics for research. Market structure principles and market quality metrics are well developed for financial markets, and technology opens new possibility in market organization. Research methods to study trading technology and its economic impacts include:

- empirical data analysis
- experimental economics
- simulation of market models

Markets are changing in response to customer needs and new trading technologies. The role of information systems has evolved from disseminating price data and supporting clerical functions of floor-based markets, to providing fully, open electronic markets. Yet, the business success of new market technologies is far from certain. A number of principles though in the design, business model, and user network appear to able to drive trading technology to greater adoption level and better contribution to the marketplace.

References

Early electronic market models

Black, F.: Toward a Fully Automated Exchange. Financial Analysts Journal, 29–44 and 25–29 (July-August and November-December, 1971)

Clemons, E., Weber, B.: London's Big Bang: A Case Study of Information Technology, Competitive Impact, and Organizational Change. Journal of Management Information Systems 6(4), 41–60 (1990)

Cohen, K., Maier, S., Schwartz, R., Whitcomb, D.: The Microstructure of Securities Markets. Prentice-Hall, Englewood Cliffs (1986)

Hamilton, J.: Electronic Market Linkages and the Distribution of Order Flow: The Case of Off-Board Trading of NYSE-listed Stocks. In: Lucas Jr, H., Schwartz, R. (eds.) The Challenge of Information Technology for the Securities Markets: Liquidity, Volatility, and Global Trading, Dow Jones-Irwin (1989)

Mendelson, M.: From Automated Quotes to Automated Trading: Restructuring the Stock Market in the U.S., New York University Bulletins Nos, pp. 80–82 (1972)

Peake, J., Mendelson, M., Williams Jr, R.: Toward a Modern Exchange: The Peake-Mendelson-Williams Proposal for an Electronically Assisted Auction Market. In: Bloch, E., Schwartz, R. (eds.) Impending Changes for Securities Markets – What Role for Exchanges?, JAI Press (1979)

Analysis of Market Quality with Trading Technology

Clemons, E., Weber, B.: Information Technology and Screen-Based Securities Trading: Pricing the Stock and Pricing the Trade. Management Science 43(12), 1693–1708 (1997)

Cohen, K., Conroy, R., Maier, S.: Order Flow and the Quality of the Market. In: Amihud, Y., Ho, T., Schwartz, R. (eds.) Market Making and the Changing Structure of the Securities Industry, Lexington Books, pp. 93–109 (1985)

Domowitz, I., Steil, B.: Automation, Trading Costs, and the Structure of the Securities Trading Industry. Working Paper, Royal Institute of International Affairs, London, UK (February 1998)

Garbade, K., Silber, W.: Technology, Communication and the Performance of Financial Markets: 1840-1975. Journal of Finance 33(3), 819–831 (1978)

Pagano, M., Steil, B.: Equity Trading I: The Evolution of European Trading Systems. In: Steil, B. (ed.) The European Equity Markets: The State of the Union and an Agenda for the Millennium. European Capital Markets Institute and the Royal Institute of International Affairs, London (1996)

Schwartz, R., Francioni, R., Weber, B.: The Equity Trader Course. John Wiley & Sons, Chichester (2006)

Weber, B.: Adoption of electronic trading at the International Securities Exchange. Decision Support Systems 41(4), 728–746 (2006)

Author Index