

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into

- seminars
- proceedings
- dissertations
- thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: <http://www.gi.de/service/publikationen/lni/>

ISSN 1617-5468

ISBN 978-3-88579-616-9

EMISA 2013 is the fifth international workshop in a series that provides a key forum for researchers and practitioners in the fields of enterprise modeling and the design of information system (IS) architectures. The workshop series emphasizes a holistic view on these fields, fostering integrated approaches that address and relate business processes, business people and information technology. EMISA'13 will provide an international forum to explore new avenues in enterprise modeling and the design of IS architectures by combining the contributions of different schools of Information Systems, Business Informatics, and Computer Science.



# GI-Edition

## Lecture Notes in Informatics

**Reinhard Jung, Manfred Reichert (Eds.)**

## Enterprise Modelling and Information Systems Architectures (EMISA 2013)

**St. Gallen, Switzerland  
September 5 - 6, 2013**

Reinhard Jung, Manfred Reichert (Eds.): EMISA 2013







Reinhard Jung, Manfred Reichert (Eds.)

**Enterprise Modelling  
and Information Systems  
Architectures  
(EMISA 2013)**

**Fifth International Workshop on Enterprise Modelling  
and Information Systems Architectures**

**September 5 – 6, 2013  
St. Gallen, Switzerland**

## **Lecture Notes in Informatics (LNI) - Proceedings**

Series of the Gesellschaft für Informatik (GI)

Volume P-222

ISBN 978-3-88579-616-9

ISSN 1617-5468

### **Volume Editors**

Prof. Dr. Reinhard Jung

University of St.Gallen

Institute of Information Management

9000 St.Gallen, Switzerland

Email: reinhard.jung@unisg.ch

Prof. Dr. Manfred Reichert

University of Ulm

Institute of Databases and Information Systems

89069 Ulm, Germany

Email: manfred.reichert@uni-ulm.de

### **Series Editorial Board**

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria

(Chairman, mayr@ifit.uni-klu.ac.at)

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, Hochschule für Technik, Stuttgart, Germany

Ulrich Frank, Universität Duisburg-Essen, Germany

Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany

Michael Goedicke, Universität Duisburg-Essen, Germany

Ralf Hofestädt, Universität Bielefeld, Germany

Michael Koch, Universität der Bundeswehr München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany

Sigrid Schubert, Universität Siegen, Germany

Ingo Timm, Universität Trier, Germany

Karin Vosseberg, Hochschule Bremerhaven, Germany

Maria Wimmer, Universität Koblenz-Landau, Germany

### **Dissertations**

Steffen Hölldobler, Technische Universität Dresden, Germany

### **Seminars**

Reinhard Wilhelm, Universität des Saarlandes, Germany

### **Thematics**

Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2013

**printed by** Köllen Druck+Verlag GmbH, Bonn

**Organizer**

University of St.Gallen (HSG)  
Institute of Information Management  
9000 St.Gallen, Switzerland

Prof. Dr. Reinhard Jung (Workshop Chair and Programme Committee Co-Chair)  
Prof. Dr. Manfred Reichert (Programme Committee Co-Chair)  
Torben Küpper (Organization)

The workshop is jointly organized by the GI Special Interest Group on Modelling Business Information Systems (*GI-SIG MobIS*) and the GI Special Interest Group on Design Methods for Information Systems (*GI-SIG EMISA*) and the Swiss Informatics Society (*SI*).

*GI-SIG EMISA*: The GI Special Interest Group on Design Methods for Information Systems provides a forum for researchers from various disciplines who develop and apply methods to support the analysis and design of information systems.

*GI-SIG MobIS*: Conceptual Modelling is pivotal for analysing and designing information systems that are in line with a company's long term strategy and that efficiently support its core business processes. The Special Interest Group on Modelling Business Information Systems (SIG MobIS) within the German Informatics Society (GI) is aimed at providing a forum for exchanging ideas and solutions on modelling research within Information Systems both for researchers at universities and experts in industry.

**Programme Committee**

Witold Abramowicz (University of Poznan)  
Antonia Albani (University of St.Gallen)  
Thomas Allweyer (University of Applied Sciences Kaiserslautern)  
Colin Atkinson (University of Mannheim)  
Lars Bækgaard (University of Aarhus)  
Ulrike Baumöel (University of Hagen)  
Khalid Benali (University of Lorraine)  
Martin Bertram (Commerzbank, Frankfurt)  
Patrick Delfmann (European Research Center for Information Systems (ERCIS))  
Jörg Desel (University of Hagen)  
Werner Esswein (Technical University of Dresden)  
Fernand Feltz (University of Belvaux)  
Ulrich Frank (University of Duisburg-Essen)  
Andreas Gadatsch (University of Applied Sciences Bonn-Rhein-Sieg)

Norbert Gronau (University of Potsdam)  
Elmar J. Sinz (University of Bamberg)  
Heinrich Jasper (Technical University of Freiberg)  
Stefan Klink (University of Karlsruhe)  
Ralf Klischewski (German University of Kairo)  
Horst Kremers (CODATA, Germany)  
John Krogstie (University of Trondheim)  
Susanne Leist (University of Regensburg)  
Heinrich C. Mayr (Alpen-Adria-University of Klagenfurt)  
Jan Mendling (University of Vienna)  
Mirjam Minor (University of Trier)  
Bernd Müller (University of Applied Sciences Wolfenbüttel)  
Bela Mutschler (University of Applied Sciences Ravensburg-Weingarten)  
Volker Nissen (Technical University of Ilmenau)  
Markus Nüttgens (University of Hamburg)  
Andreas Oberweis (Karlsruhe Institute of Technology)  
Sven Overhage (University of Augsburg)  
Hansjürgen Paul (Institute of Work and Technique)  
Erik Proper (Luxembourg and Radboud University Nijmegen)  
Michael Rebstock (University of Applied Sciences Darmstadt)  
Stefanie Rinderle-Ma (University of Vienna)  
Peter Rittgen (University of Borås)  
Michael Rosemann (Queensland University of Technology)  
Matti Rossi (Aalto University)  
Gunter Saake (University of Magdeburg)  
Eike Schallehn (University of Magdeburg)  
Carlo Simon (Proxadis University of Applied Sciences)  
Stefan Strecker (University of Hagen)  
Oliver Thomas (University of Osnabrück)  
Klaus Turowski (University of Magdeburg)  
Marten van Sinderen (University of Twente)  
Gottfried Vossen (University of Münster)  
Barbara Weber (University of Innsbruck)  
Hans Weigand (University of Tilburg)  
Mathias Weske (University of Potsdam)

## Preface

Modern organizations recognize the need for a close alignment of their business with information technology. In turn, this requires the co-design and -evolution of the organization and its information systems, considering the corporate strategy and business processes as well as the information systems supporting them. In this context, the complexity inherent to such a co-design should be reduced and the cultural chasm between business people and IT professionals be overcome.

Conceptual models of the enterprise as well as information systems architectures represent important means to tackle these challenges. Enterprise models integrate conceptual models of information systems and models of the surrounding action systems (e.g., business process models) and, hence, take technical, organizational, and economic aspects of the organization into account. Information systems architectures provide ‘blueprints’ for the design and implementation of software systems and complement enterprise models in the co-design of the organization and its information systems. Both serve as a means to foster communication and cooperation between the various stakeholders of an enterprise. At the same time, research on enterprise models and information systems architectures requires the cooperation of disciplines such as Information Systems, Business Informatics, and Computer Science.

The *5th International Workshop on Enterprise Modelling and Information Systems Architectures* (EMISA’13) addresses different aspects relevant for enterprise modelling as well as for designing enterprise architectures in general and information systems architectures in particular. It is jointly organized by the GI Special Interest Group on Modelling Business Information Systems (GI-SIG MoBIS) and the GI Special Interest Group on Design Methods for Information Systems (GI-SIG EMISA) and the Swiss Informatics Society (SI).

These proceedings feature a selection of high quality contributions from academia and practice on enterprise architecture modeling, business processes management, information systems engineering, and other relevant issues in enterprise modelling and information systems architectures. In total, we received 24 submissions that were all thoroughly reviewed by at least two experts of the program committee; 12 out of these 24 submissions were finally accepted as full paper.

We would like to thank the members of the program committee and the reviewers for their efforts in selecting the papers. They helped us to compile a high-quality workshop program. Furthermore, we want to acknowledge the splendid support of Torben Küpper who was in charge of the local organization and the preparation of the proceedings. We also thank Ulrich Frank for being the keynote speaker of EMISA’13. Finally, we are grateful for the generous support of the University of St. Gallen and its School of

Management. We hope you will find the papers in this volume interesting and stimulating.

St. Gallen, September 2013

Reinhard Jung and Manfred Reichert  
PC Co-Chairs

# Directory

## Enterprise Architecture

**Nils Labusch, Stephan Aier, Robert Winter**

*Beyond Enterprise Architecture Modeling – What are the Essentials to Support Enterprise Transformations?* ..... 13

**Ralf Abraham, Hella Niemietz, Sybren de Kinderen, Stephan Aier**

*Can boundary objects mitigate communication defects in enterprise transformation? Findings from expert interviews* ..... 27

**Thomas Benker**

*A Hybrid OLAP & OLTP Architecture Using Non-Relational Data Components* ..... 41

## Methods

**Jeannette Stark, Martin Burwitz, Richard Braun, Werner Esswein**

*Cognitive Efficient Modelling Using Tablets*..... 57

**Rafael Accorsi, Thomas Stocker**

*SecSy: Synthesizing Process Event Logs* ..... 71

**Heiko Kattenstroth, Ulrich Frank, David Heise**

*Towards a Modelling Method in Support of Evaluating Information Systems Integration* ..... 85

## Process Modelling

**Michael Fellmann, Sebastian Bittmann, Arne Karhof, Carl Stolze, Oliver Thomas**

*Do We Need a Standard for EPC Modelling? The State of Syntactic, Semantic and Pragmatic Quality* ..... 103

|                                                                                                                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Michael Fellmann, Novica Zarvic, Annelene Sudau</b><br><i>Ontology-Based Assistance for Semi-Formal Process Modeling</i> .....                                                                             | 117 |
| <b>Nico Clever, Justus Holler, Maria Shitkova, Jörg Becker</b><br><i>Towards Auto-Suggested Process Modeling – Prototypical Development of an<br/>Auto-Suggest Component for Process Modeling Tools</i> ..... | 133 |
| Business Process Modelling                                                                                                                                                                                    |     |
| <b>Dirk Van Der Linden, Henderik Proper</b><br><i>On the cognitive understanding of types in modeling languages</i> .....                                                                                     | 149 |
| <b>Nicolas Pflanzl, Gottfried Vossen</b><br><i>Human-Oriented Challenges of Social BPM: An Overview</i> .....                                                                                                 | 163 |
| <b>Matthias Geiger, Guido Wirtz</b><br><i>BPMN 2.0 Serialization - Standard Compliance Issues and Evaluation of<br/>Modeling Tools</i> .....                                                                  | 177 |

# **Enterprise Architecture**



# Beyond Enterprise Architecture Modeling – What are the Essentials to Support Enterprise Transformations?

Nils Labusch, Stephan Aier, Robert Winter

Institute of Information Management  
University of St.Gallen  
Mueller-Friedberg-Str. 8  
CH-9000 St. Gallen  
{ nils.labusch | stephan.aier | robert.winter }@unisg.ch

**Abstract:** In recent years, many different modeling techniques and languages were developed in order to allow for an efficient and appropriate enterprise architecture management (EAM). Recently, EAM is no longer only seen as a means to ensure business/IT alignment but further as a means to support fundamental changes of the organization, often called enterprise transformation (ET). In a joint project with a group of practitioners we aim at developing a framework that provides guidance on how EAM can support such ETs and thus leverage the benefit of modeling. Our design results after more than one year of research reveal different types of EAM, areas of action in ETs and ET activities that can be supported by EAM. The findings show that modeling techniques or languages should focus on constructs like transitions, benefits or risks in order to increase their value for ET support.

## 1 Introduction

While enterprise architecture (EA) describes the fundamental structures of an enterprise, EAM is concerned with the establishment and coordinated development of the EA in order to consistently respond to business and IT goals, opportunities, and necessities [Op11]. Core of EAM are manifold modeling techniques and models that document and consolidate the relevant information and allow for understanding cross-company relations [La13; St12; WF07].

In recent years, many different modeling techniques and languages were developed in order to allow for an efficient and appropriate EAM (e.g. the Archimate language as a standard [Op12] or vendor dependent languages). The traditional task of EAM is providing guidance on the achievement of business to IT alignment [BY07], many frameworks aiming at this task exist (e.g. TOGAF [Op11]). However, EAM increasingly claims to support tasks that are beyond traditional and often limited business to IT alignment challenges. One major example is the support of complex and large scale changes in an organization.

These changes, often called enterprise transformations (ET), are not routine since they substantially alter an organization's relationships with its key constituencies like

customers, suppliers or regulators. ETs may result in new value propositions, they may provide old value propositions in fundamentally new ways or they may change the inner structure of the enterprise [RB06]. Examples are transformations of the business model [As11], mergers & acquisitions [JM00] or introductions and replacements of enterprise information systems [BSS10; HTW97; SL99]. Many transformations fail for a variety of reasons [Ko95; SL99] like underestimated technical complexity, or lacks in either portfolio or benefits planning [FB12].

For these reasons, EAM is believed to support the management of ETs [ABA09] by guiding the necessary coordination efforts [AAL12; HPK09; PNL07] and providing information for management support or strategy development [ABA09]. IT is often an important part of the ET but further aspects are as important or even more important.

Thus, in the research project that we conduct with partners from corporate practice, we aim at investigating what the important parts of managing an ET are and how they can be supported by the models and techniques that EAM provides. Thus, we are guided in the project by the following research question:

*RQ: How can the management of enterprise transformations be supported by EAM?*

We proceed as follows: We present related work that explores the link between EAM and ET management (ETM). We go on with presenting our design approach and provide a brief description of the current work status concerning our designed framework. In section five we discuss the achieved state of work and conclude with a summary and implications for future work in the last section.

## **2 Related Work**

Winter et al. [Wi12] illustrate the relation of EAM and ETM. Their main findings show that current EAM is primarily conducted in the IT and operations departments (whereas ETM is often part of the business departments), EAM primarily focusses on current and target states (whereas ETM focusses on the process in between these states) and EAM requires experts with analytical experience (whereas ETM requires people with very mature communication and politics skills). Keller & Price [KP11] take a managerial perspective and include “architect” as one of five transformation stages (within “aspire”, “assess”, “act” and “advance”). They describe activities like breaking down the transformation initiative into a manageable portfolio, identifying skills or setting up formal reinforcement mechanisms as part of the “architecting”.

Authors that deal explicitly with EAM identify similar potentials. Harmsen et al. [HPK09] propose to use EAM as a governing function in order to streamline a portfolio of transformation steps that need to be well aligned in order to be successful. The authors consider EAM suitable to ensure this – especially in areas like strategic direction (investigate alternatives), gap analysis, tactical planning (identify intermediate milestones), operational planning, selection of partial solutions, e.g. based on standards [BY07], or solution crafting (which describes the identification of tasks in projects).

Radeke [Ra11] discusses, how EAM can contribute to the strategic change process. He finds that EAM has the potential to improve the strategic fit of an enterprise with its market environment, to improve business/IT alignment, and to improve the preparedness for change through standardization and modularization of the respective enterprise. According to Pulkkinen et al. [PNL07] EAM allows groups to interpret the related issues for their purposes. The guidelines and principles agreed on with the collaborative EAM work facilitate plans and designs for interoperability and synergy of systems.

Focusing on a modeling perspective, Aier & Gleichauf [AG10] describe what is necessary in order to not only model to-be and as-is states but to model the transformation between these states. They describe different types of necessary intermediary transformation models. Steinhorst et al [St12] describe an approach that allows for an analysis of models on a structural and semantic level. The approach allows for a detection of weakness and best practice patterns concerning transformations in existing models. McGinnis [Mc07] provides conditions that models need to fulfill in order to be used in ETs. He claims that such models facilitate the education of employees and can be used as a foundation for large-scale IT implementations (like ERP or CRM systems). In addition they can be used to predict how the enterprise reacts to future scenarios or assess alternative processes, customer or supplier relations. He further states that such models need to be based on syntax and semantics that are driven from a business and non-IT implementation perspective.

Summarized, EAM is considered to have a high potential to support ETs. Lots of research about EAM is conducted and many practitioners are dealing with the topic. So far, however, EAM is mostly concentrating on business/IT alignment issues. Evidence for these problems can be found in the formal specifications of the EAM frameworks currently available. For example the content meta-model of TOGAF [Op11] is comprised half of IT related and half of business related elements. Other aspects beyond business/IT alignment that may be relevant to ET, e.g. skill development or role definitions, are less explicitly illustrated. Identifying these aspects and opening up the field of ET for EAM and its modeling core would leverage the manifold research in the modeling domain.

Thus, together with consortia partners, we aim at designing and consolidating a framework that prescribes where and how models and techniques from EAM can support the management of enterprise transformations. Such a framework aims at bringing together the ET and the EAM perspective and it aims at alleviating communication defects between the stakeholder groups involved.

### **3 Research Design**

In the paper at hand we focus on the description and discussion of the design research process [Pe07]. Our research approach follows the general design cycle idea as introduced by Hevner et al. [34] and particularly follows the more recent understanding of design science that assumes alternating core activities of design and evaluation within one and the same cycle [WA12].

### 3.1 Research Setting

We conducted the design process in a consortium of researchers and practitioners [OO12] that met four times during the years 2011 and 2012 in order to develop the framework and to apply it within their organizations later on. The group was comprised of eleven experts being employed with seven organizations located in different industries (public sector, insurance, utilities). The participants had long-time experience in their domains and access to further experts within their companies [OO12]. Table 1 provides an overview about participants and their organizations.

Table 1. Design Partners

| Company | Industry      | Informants                                                                                   |
|---------|---------------|----------------------------------------------------------------------------------------------|
| A       | Insurance     | Head Enterprise Architecture<br>IT Project Manager                                           |
| B       | Insurance     | Head of IT Service Management<br>Head of Enterprise Architecture                             |
| C       | Public Sector | Head of IT Strategy                                                                          |
| D       | Utilities     | Enterprise Architect                                                                         |
| E       | Insurance     | Head of Enterprise Architecture<br>Data Architect<br>Enterprise and Data Warehouse Architect |
| F       | Utilities     | Enterprise Architect                                                                         |
| G       | Insurance     | Vice President Enterprise Architecture                                                       |

The meetings contained four major elements: First, keynotes in which one of the informants reported about practices from his organization. Second, external input by experts that where not participating regularly in the meetings but could provide the core group with new and challenging perspectives – we especially invited external experts to avoid biases and to ensure reliability. Third, academic input: the researchers presented findings and implications from theory. Finally, workshop sessions: the participants were asked to conduct different tasks and discussions moderated and supported by the research team. The meetings usually took two days.

In the time between the meetings with the practitioners, our research team conducted internal workshops and discussions in order to provide input to the practitioners. In order to avoid biases for the design decisions by the research team, one person – similar to the external experts in the practitioner workshops – was taking the role of a devil’s advocate [HE77] and thus purposefully provided an opposite opinion.

### 3.2 Design Process

The process itself contained four major iterations. In a *first iteration*, we identified the problem as stated above and ensured its relevance during discussions with the partners.

Thus, EAM is perceived to support ET, but current approaches seem not to be fully sufficient to achieve this goal due to their focus on business/IT alignment. We conducted a first design-cycle in the research team where we surveyed the EAM knowledge base for appropriate kernel theories in order to allow for an approach “informed by theory” [Gr06]. Based on its explicit link to transformation and the focus of the project, we identified the body of knowledge about dynamic capabilities as a very helpful means. Barreto [Ba10] summarizes a dynamic capability as “the firm’s potential to systematically solve problems, formed by its propensity to sense opportunities and threats, to make timely and market-oriented decisions, and to change its resource base.” Teece [Te07] subdivides dynamic capabilities into the areas of sensing, seizing and transformation and thus includes the fundamental change as core. Abraham et al. [AAW12] consider EAM as a dynamic capability and thus link the areas of ET and EAM on a sound theoretical foundation. According to Abraham et al. [AAW12] each type of change needs a different type of EAM: A fast but lean type for unpredictable changes and a rather traditional type of EAM for planned changes. We further distinguished the traditional type into a business-related and a rather IT-focused type of EAM since these are the ones regularly mentioned in theory (e.g. [La09]) or seen in practice like in our group of practitioners. We applied the wording and content used in the theories also for the discussions during the practitioner discussions. For example, the theoretical view helped to understand that transformation activities could be distinguished into sensing, seizing and actually implementing the changes.

The theoretical findings were discussed and further refined in the group of practitioners. Thus, with this step we were able to reduce the size of the relevant EAM knowledge base into smaller pieces, depending on which type of EAM should be considered in more detail.

In a *second iteration* we identified areas of action, where EAM potentially can support ETs. We first conducted this step openly in the group of practitioners to collect their experiences and perceptions. This resulted in four major areas: (1) *Rolls, skills & communication* deals with the question, which roles and skills are necessary during a transformation from the point of view of EAM and how communication with other stakeholders being involved in transformations can be improved. (2) *Governance & control* deals with the governance-processes that are necessary for transformations and can be supported by EAM. (3) *Planning & requirements management* is concerned with techniques and tasks that are relevant for planning ETs. The fourth identified area was (4) *organizational culture* that was considered to be an important context factor.

In the *third iteration*, we refined the results. For example, we considered the area of roles and communication. In the research team, we identified the kernel theory of boundary objects [DM12] as a helpful means to explain, how EA can foster the communication among different stakeholder groups. The theory shows that certain objects like models, commonly used frameworks etc. overcome barriers like different language between different areas of business or between business and IT. In the practitioner group boundary objects from the different companies were identified, e.g. capability maps, application landscapes but also more unusual objects like a project

interaction room (a fixed room that is used to discuss a certain topic by different stakeholder groups).

While the design process up to this point was driven a lot by an architectural perspective on ET, we shifted the perspective towards a more business oriented approach in the *fourth iteration* by inviting external experts that deal on the one hand with a holistic transformation perspective and on the other hand with soft factors and a change-centered perspective. The overall goal for this meeting was to become more specific about activities that are necessary during transformations and could potentially be supported by EAM. We further aimed at integrating those in a consolidated framework. As a foundation we used the BTM<sup>2</sup> framework [UG12] that aims at covering the management of transformations, based on a holistic perspective. In the team of researchers we discussed, which of the given activities in the framework could be supported by EAM. In the next step, we handed over this discussion to the EAM practitioner group. In here, we even went one step further and asked, if EAM could support the transformation activities depending on the EAM type (lean, traditional, business) that we investigated in the first iteration. We conducted three cycles in total in this workshop session, in each cycle the groups were mixed again in order to ensure a direct evaluation of the findings (world-café method [Wo13]). As the result, some of the activities were marked as not supportable by any of the types, some only by one type and some by all three types of EAM. The overall design process that we conducted so far is illustrated in Table 2.

Table 2. Design Process

| Iteration | Event                             | Meeting Program                                                                                                                                                                                                                     |
|-----------|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1         | Whole Group Meeting               | <ul style="list-style-type: none"> <li>• EAM for ET: Idea and Necessities (speech, researcher)</li> <li>• Exchange of experiences (workshop, all)</li> <li>• Consolidation of challenges and experiences (workshop, all)</li> </ul> |
|           | Research team internal discussion | <ul style="list-style-type: none"> <li>• Learning from dynamic capabilities</li> </ul>                                                                                                                                              |
| 2         | Whole Group Meeting               | <ul style="list-style-type: none"> <li>• Identification of EAM capabilities for ET (workshop, all)</li> <li>• Mapping of EAM capabilities to solution areas (workshop, all)</li> <li>• Derivation of “areas of action”</li> </ul>   |
|           | Research team internal discussion | <ul style="list-style-type: none"> <li>• Learning from boundary objects</li> </ul>                                                                                                                                                  |

| Iteration | Event                             | Meeting Program                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3         | Whole Group Meeting               | <ul style="list-style-type: none"> <li>• Architectural Support of ET – The perspective of transformation management (speech, researcher)</li> <li>• Boundary objects: roles, skills and communication (speech, informant)</li> <li>• Boundary objects: roles, skills and communication (workshop, all)</li> <li>• Controller and IT: the changing roles (speech, external expert)</li> <li>• Governance processes (speech, informant)</li> <li>• Governance processes (workshop, all)</li> <li>• EAM planning processes (workshop, all)</li> </ul>                        |
|           | Research team internal discussion | <ul style="list-style-type: none"> <li>• Pre-consolidation of the framework</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 4         | Whole Group Meeting               | <ul style="list-style-type: none"> <li>• BTM<sup>2</sup>: Overview, development and application (speech, external expert)</li> <li>• Capability models as a management tool (workshop, all)</li> <li>• Value management as a connection to the business (workshop, all)</li> <li>• Change management: a practical view behind the scenes (speech, external expert)</li> <li>• Theoretical perspectives on culture and acceptance (speech, researcher)</li> <li>• Consolidating the Framework: Transformation from an architectural perspective (workshop, all)</li> </ul> |
|           | Research team internal discussion | <ul style="list-style-type: none"> <li>• Refinement of the framework</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

## 4 Towards a Framework for the Architectural Support of Enterprise Transformations

In the following section we provide a brief overview of the results that we achieved so far with our partners in order to allow for an understanding of the framework.

### 4.1 Overall Structure

The main constructs in the framework are the identified types of EAM, the respective areas of action concerning ETM and the ETM activities that according to our research process (described above) can be supported by EAM. Like described above, we identified a lean type of EAM to deal with rather sudden transformations, a traditional type of EAM for planned changes concerning mostly IT issues and a business-related

type of EAM. The framework is further comprised of areas of action that can be supported by EAM during a transformation. These are “rolls, skills & communication”, “governance & control”, “planning & requirements management” and “organizational culture”. In order to deal with the complexity, each ETM activity belongs to one ET area of action but can, of course, be supported by more than one type of EAM. This is summarized in Table 3.

Table 3. Overall Relation

| Transformation Activity                   | EAM Type        |                       |          | ET Area of Action             |                      |                                    |                        |
|-------------------------------------------|-----------------|-----------------------|----------|-------------------------------|----------------------|------------------------------------|------------------------|
|                                           | IT oriented EAM | Business oriented EAM | Lean EAM | Rolls, skills & communication | Governance & Control | Planning & requirements management | Organizational culture |
| Setup communication                       | X               |                       |          | X                             |                      |                                    |                        |
| Establish communities of practice         | X               |                       |          | X                             |                      |                                    |                        |
| Manage training                           | X               |                       |          | X                             |                      |                                    |                        |
| Orchestrate skills & disciplines          | X               |                       |          |                               | X                    |                                    |                        |
| Manage and measure principles             | X               |                       |          |                               | X                    |                                    |                        |
| Conduct lifecycle management              | X               |                       |          |                               | X                    |                                    |                        |
| Monitor change                            | X               |                       |          |                               | X                    |                                    |                        |
| Develop roadmap plan                      | X               |                       |          |                               |                      | X                                  |                        |
| Analyze cultural environment              |                 |                       |          |                               |                      |                                    | X                      |
| Establish common language                 | X               | X                     |          | X                             |                      |                                    |                        |
| Conduct stakeholder management            | X               | X                     |          | X                             |                      |                                    |                        |
| Establish transformation lifecycle        | X               | X                     |          |                               | X                    |                                    |                        |
| Establish potentials for further benefits | X               | X                     |          |                               | X                    |                                    |                        |
| Analyze initial situation                 | X               | X                     |          |                               |                      | X                                  |                        |
| Develop integrated transformation plan    | X               | X                     |          |                               |                      | X                                  |                        |
| Analyze needs & maturity level            | X               | X                     |          |                               |                      | X                                  |                        |
| Develop detailed business case            | X               | X                     |          |                               |                      | X                                  |                        |
| Manage requirements                       | X               | X                     |          |                               |                      | X                                  |                        |
| Define overall goals                      |                 | X                     |          |                               |                      | X                                  |                        |
| Develop high-level business case          |                 | X                     |          |                               |                      | X                                  |                        |

| Transformation Activity                           | EAM Type        |                       |          | ET Area of Action             |                      |                                    |                        |
|---------------------------------------------------|-----------------|-----------------------|----------|-------------------------------|----------------------|------------------------------------|------------------------|
|                                                   | IT oriented EAM | Business oriented EAM | Lean EAM | Rolls, skills & communication | Governance & Control | Planning & requirements management | Organizational culture |
| Define KPIs and benchmark                         |                 | X                     |          |                               |                      | X                                  |                        |
| Plan benefit realization                          |                 | X                     |          |                               |                      | X                                  |                        |
| Perform 360° strategic risk assessment            |                 | X                     |          |                               |                      | X                                  |                        |
| Define risk strategy                              |                 | X                     |          |                               |                      | X                                  |                        |
| Conduct program planning                          |                 | X                     |          |                               |                      | X                                  |                        |
| Manage program scope                              |                 | X                     |          |                               |                      | X                                  |                        |
| Assess change readiness                           |                 | X                     |          |                               |                      | X                                  |                        |
| Analyze & set cultural environment                |                 | X                     |          |                               |                      |                                    | X                      |
| Orchestrate skills & disciplines: propose experts |                 | X                     | X        |                               | X                    |                                    |                        |
| Evaluate risk for transformation business case    |                 | X                     | X        |                               |                      | X                                  |                        |
| Manage communication and translation              |                 |                       | X        | X                             |                      |                                    |                        |
| Review and evaluate results                       |                 |                       | X        |                               | X                    |                                    |                        |
| Conduct ex-post program alignment                 |                 |                       | X        |                               | X                    |                                    |                        |
| Conduct risk monitoring                           | X               |                       | X        |                               | X                    |                                    |                        |
| Manage communication                              | X               | X                     | X        | X                             |                      |                                    |                        |
| Align with risk management                        | X               | X                     | X        |                               |                      | X                                  |                        |
| Assess as-is capabilities                         | X               | X                     | X        |                               |                      | X                                  |                        |
| Design to-be architecture                         | X               | X                     | X        |                               |                      | X                                  |                        |
| Perform gap analysis                              | X               | X                     | X        |                               |                      | X                                  |                        |

Consider, for example, the activity “Manage Communication”. In the table above we can see that it was considered important concerning all three types of EAM and belongs to the “rolls, skills & communication” area of action. At the current stage of the research project, we aim at investigating, how EAM can support ETM in terms of currently available techniques and models. Thus, we discussed this issue in the workshop and documented the findings in a structured way. We used a structure that is similar to Bucher & Dinter [BD12] and the TOGAF framework [Op11]. The most important parts are the actual activities (what is done) and techniques (how is it done). Figure 1 provides an example of the “manage communication” activity.

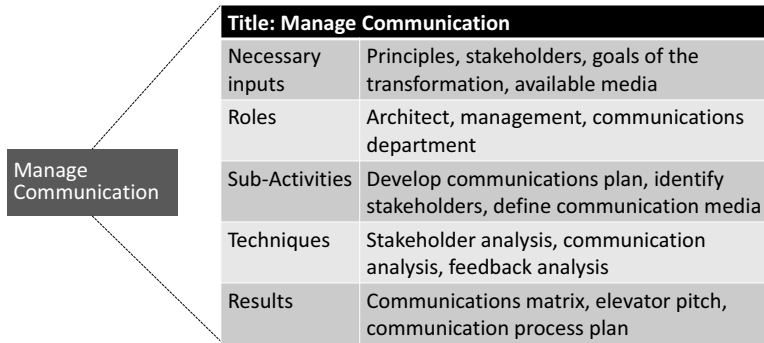


Figure 1. Example of an Activity

The example shows that transformation managers need inputs like principles, stakeholders etc. in order to successfully conduct the activity. EAM can offer support by e.g. using models that include the stakeholders and provide those to ET managers or, depending on how EAM is established in the organization address the stakeholders by themselves.

Summarized, the framework provides two degrees of abstraction: (1) The overview level with types and areas of actions shows where the EAM knowledge and models could be used in order to support ETM. (2) The concrete activities and the included details illustrate how EAM could be used in detail.

## 4.2 Framework Application

The proposed framework can be further customized for single organizations and the scope of the future EAM in the specific company can be determined. If e.g. a company has an EAM rooted in the IT department and dealing with IT questions only, the information that such EAM can provide are relevant for most transformations. However, the department should focus on supporting ET activities that they are familiar with (e.g. development of roadmaps). The situation is different, if EAM should be set up as an ET supporting discipline, rooted in the business departments. Further activities, like defining key performance indicators (KPI), should be considered relevant. These two examples illustrate that an incorporation of the framework differs on the concrete enterprise it is applied in. The artifact can provide guidance in both (and more) cases by providing direction on techniques and results that the EAM knowledge base can provide.

An implementation in one specific enterprise requires workshop sessions with the most important stakeholders in this organizational environment in order to get insights about necessary priorities and preferences.

## 5 Discussion

In the paper at hand we ask, how important parts of managing enterprise transformations can be supported by EAM and thus, how the potential of the manifold existing modeling techniques and languages can be leveraged in a promising field.

Our research shows that EAM can provide useful inputs to the management of transformations – not just business/IT alignment but also business/business alignment. In the group discussions it became clear that “the business” is not a monolithic unit and their interests are heterogeneous. The term business to business alignment might be sufficient when it comes to transformations. It became clear during the discussions that certain transformation activities are supported rather natively by EAM. These are characterized by a utilization of “native” EAM outputs like risk assessments, IT-application landscapes etc. Concerning these activities it is straight forward for EAM to support the ET and provide appropriate models.

Concerning other ET activities, EAM has a high development potential in order to become a supportive means. Such activities can be the support of staffing (by providing capability and skill information on an employee level) or establishing a common language crossing organizational boundaries (e.g. by providing catalogues or corporate languages). Further, EAM could focus on the assessment and modeling of benefits that certain stakeholders want to achieve during the transformation. Such models could support the early identification of conflicts.

Some activities that ET management needs to conduct however, are hardly supported and will be hard to support by future EAM. These are especially related to psychological aspects like management of employee’s perception of work or further ones. Modeling such aspects might be an interesting future field for the development of new modeling languages and techniques.

The field of enterprise transformation seems to represent a significant potential for modeling research. So far, most languages focus on the different states during an ET. What is still lacking are approaches that allow for modeling the transformation itself, including factors like risks, benefits and other critical issues. When conducting a search on modeling and transformation literature on Google Scholar, most approaches found focus on transforming models by themselves. Solid work on specifics of transformations and their documentation are underrepresented. This offers an interesting and relevant field for future work.

Apart from the potential that new modeling languages or methods might have, the existing ones could be leveraged further by understanding the activities that are conducted during transformations and the capabilities that are needed. The results of the first design iterations that we present in this paper could be a helpful means for practitioners in order to identify, which activities they might be able to support with enterprise models. In consequence, the framework provides guidance where “self-marketing” activities can be applied or in which areas the internal EAM approach could become improved.

## 6 Summary & Conclusion

In the paper at hand we presented results from a research project that aims at the development of a framework for the architectural support of enterprise transformations. We focused on the presentation of the design process itself and presented some of the recently achieved results.

Of course, some limitations occur. Some colleagues might consider the lack of a one-time large-scale evaluation of the presented artifact. Nonetheless, due to the chosen one-cycle approach of design science research, the single steps were evaluated intermediately and immediately during design. Thus, the validity and reliability of the artifact is ensured. In addition, two of the research partners are planning to incorporate a version of the framework in their companies. We will report on the experiences in future work. We are aware that the details presented in the paper at hand can only be part of the whole artifact description and result. However, such reduction is necessary due to space limitations and the current state of the project. The industry mix of the research partners might also have an influence on the result. Nevertheless, not only partners that are primarily dealing with information (e.g. banks or insurances) but also more production oriented companies like utilities participated.

The next steps in the research process will be a more detailed catalogue that includes, how exactly (e.g. by which EAM artifacts) the identified ET activities can be supported. Further we aim at identifying context factors that lead to favor one introduced EAM approach over the other.

### Acknowledgement

This work has been funded by the Swiss National Science Foundation (SNSF).

## References

- [AAL12] Abraham, R.; Aier, S.; Labusch, N.: Enterprise Architecture as a Means for Coordination – An Empirical Study on Actual and Potential Practice. In: Proceedings, The 7th Mediterranean Conference on Information Systems, Guimarães 2012, S. Paper 33.
- [AAW12] Abraham, R.; Aier, S.; Winter, R.: Two Speeds of EAM – A Dynamic Capabilities Perspective. In: Proceedings, 7th Workshop, TEAR 2012, and 5th Working Conference, PRET 2012 at The Open Group Conference 2012, Barcelona 2012, S. 111–128.
- [ABA09] Asfaw, T.; Bada, A.; Allario, F.: Enablers and Challenges in Using Enterprise Architecture Concepts to Drive Transformation: Perspectives from Private Organizations and Federal Government Agencies. In: Journal Of Enterprise Architecture 5 (2009) 3, S. 18-28.
- [AG10] Aier, S.; Gleichauf, B.: Application of Enterprise Models for Engineering Enterprise Transformation. In: Enterprise Modelling And Information Systems Architectures 5 (2010) 1, S. 58-75.
- [As11] Aspara, J. et al.: Strategic management of business model transformation: lessons from Nokia. In: Management Decision 49 (2011) 4, S. 622-647.
- [Ba10] Barreto, I.: Dynamic Capabilities: A Review of Past Research and an Agenda for the Future. In: Journal Of Management 36 (2010) 1, S. 256-280.

- [BD12] Bucher, T.; Dinter, B.: Situational Method Engineering to Support Process-oriented Information Logistics: Identification of Development Situations. In: *Journal Of Database Management* 23 (2012) 1, S. 31-48.
- [BSS10] Bhattacharya, P. J.; Seddon, P. B.; Scheepers, R.: Enabling Strategic Transformations with Enterprise Systems: Beyond Operational Efficiency. In: *Proceedings, ICIS 2010* Proceedings 2010, S. Paper 55.
- [BY07] Boh, W. F.; Yellin, D.: Using Enterprise Architecture Standards in Managing Information Technology. In: *Journal Of Management Information Systems* 23 (2007) 3, S. 163-207.
- [DM12] Doolin, B.; McLeod, L.: Sociomateriality and boundary objects in information systems development. In: *European Journal Of Information Systems* 21 (2012) 5, S. 570-586.
- [FB12] Flyvbjerg, B.; Budzier, A.: Why Your IT Project May Be Riskier Than You Think. In: *Harvard Business Review* 89 (2012) 9, S. 23-25.
- [Gr06] Gregor, S.: The Nature of Theory in Information Systems. In: *MIS Quarterly* 30 (2006) 3, S. 611-642.
- [HE77] Herbert, T. T.; Estes, R. W.: Improving Executive Decisions by Formalizing Dissent: The Corporate Devil's Advocate. In: *Academy Of Management Review* 2 (1977) 4, S. 662-667.
- [HPK09] Harmsen, F.; Proper, E.; Kok, N.: Informed Governance of Enterprise Transformations. In: Proper, E.; Harmsen, F.; Dietz, J. L. G. (Hrsg.): *Advances in Enterprise Engineering II*. Springer, Berlin Heidelberg 2009, S. 155-180.
- [HTW97] Hock-Hai Teo, B.; Tan, C. Y.; Wei, K.-K.: Organizational Transformation Using Electronic Data Interchange: The Case of TradeNet in Singapore. In: *Journal Of Management Information Systems* 13 (1997) 4, S. 139-165.
- [JM00] Johnston, J.; Madura, J.: Valuing the Potential Transformation of Banks into Financial Service Conglomerates: Evidence from the Citigroup Merger. In: *The Financial Review* 35 (2000) 2, S. 17-36.
- [Ko95] Kotter, J. P.: Leading Change: Why Transformation Efforts Fail. In: *Harvard Business Review* 73 (1995) 2, S. 59-67.
- [KP11] Keller, S.; Price, C.: *Beyond performance: How great organizations build ultimate competitive advantage*. 1 Aufl., John Wiley & Sons, Hoboken, NJ 2011.
- [La09] Lankhorst, M.: *Enterprise Architecture at Work: Modelling, Communication and Analysis*. 2 Aufl., Springer, Berlin, Heidelberg 2009.
- [La13] Lankhorst, M.: *Enterprise Architecture at Work: Modelling, Communication and Analysis*. 3 Aufl., Springer, Berlin, Heidelberg 2013.
- [Mc07] McGinnis, L. F.: Enterprise modeling and enterprise transformation. In: *Information, Knowledge, Systems Management* 6 (2007) 1-2, S. 123-143.
- [OO12] Otto, B.; Oesterle, H.: Principles for Knowledge Creation in Collaborative Design Science Research. In: *Proceedings, 33. International Conference on Information Systems (ICIS)*, Orlando, CA 2012.
- [Op11] The Open Group: *TOGAF Version 9.1*. 2011.
- [Op12] The Open, G.: *ArchiMate® 2.0 Specification*. The Open Group, Berkshire, UK 2012.
- [Pe07] Peffers, K. et al.: A Design Science Research Methodology for Information Systems Research. In: *Journal Of Management Information Systems* 24 (2007) 3, S. 45-77.
- [PNL07] Pulkkinen, M.; Naumenko, A.; Luostarinen, K.: Managing information security in a business network of machinery maintenance services business - Enterprise architecture as a coordination tool. In: *Journal Of Systems And Software* 80 (2007) 10, S. 1607-1620.
- [Ra11] Radeke, F.: Toward Understanding Enterprise Architecture Management's Role in Strategic Change: Antecedents, Processes, Outcomes. In: *Proceedings, Wirtschaftsinformatik, Zuerich 2011*, S. Paper 62.

- [RB06] Rouse, W. B.; Baba, M. L.: Enterprise transformation. In: Communications Of The ACM 49 (2006) 7, S. 67-72.
- [SL99] Sarker, S.; Lee, A. S.: IT-enabled organizational transformation: a case study of BPR failure at TELECO. In: The Journal of Strategic Information Systems 8 (1999) 1, S. 83-103.
- [Se11] Sein, M. et al.: Action Design Research. In: MIS Quarterly 35 (2011) 1, S. 37-56.
- [St12] Steinhorst, M. et al.: Supporting Enterprise Transformation using a Universal Model Analysis Approach. In: Proceedings, Ecis 2012 2012, S. Paper 147.
- [Te07] Teece, D. J.: Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. In: Strategic Management Journal 28 (2007) 13, S. 1319-1350.
- [UG12] Uhl, A.; Gollenia, L. A. (Hrsg.): A Handbook of Business Transformation Management Methodology. Gower Publishing, Farnham, UK 2012.
- [WA12] Winter, R.; Albani, A.: Restructuring the Design Science Research Knowledge Base – A One-Cycle View of Design Science Research and its Consequences for Understanding Organizational Design Problems. In: Spagnoletti, P.; Baskerville, R. (Hrsg.): Designing Organizational Systems: an Interdisciplinary Discourse. 2012.
- [WF07] Winter, R.; Fischer, R.: Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. In: Journal Of Enterprise Architecture 3 (2007) 2, S. 7-18.
- [Wi12] Winter, R. et al.: Enterprise Architecture and Transformation: The Differences and the Synergy Potential of Enterprise Architecture and Business Transformation Management. In: 360° – The Business Transformation Journal (2012) 5, S. 22-31.
- [Wo13] World Café Community Foundation: The World Café. <http://www.theworldcafe.com>, Accessed 01.02.2013.

# Can boundary objects mitigate communication defects in enterprise transformation? Findings from expert interviews

Ralf Abraham<sup>1</sup>, Hella Niemietz<sup>2,3,4</sup>, Sybren de Kinderen<sup>4</sup>, Stephan Aier<sup>1</sup>

<sup>1</sup>University of St. Gallen, St. Gallen, Switzerland

<sup>2</sup>CRP Henri Tudor, Luxembourg, Luxembourg

<sup>3</sup>Radboud University Nijmegen, Nijmegen, The Netherlands

<sup>4</sup>EE-Team, Luxembourg, Luxembourg

{ ralf.abraham | stephan.aier }@unisg.ch

hella.niemietz@tudor.lu

sybren.dekinderen@gmail.com

**Abstract:** Inappropriate communication is a major threat to enterprise transformations. While enterprise architecture (EA) models may be helpful to support communication, these models are often tailored to the needs of specialists like enterprise architects. Based on empirical data from 12 expert interviews, we analyze how EA models can become boundary objects that span knowledge boundaries and alleviate communication defects among heterogeneous stakeholder groups in enterprise transformations. We contribute a framework that maps six communication defects to three knowledge boundaries and to 12 boundary object properties as a foundation for future EA model design. Our findings also indicate that EA models alone are not sufficient for overcoming communication defects, but that facilitators like architects are needed in addition.

## 1 Introduction

Induced by various environmental pressures (originating from markets, regulators, customers etc.), enterprises face a constant need for change that often affects large parts of an enterprise. This kind of large-scale change is referred to as *enterprise transformation* [Ro05]. An enterprise transformation typically is a collaborative endeavor of diverse stakeholder communities. These communities are diverse with respect to their know-ledge, values, and goals. The need for collaboration among diverse communities is well-recognized in organizational literature [Ca04; Ka01; NMS12]. To coordinate collaborative efforts during enterprise transformation, communication is a key success factor. Conversely, *communication defects* are a major threat to successful transformation. Communication defects lead to delays in transformation, increases in costs, and ultimately to struggles or even failure of transformation [FF95; HPK09].

---

This work has been funded by the Swiss National Science Foundation (SNSF) and the Fonds National de la Recherche (FNR) Luxembourg.

Enterprise Architecture (EA) models support such communication [Va11]. In line with [WF07], we understand an EA model to be a representation of an as-is or to-be state of an organization in its business to IT stack. However, we argue that not all EA models are particularly suitable for mitigating communication defects among diverse communities in enterprise transformations. Particularly, a model cannot be considered separately from its context of use: the role of the modeler (in our case: the architect) is paramount for the usefulness of EA models [SP12], while at the same time the fitness of a particular EA model depends on factors such as the addressed community or the purpose and scope of the model [An10]. Therefore, we take a boundary object perspective to analyze how EA models can help to prevent communication defects in enterprise transformations. Boundary objects, a concept from organizational science [SG89], aim at providing interfaces between different communities of practice and thus support knowledge sharing communication.

Our research question is the following: *To what extent can communication defects in enterprise transformations be mitigated by using EA models as boundary objects?*

In addressing this question, we build upon a set of boundary object properties pertaining to both (1) the model itself, and (2) the role of the architect, and discuss to what extent these properties can address communication defects in enterprise transformations. In addition to this theoretical grounding of our research, we provide empirical grounding of the proposed boundary object properties by illustrating each property with qualitative data on practical modeling experiences from senior enterprise architects. In so doing, we extend our earlier work [Ab13; NKC13] by (1) explicitly linking the role of the enterprise architect to boundary objects, (2) applying the boundary object perspective to specifically mitigate communication defects, and (3) providing an empirical grounding of the theoretically-derived boundary object properties found in [Ab13].

Note that work has been done on models that involve different stakeholders, prominently by [SP12; SPS07]. Yet, this work concerns collaborative modeling in general, while our focus is on EA models' application as far as they involve different stakeholder communities. Furthermore, work on model quality [Kr12; KSJ06; Mo98; Ne12] provides properties to assess both the process of modeling as well as the resulting model itself [Mo05]. However, in [Mo05] the focus is on evaluating the model quality for a single user community, while in the paper at hand we are interested in the interplay among different communities. The rest of this paper is structured as follows.

In section two, we discuss conceptual foundations of communication defects and boundary objects. Section three presents our research approach. In section four, we describe, based on empirical data, how various boundary objects have helped architects in overcoming communication defects. We discuss our findings, especially with regard to the role of the architect, in section five. Section six concludes our paper.

## 2 Conceptual foundations

### 2.1 Communication defects in enterprise transformations

Change-management literature shows that transformations often fail because of poor or too little communication [Ba02; Gi98; KD02; Ko95]. Elving [El05] for example points out that “poorly managed change communication results in rumors and resistance to change”. This raises a need to prevent or at least mitigate pitfalls such as communication defects among the involved communities [HPK09].

The EA function is consistently positioned as an instrument to improve communication in enterprise transformations [Ra10; Ta11]. Yet, recent research shows that communication defects also occur in EA-driven enterprise transformations. They contribute significantly to the struggling or failure of EA-driven enterprise transformations [NKC13]. Examples of such transformation struggles include delays in the transformation and not fulfilling the transformation goals. Based on qualitative data from interviews with mostly enterprise architects, [NKC13] provide a list of communication defects. They categorize those specific defects into three groups, namely *lack of communication*, *inappropriate communication* and *over-communication*. In the paper at hand we focus on the question of how EA models can be employed to overcome inappropriate communication. Inappropriate communication is found in the following communication defects [NKC13]: inappropriate communication means, inappropriate communication style, no shared frame of reference, communication against the transformation, non-aligned implicit and explicit communication, and dishonest communication.

The finding that communication defects are an important reason for struggles in or failure of EA-driven enterprise transformations is particularly interesting when considering that EA models are supposed to support communication [Va11]. This raises the question to what extent communication defects in enterprise transformations can be mitigated by using EA models, and whether existing models need to be changed.

### 2.2 Boundary objects

The differences of communities regarding their knowledge, values, and goals are manifested in knowledge boundaries. Carlile [Ca04] distinguishes three types of such boundaries: syntactic, semantic, and pragmatic boundaries. *Syntactic boundaries* exist due to different terminology among communities. *Semantic boundaries* are boundaries of interpretation, which can be crossed by identifying differences and dependencies of the communities and the creation of common meaning based on common terminology. *Pragmatic boundaries* are political boundaries. They represent differences in goals and interests. Communities have their own agenda and see their power position “at stake” [Ca04].

The three boundary types have an increasing level of complexity with syntactic boundaries at the lowest and pragmatic boundaries at the highest level. Crossing a higher-level boundary invariably involves crossing lower-level boundaries: To identify

differences in meaning at a semantic boundary a common terminology must be provided first; being able to negotiate common solutions at a pragmatic boundary also involves crossing syntactic and semantic boundaries.

To aid in crossing knowledge boundaries, boundary objects (originally introduced by Star and Griesemer [SG89]) are a widely-employed concept. Boundary objects are abstract or physical artifacts that support knowledge sharing and collaboration between different stakeholder communities by providing interfaces for communication. Examples of boundary objects include physical objects such as prototypes [Ca04], intangible objects like shared IT applications [PR04], maps and models [SG89], and abstract conceptualizations such as standardized forms and repositories [Ca04; SG89]. According to Winter and Butler [WB11], “boundary objects provide a sufficient platform for cooperative action – but they do so without requiring the individuals involved to abandon the distinctive perspectives, positions, and practices of their ‘base’ social world”. Boundary objects are emergent, and “designated boundary objects” only become “boundary-objects-in-use” when they are incorporated into the local practice of a stakeholder community [LV05]. A specific boundary object may therefore be used by two or more different communities of practice (see [Ab13] for examples).

Abraham [Ab13] identifies an initial set of boundary object properties based on a literature review. The identified properties can be classified into two groups: *object properties* that concern the construction of an object, and *management properties* that describe the way an object is used and managed in an organization. Depending on the type of knowledge boundary to be crossed, different properties may be required [Ca04].

The object properties are described as follows. For a detailed description, see [Ab13]:

**Modularity** enables communities to attend to specific areas of a boundary object independently from each other, such as attending to individual portions of an ERP system.

**Abstraction** serves the interests of all involved communities by providing a common reference point on a high level of abstraction. Local contingencies are eliminated from high-level views to highlight the commonalities.

**Concreteness** addresses specific problems relevant to specific communities. Communities are able to specify their concerns and express their knowledge related to the problem at hand. Thus, interpretive flexibility is provided.

**Shared syntax** provides a common schema of information elements, so that local use of information objects is uniform across communities.

**Malleability** entails that boundary objects are jointly transformable to support the detection of dependencies and the negotiation of solutions.

**Visualization** entails that boundary objects do not rely on verbal definitions, but possess a graphical or physical representation (e.g., a drawing or a prototype).

**Annotation** enriches boundary objects with additional information by individual communities in order to provide context for local use.

The management properties are described as follows. For more details, see [Ab13]:

**Versioning** traces changes to boundary objects, along with their rationale. Additional context is provided by reconstructing the chronological evolution of the boundary object.

**Accessibility** includes informing interested communities about the boundary object using appropriate communication channels and other measures aimed at helping communities to use the boundary object, such as trainings. As a result, the boundary object is easier to access for the involved communities.

**Up-to-dateness** includes timely communication of changes to the involved communities as well as responsibilities and processes for updating the boundary object.

**Stability** implies that the structure and underlying information objects of a boundary object remain stable over time. Despite different local uses and annotations, boundary objects provide a stable reference frame: While changes at the periphery are possible, the core of the boundary object remains stable and recognizable.

**Participation** means that relevant communities should be involved in the creation and maintenance of the boundary object, and that users should also include top management.

### 3 Research approach

To address our research question we conducted a series of semi-structured qualitative expert interviews. We chose this approach because it provides in-depth insights into complex phenomena [GG10] such as communication defects in EA-driven enterprise transformations. Semi-structured interviews are focused on the research problem while at the same time allowing for exploration of the field of research [F109]. This combination allowed us to focus on our research problem and be open to new ideas.

We used the snowball sampling technique. According to Miles and Huberman [MH94] snowball sampling is useful to identify experts that have a lot of experience concerning the phenomenon studied. In total, we interviewed twelve experts. We stopped collecting data when we did not gain any new insights into our research problem, i.e., at the point of theoretical saturation [Ei89]. Table 4 gives an overview of the interviewees.

Each interview took between 60 and 90 minutes. All the interviews were recorded and transcribed. We started the interview analysis with open coding following Flick [F109]. That is, the first codes were linked closely to the transcripts. For the specific purpose of this paper, however, we analyzed the interviews in a second step with regard to the three types of knowledge boundaries and with regard to the use of boundary objects. Furthermore, we specifically looked into the role of the enterprise architects in the context of mitigating communication defects. Eventually, this helped us to extract, combine and interpret the relevant data from the interviews.

The coded data was analyzed in two steps: (1) Based on the interviews we classified the specific communication defects presented in section 2 according to Carlile’s knowledge boundaries [Ca04]. The mapping of the six specific communication defects to the knowledge-boundary types is presented in section 4.1. (2) As a second step we analyzed the interviews with regard to the boundary object properties. We added those properties we could find support for in our interviews to the map created in step one. The mapping of the properties to the communication defects and the knowledge boundaries is described in sections 4.2 and 4.3.

Table 4. Characteristics of the experts.

| Expert   | Position              | Industry      | Works in        | Nationality     |
|----------|-----------------------|---------------|-----------------|-----------------|
| #1       | Enterprise architect  | Consulting    | Australia       | The Netherlands |
| #2       | Enterprise architect  | Energy        | The Netherlands | The Netherlands |
| #3       | Enterprise architect  | Insurance     | USA             | USA             |
| #4; #7-9 | Enterprise architect  | Consulting    | The Netherlands | The Netherlands |
| #5       | Enterprise architect  | Consulting    | USA             | USA             |
| #6       | Enterprise architect  | Public sector | The Netherlands | The Netherlands |
| #10      | Management consultant | Consulting    | The Netherlands | The Netherlands |
| #11      | Enterprise architect  | Banking       | Luxembourg      | France          |
| #12      | Enterprise architect  | Consulting    | The Netherlands | The Netherlands |

## 4 Results

### 4.1 Mapping communication defects to knowledge boundaries

In our analysis we focused on the category ‘inappropriate communication’ [NKC13], because the other two categories (‘lack of communication’ and ‘over-communication’) refer to the amount of communication, which is unlikely to be addressed by EA models and therefore is not relevant to the purpose of this paper. To answer the question to what extent EA models can be used as boundary objects to mitigate inappropriate communication, we first map the specific communication defects according to Carlile [Ca04] in syntactic, semantic and pragmatic knowledge boundaries (Table 5).

**Inappropriate communication means.** Communication means, such as face-to-face communication, newsletters or intranets, are used to transfer knowledge. However, our interview results indicate that the appropriateness of one and the same communication mean depends on the purpose of the communication and on the target group. For instance, one enterprise architect contrasted IT-oriented people and business-oriented people: While email, social-media channels or internet would work well for IT people, for business people he would rather arrange meetings where they could discuss things and have coffee together. The use of inappropriate communication means, e.g. using email threads towards business people, disturbs the transfer of knowledge and can thus be interpreted as an information-processing or *syntactic boundary*.

**Inappropriate communication style.** Different stakeholder communities have different preferences regarding the communication style. “*When you talk to architects you have to be sure that there is one mistake in the picture, so they really can find the mistake in the picture. And they are proud they found that mistake. [...] An architect thinks that’s funny. That’s the way they look at things. [...] When you are introducing a mistake in the picture when you are talking to managers, they are becoming insecure. They are not sure anymore that you really know what you are doing. [...] You have to fit your communication with the one you are talking to*” [expert #6]. This quote shows that the inclusion of a mistake in a picture has a different meaning for enterprise architects than it has for management. Therefore, we classify this defect as *semantic boundary*.

**No shared frame of reference.** [NKC13] illustrate that different stakeholder communities can have different frames of reference. They distinguish two levels of differences: the level of vocabulary and the level of understanding. If two communities differ in terms of vocabulary, the *boundary is syntactic*. If they use the same terminology but have a different understanding of it, there is a *semantic boundary* between them.

**Communication against the transformation.** Stakeholders who do not want the transformation to happen sometimes try to stop it by communicating against it. As this communication defect is based on conflicting interests and can be understood as political intervention, we classify it as a *pragmatic boundary*.

**Implicit and explicit communication not aligned.** This communication defect concerns inconsistencies between what is communicated through explicit statements and what is communicated implicitly through actions, symbols, etcetera [NKC13]. If, for example, senior management explicitly declares the increase of service quality as a main goal and at the same time introduces the reduction of costs as a new KPI, the explicit and implicit communication is not aligned. This can be interpreted as a difference in interest and therefore as a *pragmatic boundary*.

**Dishonest communication.** An example for dishonest communication that was mentioned during our interviews is not telling the negative consequences of a transformation initiative. Such a communication defect is politically motivated and can therefore be labeled as *pragmatic boundary*.

Table 5. Communication defects and knowledge boundaries.

| Communication defect                            | Syntactic boundary | Semantic boundary | Pragmatic boundary |
|-------------------------------------------------|--------------------|-------------------|--------------------|
| Inappropriate means of communication            | x                  |                   |                    |
| Inappropriate communication style               |                    | x                 |                    |
| No shared frame of reference                    | x                  | x                 |                    |
| Communication against the transformation        |                    |                   | x                  |
| Implicit and explicit communication not aligned |                    |                   | x                  |
| Dishonest communication                         |                    |                   | x                  |

## 4.2 Mapping object properties to knowledge boundaries and communication defects

After mapping specific communication defects to knowledge boundaries we analyzed the interviews regarding boundary object properties. Table 6 shows which properties were supported as being relevant for bridging certain knowledge boundaries and for mitigating specific communication defects. When assessing the mapping of properties to knowledge boundaries it must be considered that crossing knowledge boundaries is a cumulative process. To cross a pragmatic boundary for example, semantic and syntactic boundaries must be crossed before. However, to not clutter the table only the properties that have been explicitly identified as relevant to a specific defect or boundary are marked (i.e., no accumulation effects are represented in the table). In our analysis we distinguish object properties (this section) and management properties (section 4.3). The latter are marked grey in Table 6.

Table 6. Boundary object properties, knowledge boundaries and communication defects.

| Communication defect                            | Syntactic boundary | Semantic boundary                                                           | Pragmatic boundary |
|-------------------------------------------------|--------------------|-----------------------------------------------------------------------------|--------------------|
| Inappropriate means of communication            | Accessibility      |                                                                             |                    |
| Inappropriate communication style               |                    | Visualization                                                               |                    |
| No shared frame of reference                    | Shared syntax      | Modularity,<br>Abstraction/<br>Concreteness,<br>Visualization,<br>Stability |                    |
| Communication against the transformation        |                    |                                                                             | Participation      |
| Implicit and explicit communication not aligned |                    |                                                                             |                    |
| Dishonest communication                         |                    |                                                                             | Up-to-dateness     |

**Shared syntax.** If an EA model uses vocabulary that is common among different stakeholder groups in an organization, shared syntax and a common frame of reference can be achieved. This can, for example, be accomplished by agreeing on the terminology of one stakeholder community (“*you need to talk purely in business terms*” [expert #1]) and capturing this in an information model. Another possibility is to use external standards (if this is compatible with the organizational culture). One interviewee described an engineering-driven organization that was very keen on complying with ISO naming standards in their models. A communication defect resulting from a lack of a shared frame of reference due to different vocabulary can thus be solved, which means that shared syntax is primarily associated with a syntactic boundary.

**Modularity.** By providing different views of an EA model, where each view captures the concerns of a particular stakeholder community, these communities can focus on different parts of an EA model. Moreover, stakeholder communities can hide parts of an overall model they are not interested in. Since views enable each community to explicate their understanding, perspectives can be compared and differences detected. One

interviewee explained that views enabled him to communicate architecture “*in terms that can be understood by other stakeholders*” [expert #7]. A communication defect due to a lack of a shared frame of reference caused by different interpretations can be overcome by using an EA model and providing appropriate views to the involved communities. This property is therefore linked to a semantic knowledge boundary.

**Abstraction/concreteness.** Our informants stated that stakeholders needed to be able to overlook a transformation project at an early phase to “*really understand what this change means*” [expert #6]. As a specific example, the ArchiMate layer diagram was mentioned as being helpful in linking business processes at a high level to the organizational structure, and eventually to the technology layer. For an EA model as a boundary object this implies that architects should be able to provide a short and concise overview that will necessarily be at a high level of abstraction. Some informants mentioned a one-page overview, i.e., an architectural model showing the envisioned transformation in a very concise way, as an important success factor.

In an early phase of a transformation, a high level of abstraction is important for communicating the transformation goal to stakeholders. However, as a transformation goes on and concrete decision alternatives become available, more concrete models are becoming increasingly useful. Since EA models with an appropriate balance between abstraction and concreteness can help communities in quickly assessing differences in interpretation, they contribute to a shared frame of reference by providing common meanings. Therefore, abstraction/concreteness are associated with semantic boundaries.

**Visualization.** The interviewees argued for visually appealing, graphical representations of architectural models. From the majority of our interviews, it emerged that a visual representation is more highly valued than plain text. Cognitively effective models can address communication defects caused by inappropriate means of communication, as well as those caused by a lack of a shared frame of reference. Visualization is therefore particularly useful for crossing semantic boundaries. However, one interviewee reported on an organization that preferred written text over graphics. This shows that the usefulness of visualization depends to a certain degree on the communities’ preferences.

Our interviews did not support the properties of **malleability** and **annotation**.

### 4.3 Management properties and the role of the enterprise architects

We have shown that most of the object properties discussed in section 2.2 are supported by empirical data. However, we also must consider *management properties* of boundary objects (section 2.2). The idea of management properties is supported by our interviewees who point out that, next to the design of appropriate EA models, their management by the enterprise architects is particularly important. In the interviews we found evidence for four of the five management properties.

**Stability/up-to-dateness.** According to the experts an enterprise architect has to deal with the trade-off between stability and up-to-dateness of an EA model. They indicated that the more unstable a model was, i.e., the higher the update frequency, the more they

had to invest in communication to maintain a shared understanding among the stakeholders. Conversely, a high degree of stability of a model can lead to a structure that is well-recognized among diverse communities, and thus contributes to establishing a common meaning. Therefore, stability contributes to crossing semantic boundaries. However, EA models also have to be updated regularly *“because the world is changing, your enterprise is changing, people are changing”* [expert #9]. If an enterprise architect does not update an EA model although s/he knows that something relevant has changed, this can be interpreted as dishonest communication. Hence, up-to-dateness can help in crossing a pragmatic boundary.

**Accessibility.** According to our interviews making an EA model accessible does not only include putting it on a server that everyone can access. Depending on their preferences regarding the means of communication people would or would not access that server (section 4.1). One architect illustrated his way to make his one-page overview of a future architecture accessible: *“So using that paper everywhere you go and put it on the table, leave it on the walls is helping you because everyone knows where his part is on the project”* [expert #8]. This quote shows that accessibility can also be reached by personal interaction. By ensuring that everyone receives the information accessibility helps crossing syntactic boundaries. Moreover, if an EA model is made accessible by personal interaction of the architect, this can also help crossing semantic boundaries because the architect can directly explain or translate the model to the respective audience: *“I was naïve enough in the beginning to think that if it was written down and if you have models then people would understand what was written down and they would understand those models. And then everybody would be on the same page. But that turns out not to be the case. So it’s a lot of personal interaction with people”* [expert #1].

**Participation.** The interviewees mentioned participation as an essential condition for the acceptance of an EA model. *“People need to know about it, but also need to have the feeling that their voice is heard and that their concerns are included in the architecture”* [expert #7]. If a stakeholder has participated in creating an EA model for the transformation, s/he is more likely to support the transformation. Thus, the probability for communication against the transformation is decreased. Participation in the process of creating and maintaining an EA model of a future architecture by a broad number of stakeholders can therefore help in bridging pragmatic boundaries. However, developing common interests among different stakeholder communities requires strong personal involvement of the architects. Virtually all of our interviewees stressed the importance of conducting workshops in order to engage stakeholders instead of preparing a target architecture themselves and presenting the finished result.

However, the experts also pointed out that, if the architect does not intervene, letting different stakeholders participate in creating an EA model and conducting joint workshops often results in communication defects. The interviewees emphasized the architects’ role as a translator or mediator between the different groups: *“And I invite people to do not judge or make assumptions but listen what other people have to say before they reply. So don’t have their answers ready and just wait for the right keyword to jump in. Let other people speak. But I will always take the role of summarizer only. So whenever someone*

*has spoken I will then give a summary. [...] So I will repeat what was said and use different words so that the different people will be able to pick it up” [expert #1].*

**Versioning.** We did not find support for this property in our interviews.

We have illustrated that solely having EA models whose construction adheres to certain properties may not be sufficient to mitigate certain communication defects. The examples above show that to implement the management properties for EA models, the architect’s personal interaction is needed. They have to make sure their models are recognized and understood by the relevant stakeholders. Furthermore, architects act as translators between different stakeholder communities. The property ‘participation’ expresses the need for also engaging stakeholders when creating EA models. In summary, the involvement of enterprise architects is needed to cross all three boundary types (Table 63).

Furthermore, in the interviews, no management property was directly linked to the defect ‘explicit and implicit communication not aligned’. However, we argue that the combination of stakeholder engagement and personal interaction of the architects may also help to mitigate this defect. Through participation stakeholders express and document their concerns explicitly. If later the implicit communication contradicts those explicitly stated goals, the architects can point that out through personal interaction.

Finally, the interview analysis shows that pragmatic boundaries cannot be bridged by EA models and architects alone. However, an EA model that has invited participation from many stakeholders may help convey a de-politicized perspective that is considered more objective. When an EA model is recognized among stakeholder communities as providing a neutral, commonly agreed-upon perspective (instead of being perceived as disproportionately representing a single community’s perspective), it will be harder (albeit not impossible) for politically-motivated stakeholders to push their own agenda and communicate against the transformation.

## 5 Discussion

Our data suggest that EA models as boundary objects are particularly helpful for crossing semantic boundaries, especially when helping to identify differences in interpretation and thereby preventing communication defects caused by the lack of a shared frame of reference. For addressing semantic boundaries, visualization, abstraction, and modularity emerged as the most important properties of EA models. These properties help different stakeholder communities to identify misunderstandings and re-align their interpretations towards a shared understanding. To maintain shared understanding and a common frame of reference, mock-ups of user-interface screens or roadmaps have been named by interviewees as valuable tools supporting EA models. Particularly roadmaps received attention for highlighting dependencies between the actions of various stakeholder communities, and also for showing individual stakeholders when and to what degree their areas of concern will be affected by a transformation. However, also at a semantic bound-

dary, the involvement of the architect is required to ensure that these boundary objects are actually used and that stakeholder concerns are adequately reflected.

Overall, the harder a communication defect is to fix, the more involvement of the architect is required: While an established and well-communicated information model, crossing a syntactic boundary, functions largely without the architect's intervention, overcoming pragmatic boundaries requires strong personal and time-consuming involvement, e.g., by conducting face-to-face meetings, negotiations, and workshops.

To cross pragmatic boundaries and address communication defects such as dishonest communication, or communication against the transformation, requires both EA models that are trusted for the accuracy of their content, and skilled architects that are trusted for the way they are managing and using these models. Especially when pragmatic boundaries are encountered, the architect needs to perform a role as a 'boundary spanner', i.e., s/he needs to invest significant effort in upholding communication between diverse stakeholder communities.

Levina and Vaast [LV05] identified three conditions effective boundary spanners need to meet: (1) Boundary spanners need to have sufficient knowledge and understanding of each of the fields they are about to span to be perceived as legitimate and competent. For architects, this means that they need to have an understanding of both business and technological terms to be respected as competent by both business-unit and IT-unit members and to be able to fulfill their translator role. (2) Boundary spanners need to be considered legitimate negotiators of their own field. For architects, this means that they need to be perceived as authorized to make architectural decisions, to give dependable advice. An example would be granting exceptions from architectural principles. (3) Boundary spanners need to possess the required communication and negotiation skills associated with this role. And they need to be willing to perform a boundary spanning role between two different fields instead of becoming functional experts in one field alone. For architects, this brings the risk that they are considered as being caught in the middle between the business and IT domain, without being seen as legitimate and competent members of either one. To mitigate this risk and support the boundary-spanning function, skilled architects need to be able to rely on the proper tools, one of which are boundary objects.

## 6 Conclusion

In this paper, we have analyzed the knowledge boundaries lying beneath various communication defects. In a second step we have shown how EA models acting as boundary objects can be used to overcome these communication defects. We conclude that EA models are most useful for crossing semantic boundaries when they possess the object properties abstraction, modularity, and visualization, whereas for communication defects residing at pragmatic boundaries enterprise architects are required to perform a boundary-spanning function.

Although this inquiry builds on both theoretical considerations and empirical data, the findings are still preliminary. Further in-depth studies are necessary to understand the

construction of those (few) EA models that work as boundary objects, and the context of their use. Also, the EA function is implemented differently across organizations, e.g., with respect to governance structures. Thus, the architect's role, organizational position, tasks and responsibilities, as well as his/her interface with specific stakeholder communities (e.g., project managers or HR managers) needs to be further analyzed. Finally, more work needs to be done regarding the theoretical framing of communication (defects) in EA-driven enterprise transformations. This will help in classifying specific situations as being communication defects or not and thus in identifying those situations where EA models are likely to work as boundary objects.

Summarizing, this inquiry is a step to eventually derive design principles for EA model construction and management that are particularly suitable for overcoming communication defects.

## References

- [Ab13] Abraham, R.: Enterprise Architecture Artifacts as Boundary Objects – A Framework of Properties. In: Proceedings, 21th European Conference on Information Systems, Utrecht, NL 2013.
- [An10] Anaby-Tavor, A. et al.: Insights into enterprise conceptual modeling. In: *Data & Knowledge Engineering* 69 (12), 2010, pp. 1302-1318.
- [Ba02] Barrett, D. J.: Change communication: using strategic employee communication to facilitate major change. In: *Corporate Communications: An International Journal* 7 (4), 2002, pp. 219-231.
- [Ca04] Carlile, P. R.: Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. In: *Organization Science* 15 (5), 2004, pp. 555-568.
- [Ei89] Eisenhardt, K. M.: Building Theories from Case Study Research. In: *Academy of Management Review* 14 (4), 1989, pp. 532-550.
- [El05] Elving, W. J. L.: The role of communication in organisational change. In: *Corporate Communications: An International Journal* 10 (2), 2005, pp. 129-138.
- [FF95] Ford, J. D.; Ford, L. W.: The role of conversations in producing intentional change in organizations. In: *Academy of Management Review* 20 (3), 1995, pp. 541-570.
- [Fl09] Flick, U.: *An Introduction to Qualitative Research*. SAGE Publications, Los Angeles et al. 2009.
- [GG10] Ghauri, P.; Grønhaug, K.: *Research Methods In Business Studies: A Practical Guide*. 2010.
- [Gi98] Gilsdorf, J. W.: Organizational Rules on Communicating: How Employees Are - and Are Not- Learning the Ropes. In: *Journal of Business Communication* 35 (2), 1998, pp. 173-201.
- [HPK09] Harmsen, F.; Proper, E.; Kok, N.: Informed Governance of Enterprise Transformations. In: Proper, E.; Harmsen, F.; Dietz, J. L. G. (eds.): *Advances in Enterprise Engineering II*. Springer, Berlin Heidelberg 2009, pp. 155-180.
- [Ka01] Karsten, H. et al.: Crossing boundaries and conscripting participation: representing and integrating knowledge in a paper machinery project. In: *European Journal of Information Systems* 10 (2), 2001, pp. 89-98.
- [KD02] Kitchen, P. J.; Daly, F.: Internal communication during change management. In: *Corporate Communications: An International Journal* 7 (1), 2002, pp. 46-53.

- [Ko95] Kotter, J. P.: Leading Change: Why Transformation Efforts Fail. In: Harvard Business Review 73 (2), 1995, pp. 59-67.
- [Kr12] Krogstie, J.: Quality of Business Process Models. In: Sandkuhl, K.; Seigerroth, U.; Stirna, J. (eds.): The Practice of Enterprise Modeling. Springer Berlin Heidelberg 2012, pp. 76-90.
- [KSJ06] Krogstie, J.; Sindre, G.; Jørgensen, H.: Process models representing knowledge for action: a revised quality framework. In: European Journal of Information Systems 15 (1), 2006, pp. 91-102.
- [LV05] Levina, N.; Vaast, E.: The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. In: MIS Quarterly 29 (2), 2005, pp. 335-363.
- [MH94] Miles, M. B.; Huberman, A. M.: Qualitative Data Analysis: An Expanded Sourcebook. 2nd ed., Sage Publications Inc 1994.
- [Mo05] Moody, D. L.: Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. In: Data & Knowledge Engineering 55 (3), 2005, pp. 243-276.
- [Mo98] Moody, D. L.: Metrics for Evaluating the Quality of Entity Relationship Models. In: Proceedings, 17th International Conference on Conceptual Modeling - ER '98, Singapore 1998, pp. 211-225.
- [Ne12] Nelson, H. J. et al.: A conceptual modeling quality framework. In: Software Quality Journal 20 (1), 2012, pp. 201-228.
- [NKC13] Niemietz, H.; de Kinderen, S.; Constantinidis, C.: Understanding the role of subcultures in the enterprise architecture process. In: Proceedings, 21th European Conference on Information Systems, Utrecht, NL 2013.
- [NMS12] Nicolini, D.; Mengis, J.; Swan, J.: Understanding the Role of Objects in Cross-Disciplinary Collaboration. In: Organization Science 23 (3), 2012, pp. 612-629.
- [PR04] Pawlowski, S. D.; Robey, D.: Bridging User Organizations: Knowledge Brokering and the Work of Information Technology Professionals. In: MIS Quarterly 28 (4), 2004, pp. 645-672.
- [Ra10] van der Raadt, B. et al.: The relation between EA effectiveness and stakeholder satisfaction. In: The Journal of Systems and Software 83, 2010, pp. 1954-1969.
- [Ro05] Rouse, W. B.: A Theory of Enterprise Transformation. In: Systems Engineering 8 (4), 2005, pp. 279-295.
- [SG89] Star, S. L.; Griesemer, J. R.: Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. In: Social Studies of Science 19 (3), 1989, pp. 387-420.
- [SP12] Stirna, J.; Persson, A.: Purpose Driven Competency Planning for Enterprise Modeling Projects. In: Ralyté, J. et al. (eds.): Advanced Information Systems Engineering. Springer Berlin Heidelberg 2012, pp. 662-677.
- [SPS07] Stirna, J.; Persson, A.; Sandkuhl, K.: Participative Enterprise Modeling: Experiences and Recommendations. In: Krogstie, J.; Opdahl, A.; Sindre, G. (eds.): Advanced Information Systems Engineering. Springer Berlin Heidelberg 2007, pp. 546-560.
- [Ta11] Tamm, T. et al.: How Does Enterprise Architecture Add Value to Organisations? In: Communications Of The Association For Information Systems 28 (1), 2011, pp. 141-168.
- [Va11] Valorinta, M.: IT alignment and the boundaries of the IT function. In: Journal Of Information Technology 26 (1), 2011, pp. 46-59.
- [WB11] Winter, S. J.; Butler, B. S.: Creating bigger problems: grand challenges as boundary objects and the legitimacy of the information systems field. In: Journal Of Information Technology 26 (2), 2011, pp. 99-108.
- [WF07] Winter, R.; Fischer, R.: Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. In: Journal Of Enterprise Architecture 3 (2), 2007, pp. 7-18.

# A Hybrid OLAP & OLTP Architecture Using Non-Relational Data Components

Thomas Benker

Chair of Information Systems - Systems Engineering  
University of Bamberg  
An der Weberei 5  
96047 Bamberg, Germany  
thomas.benker@uni-bamberg.de

**Abstract:** Relational database systems are still the first choice for the realization of business application systems. They are used as operational databases to persist business process data or as the basis for data warehouse systems. In recent years, the NoSQL-movement proposed a number of new non-relational data stores because of changing requirements of Web 2.0 applications and related limitations of relational systems. We argue that these new systems can also be used in OLTP- and right-time OLAP-scenarios. We demonstrate how enterprise application systems can be realized based on concepts of the NoSQL-movement. We propose an architecture that enables the deployment of an appropriate data model to a certain business domain and the integration of OLTP- and OLAP-functionality. Components of the architecture are suited to enhance business data with analysis-relevant information and to perform analysis tasks. To demonstrate the application of the architecture and its analytic features, we introduce a short case study.

## 1 Introduction

Since the publication of the relational model of Codd [Cod70] in 1970 most database systems were built upon that theoretical foundation. These systems were the first choice for realizing the persistence layer of business application systems for at least 20 years. Within the last few years, a new generation of non-relational data stores, known as NoSQL (often interpreted as “Not only SQL”) emerged. The reasons for their development are changing requirements, e.g. big data, scalability and fewer schema restrictions in Web 2.0 applications. To fulfill these requirements, they offer different system characteristics compared to traditional relational database systems. NoSQL data stores use non-relational data models and can be categorized as document stores, column-oriented data stores, graph data stores or key/value-stores [Ind12]. They are mostly designed for high availability and partition tolerance [Ind12]. This means, according to Brewer's CAP-Theorem [Bre00], that they are restricted for strong consistency, a key feature of relational database and enterprise application systems. The CAP-Theorem states that just two of the consistency, availability and partition tolerance requirements can be fully satisfied [GL02]. However, we claim that non-relational data stores can also be used in enterprise application systems, despite their specific characteristics. Instead, we argue that alternative data models and varying system characteristics of these data stores offer further benefits. But currently, there are only

few contributions considering these topics (see section 2). For that, we investigate how these non-relational technologies can be used in combined OLTP- (Online Transaction Processing) and OLAP-scenarios (Online Analytical Processing) and contribute to reducing the mismatch between application domains and persistent data models.

Today, object-oriented programming languages are dominant for the realization of business application systems. But storing objects in a relational manner often means breaking up complex hierarchical domain structures into a normalized schema. This phenomenon is known as object-relational impedance mismatch. Currently, this is often treated by an additional mapping layer between application and persistence logic. But these mappers introduce additional complexity in development and programming logic [Gho10]. Further, the relational model is a general data model and not aligned to a certain application domain and its characteristics; e.g. social networks, route planning problems, and business entities, as invoices or orders, are quite distant from relational concepts. Storing them in a normalized schema means violating their natural structures. This manifests a further mismatch with additional complexity in data usage [Gho10]. The authors of [SF12] propose to choose an appropriate data model/store fitting the requirements of each particular domain model. This concept is called polyglot persistence. It implies a movement from integration to application databases. Integration databases are designed to fulfill requirements of a number of application systems. Application databases can be designed and modified just according to the requirements of a single application [SF12]. Those requirements can be technical, concerning availability, consistency, backup strategy, scalability; or they can be structural, concerning data structures (structured vs. semi-structured data, highly linked entities vs. coherent hierarchical encapsulated data). The NoSQL-movement offers a number of alternatives when choosing the appropriate data model/store for a certain business domain.

Enterprise application systems can be categorized as OLTP- and OLAP-systems. OLTP-systems are used to realize the functionality of operational business processes. OLAP-systems perform business analysis and deliver decision-relevant information. Data extracted from operational data structures offers a common and valuable input for those systems. Today, enterprises have to be flexible in order to detect and react to certain operational business events with low latency [ACN05]. In [Hac04] this latency is split up in *capture latency*, describing the time between the occurrence of an event and its preparedness for analysis, *analysis latency*, meaning the interval until the information is delivered, and *decision latency* until an action is taken. The authors of [Rus11] examined the relevance of a concept called operational data warehousing in praxis. Key features are the bi-directional integration of operational business processes with data warehouse capabilities, in order to react to certain business events within an appropriate latency. One result is that almost of two-thirds of the participating organizations practice some form of operational data warehousing, and most of the others are planning to start related projects.

For this, one contribution of that paper is the proposal of a software architecture that enables the integration of OLTP- and operational OLAP-capabilities for right-time decision making. The term “right-time” is used to state, that whenever a decision maker is doing some analyses, these analyses are based on up-to-date data. The second contribution is the identification of concepts for the application of non-relational technologies (NoSQL) in enterprise application systems in order to realize benefits of

polyglot persistence. This work manifests a first step in research. The architectural blueprint should be used as basis to discuss and motivate further research with the objective to refine the presented concepts and to enhance applicability in integrated OLTP- and OLAP-scenarios. For that first step the proposed architecture focuses on a single business process scenario.

With respect to the mentioned contributions, the paper is structured as follows: In section 2 we investigate the state-of-the-art concerning component architectures to integrate OLTP- and OLAP-scenarios and using non-relational data stores. In section 3, we present the blueprint of the hybrid architecture using three types of components to integrate OLTP- and OLAP-scenarios. Section 4 presents the operational, monitoring and analytic data components of the architecture in more detail. In section 5 we present a use case to demonstrate the application of the architectural concept and its analytic capabilities. Section 6 closes the paper with a discussion of the presented architectural blueprint and proposal for future research projects.

## 2 State-of-the-Art

This section reflects the literature related to the presented work. Few publications discuss the usage of non-relational technologies in order to enable polyglot persistence or to integrate OLAP- and OLTP-scenarios. Component or service-oriented software architectures using non-relational technologies or integrating OLAP and OLTP are not discussed in literature.

In [Gho10] the author presents an infrastructural concept for multiparadigm data storage in enterprise applications. NoSQL data stores are used as cache systems to bridge the gap between the domain model of the application system and relational back-end data storage. The relational database system is used as a centralized data store to integrate data from heterogeneous domain models and application systems. The author also proposes using the relational structures for the generation of reports by using SQL or other relational tasks. The synchronization between cache and back-end databases is done in an asynchronous way. So, the infrastructure is only suitable for systems with eventually consistent characteristics. This means that there may be a short time span in which non-relational cache and relational back-end systems are not consistent. This work focuses solely on the integration of heterogeneous NoSQL data stores in enterprise scenarios.

In [SF12] the authors discuss a wide range of aspects for using NoSQL data stores in enterprise application systems. Among others, they make proposals concerning polyglot persistence, designing non-relational data structures and analyzing them using Map/Reduce [DG08]. They propose pooling data in a common entity type that is used together by the application. Further, they suggest defining services to manage each data type. Each service should encapsulate a data store best fitting the technical and structural requirements of the managed data structures. The results are used as input to the presented work and are integrated in an architectural concept for hybrid OLTP and OLAP usage.

Additionally, some research groups work on a more technical level, e.g. published in [KN11], [KHB<sup>+</sup>11] or [Pla09]. Their goal is to integrate OLAP and OLTP capabilities based on a single database system with a non-relational data model. The development is

mainly driven by the technical evolution of the last few years. The characteristics of these database systems are in-memory technologies and/or column-oriented data models. The concept of polyglot persistence in enterprise application architectures is not covered.

### 3 The Hybrid Component Architecture

The term “software architecture” is widely understood as a set of system components and their relations in order to realize a certain functionality of a software system. Nowadays, a huge number of business application systems are realized as service-oriented architectures. For the presented work we use the term *component architecture*. A service is characterized as an interface offering certain functionality while hiding the realization. In this and the future work, the realization of such an interface is of huge relevance and the inner structure of a component has to be discussed; for that we use the term “component”. Furthermore, a component is an independent part of a software system offering a dedicated functionality by its interface. It can be composed with and reused by other components in order to deliver higher functionality [Som11]. With respect to that definition and to reduce the mismatch of data and domain model, as well as integrating OLAP and OLTP in order to reduce latency between recognition of and reaction on business events, the requirements of the architecture can be summarized as follows:

support of polyglot persistence by keeping the business process execution (workflow) independent of a concrete data store/model;

enable the integration of reusable data components in higher workflows while hiding their implementation details behind an interface;

enabling the integration of OLTP- and OLAP-functionality in order to deliver up-to-date data for right-time decision situations;

In [SF12] the authors motivate to encapsulate heterogeneous database systems behind service interfaces to foster the reuse of data and to enhance the independence of application systems from specific data models. Based on this idea, we propose a component architecture in order to realize polyglot persistence, as well as right-time business analysis, in the context of a single business process. This paper focuses explicitly on data components supporting business processes, tracking monitoring data at run-time and business analysis. In figure 1 a blueprint of the proposed architecture is provided.

During *business process execution* by a workflow system, several *operational data components* are orchestrated to fulfill a common business functionality. The purpose of an operational data component is to manage the persistent business data. To integrate these data in a higher process, an operational component offers an OLTP-interface mainly supporting CRUD-operations (create, read, update, delete) or some advanced business logic on its managed business entities. The offered operations are implemented on an appropriate non-relational data model, fitting the technical and structural requirements of the business entity and its domain. Therefore, the mismatch between domain and data model can be reduced. To be orchestrated by a workflow, operational

data components also have to deal with the heterogeneity of data models and be prepared to be integrated in a transaction context.

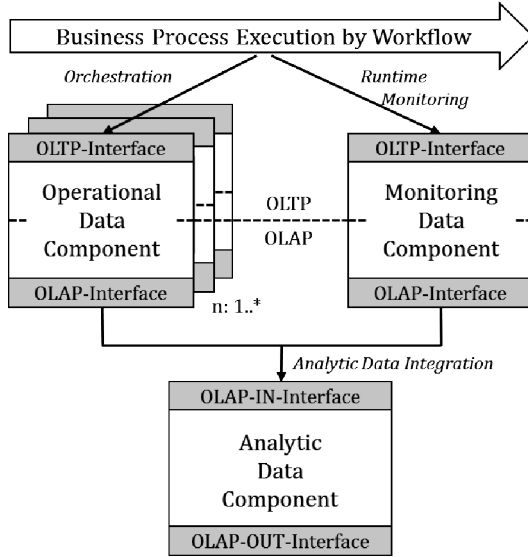


Figure 1: The blueprint of the hybrid OLAP & OLTP architecture.

In order to integrate OLTP- and OLAP-scenarios, each component has to support OLAP-functionality, too. To be flexible in data analysis, data warehouse systems use the multidimensional data model of Codd et al. [CCS93]. To provide analytic capabilities, quantitative measures are analyzed and aggregated following qualitative dimensions and their levels. The multidimensional model is also used for analysis in the proposed architecture. Each operational data component can provide measures and dimensional information for its managed business entities. To provide a complete analytic view of the business process, the analytic information of all operational components has to be integrated. This is done by the *analytic integration process*. The data integration is done in two steps: The first step is internal to the operational data component. It involves extracting the analysis-relevant data and enriching it with multidimensional information, e.g. dimensions and/or pre-calculated measures. Operational data components are able to deliver analytic data without negatively influencing the OLTP-functions. We argue this is due to typical system characteristics as scalability, parallel processing and new concurrency concepts (e.g. Multiversion Concurrency Control) of NoSQL-systems. In the second step, external to the operational components, the data is integrated by the application integration process, triggered by certain events. This process calls the OLAP-interface of the operational data components to get analysis-relevant information, integrates this information and hands it to an analytic data component using the OLAP-IN-interface. Within a single business process, data is assumed to be consolidated and homogeneous. This means it can be integrated with low complexity by using the analytic integration process. The analytic data component is responsible for the realization of

analytic functions. In order to offer these functions and to deliver the results, it is prepared with an OLAP-OUT-interface.

*Monitoring data components* are introduced to the architecture in order to enhance the analytic capabilities with run-time information. Monitoring data components are also divided in OLTP- and OLAP-functionality. Their purpose can be described as tracking run-time data from the business process executions and enriching it with analytic information.

To be flexible in delivering the right information at the right time when it is required for analysis, we propose four types of events. They are specified to trigger the analytic data integration process: (1) *Temporal events* are formulated as time periods or as certain points in time (e.g. every 10 hours certain process data has to be analyzed). (2) *Business events* show that a certain state of the business process has been reached. An incoming order or payment are examples of business events. (3) *Run-time events* represent more technical events in workflow instances executing the business process, e.g. a process instance is idle for a certain time or a run-time exception occurred. (4) *Consumer events* are related to the analytic data component. They represent ad-hoc consumer requests using the OLAP-OUT-interface.

## 4 Operational, Monitoring and Analytic Data Components

This section is used to discuss the data components in detail. Operational data components are orchestrated by workflows and manage the business process data while monitoring data are used to keep track of workflow executions. Analytic data components perform the analysis on the integrated business and logging data.

### 4.1 The Concept of Operational Data Components

The blueprint of an operational data component is shown in figure 2. There are four main parts to be examined in detail: the OLTP-part, realizing the business logic; the OLAP-part, realizing the analytic data enrichment; the interface to configure the component and the non-relational data store at its core.

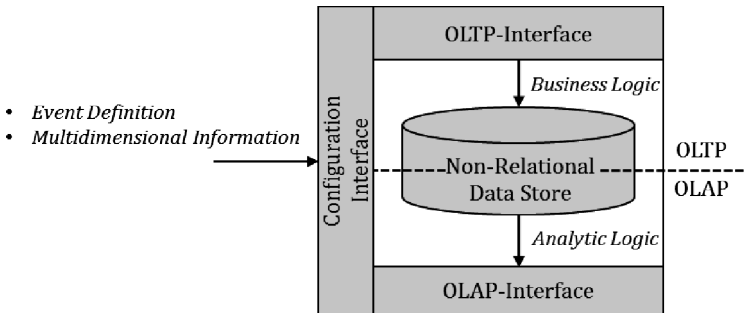


Figure 2: The blueprint of an operational data component.

According to [SF12], a common feature of most non-relational data stores is what they call the aggregate data model. This means that data which is handled (read, written)

together by the application logic is persisted in a single denormalized entity type. We extend this definition to the business context by defining business entities, as they occur in enterprise reality, like order, bill or customer as aggregate data. For each aggregate business entity we suggest using one managing component. It is responsible for offering a manageable set of operations on coherent data without the necessity of calling other components.

The *OLTP-interface* of an operational component defines operations which can be used by workflows in order to realize a certain composed business functionality. This functionality is often realized by the means of transactions. The overall result of a transaction is a successful ending or an abort with rollback to the initial state. All components participating in a transaction have to be consistent with that result. For that it must be possible to integrate an operational component in the transaction context of the orchestrating workflow. This is possible by providing an implementation of the Do-Undo-Redo-protocol as proposed in [GR93]. In case of a failing overall transaction, the local results of an operational data component can be undone. The component's internal consistence can be realized by the aggregate data design. Most non-relational data stores do not offer ACID-transactions (atomicity, consistency, isolation, durability) on more than one item but modify a single aggregate with transactional characteristics [SF12]. In order to manage the heterogeneity between data components, a common data format has to be used. JSON (Java Script Object Notations, [Cro06]) is a data format supported by most of the non-relational data store APIs. Workflow-systems have to be evaluated for their JSON support in order to use that format for persistence and data exchange. Doing so allows programming complexity within the components to be reduced.

The *OLAP-interface* has to offer operations in order to provide data for the analytic integration process. The analytic part of the component has to extract relevant data from the non-relational data store and enhance it with additional multidimensional information, e.g. the aggregation hierarchy of a dimension or pre-calculated measures. Each data component has to be configured in order to be able to prepare its business entities with analytic information. For this reason each data component is designed to have a *configuration interface* where this information can be specified. Two general possibilities are imaginable for this purpose. First, a data component can be parameterized with the location of a repository or a service where additional multidimensional information can be requested at run-time. Second, the multidimensional information is specified directly on each data component before business process executions. Which alternative fits best and the design of a concrete solution are tasks for future work. A further feature of the configuration interface is the ability to define the events (temporal, business, run-time) that trigger the analytic integration process. By using these events, the architecture can be configured to react to certain business events and subsequently reduce analysis and capture latency.

The central part of the component is a *non-relational data store*. To reduce the impedance mismatch and to benefit from polyglot persistence, an appropriate data model has to be chosen. Non-relational candidates are document stores, column-oriented data stores, graph data stores or key/value-stores. Which one to choose depends on technical and structural requirements of the domain model and on the features supported by a concrete data store. In [Hec11] and [Ind12] the authors characterize NoSQL-systems according to their data model and technical characteristics. These publications can be used to support the decision for a dedicated non-relational data store.

## 4.2 The Concept of Monitoring Data Components

The concept of monitoring data components is quite similar to operational data components. As mentioned earlier, their purpose is to log run-time data of the execution of workflow instances. Examples are entering a certain part of the workflow, calling a certain data component, starting, committing or aborting a transaction, ending a workflow successfully or in failure state. Monitoring components are part of the architecture in order to add further analytic capabilities and to react to run-time events in an adequate manner. The OLTP-interface is different compared with operational components and offers operations to deliver logging information of the execution environment. Monitoring data components also provide a configuration interface in order to define multidimensional information and to specify run-time events. The OLAP-interface is used to add analytically enriched monitoring data to the business data by the analytic integration process.

## 4.3 The Concept of Analytic Data Components

All analysis-relevant information (measures and dimensions) are integrated by the analytic integration process to a single data set, e.g. to a JSON-document, and given to the *OLAP-IN-interface* of the analytic data component. The blueprint of the analytic data component is visualized in figure 3. The purpose of that component is to perform the analysis based on non-relational data in a NoSQL data store. Using a relational database system would introduce further complexity, e.g. for transforming non-relational business data to relational analytic data.

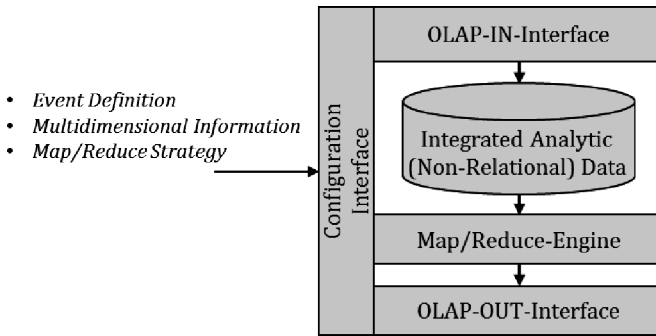


Figure 3: The blueprint of an analytic data component.

The integrated data is stored in the *non-relational data store* of the component and fits the requirements of the used data processing technology – Map/Reduce. Similar to relational OLAP, where data is structured as star or snowflake schema with respect to efficient SQL-queries, the non-relational data is prepared to be evaluated by the *Map/Reduce-engine*. Map/Reduce as proposed in [DG08] is a paradigm for parallel processing of huge data sets. It is often used in context of NoSQL data stores because it fits the characteristics of the non-relational data models. Additionally, the analysis latency can be reduced by easily increasing the number of nodes performing the

Map/Reduce-task. Map/Reduce offers two functions: first, incoming data sets are *mapped* to key-value pairs and second, the values of equivalent keys are aggregated and *reduced* to a single value [SF12]. The analytic data component has to offer analytic capabilities based on business process data. We propose how to use Map/Reduce for this purpose and show, thereby, that it is possible to perform that kind of analysis in the proposed architecture. As mentioned earlier, the analytic data consists of measures and dimensions, as known from multidimensional data structures. To describe the map-function, we extend the proposal in [SF12]. We suggest building up the keys with extensions of the levels of the associated dimensions and separating them by “.”. So, a key represents a possible extension of the dimension levels, one from each dimension. The values in those key/value-pairs are the measures which remain untouched during the execution of a map-function. As a result, there is a key/value-pair for each possible dimension level combination. The reduce-function uses the output of the map-function as input and applies the aggregate function to the measures of the key/value-pairs which share the same key. A practical example is given in section 5.2. It is not necessary to apply the map- and reduce-functions to all analytic data sets each time the analytic data component is updated. There are different strategies to apply the map/reduce-pattern, for example temporal decoupling of map and reduce or an incremental application only to updated data [SF12]. These strategies can be defined using the *configuration interface* and contribute to reducing the analysis latency. The configuration interface can also be used to define events in order to trigger further analysis (e.g. on new incoming data or on certain analytic results) or to send a message to certain observers. Using the *OLAP-OUT-Interface*, observers can register to be informed in case of certain results. Furthermore, the OLAP-OUT-interface can be used to trigger ad-hoc analysis functions (defined as consumer event).

## 5 Case Study

We present a case study in order to demonstrate the application of the hybrid OLAP and OLTP component architecture. We show how it can be used in order to manage and to analyze business process data. For the sake of confirmability, the case study is kept simple and focuses on the introduced concepts of the architectural blueprint.

### 5.1 Introduction of the Business Process “Procurement of Flights”

Figure 4 shows in a schematic manner the interactions within a business process for the procurement of flights from airlines to customers by using a web portal. The case focuses on the interaction between customer and flight portal, which should be automated using the presented hybrid OLAP and OLTP architecture. To procure a flight, there are two main interaction activities between portal and customer.

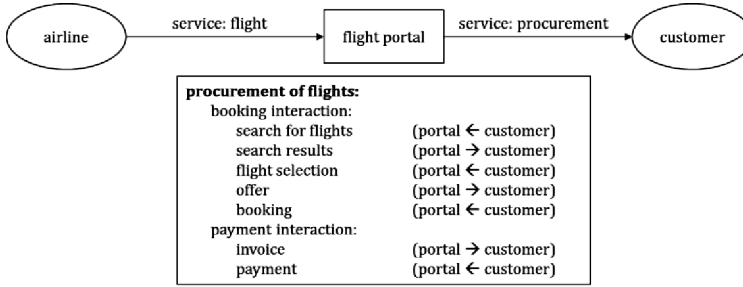


Figure 4: Interaction of the case study “procurement of flights”.

First, to book a flight, the customer can search for flights, select one, and after getting a flight offered, book this flight. The second interaction describes the payment. The flight portal sends an invoice to the customer who reacts with the payment. The interaction between flight portal and airline is mainly used for synchronizing data. After a customer has chosen a certain flight, the flight portal requests up-to-date flight information. After the completion of the booking by the customer, the flight portal and the airline synchronize the booking and customer information.

## 5.2 The Hybrid Component Architecture for the Case Study

Figure 5 shows the proposed architecture to realize the business process “procurement of flights.” It is necessary to manage information on flights, the airlines and the customers. We propose one operational component for each of these master data entity types. We propose two further components, one to manage the booking and one to manage the payment information. We argue that data that is treated together by the application workflow should be persisted and managed by a dedicated operational data component. The workflow can be realized using an executable language, e.g. BPMN (Business Process Model and Notation, [OMG11]) and orchestrates the data components by using a REST-based (Representational State Transfer, [Fie00]) OLTP-interface and JSON-documents. Based on this, a document store fits best for persistence. As a result, business entities are managed according to the structures of the domain model and close to the data usage of the orchestrating workflow. The monitoring data component is realized on a key/value-store. It fits best for logging monitoring information. Figure 5 also shows examples of events, defined on operational data components in order to trigger the analytic data integration process (realized as BPMN-workflow). If a booking or a payment is completed, the concerning data components start to create a JSON-document with the measures (volume of the booking, the days sales outstanding of the payment) and hand it to the BPMN-process. During its execution it requests further analytic data from the other data components, e.g. the dimension geography of the customer from the customer component in order to enable aggregate functions on city, state and country.

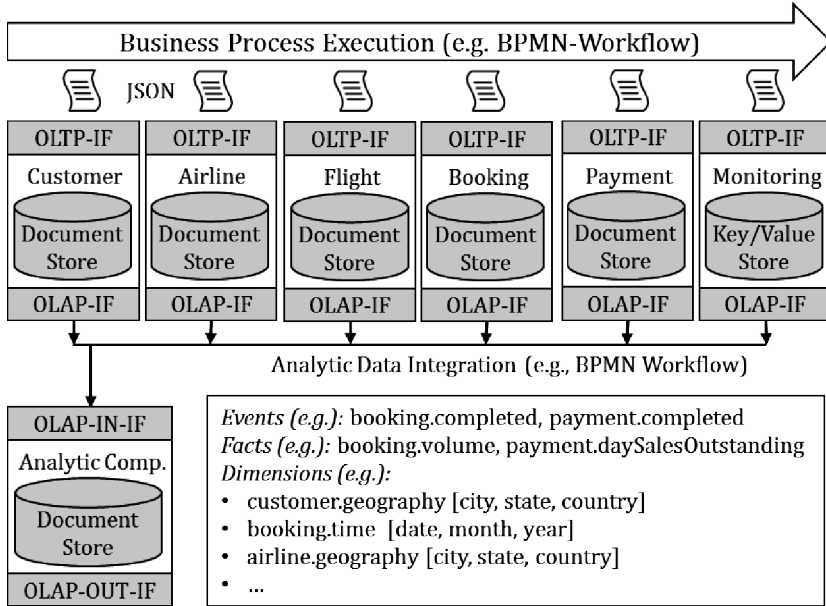


Figure 5: The proposed architecture to automate the business process on procurement of flights.

Figure 6 shows by example how the analytic data component can use the Map/Reduce-paradigm in order to perform business analysis. On the left side, the incoming integrated analytic data is shown. This document has the two named measures and dimensional information for geography and time. The map-function is used to build key/value-pairs. Keys are derived from the extensions of the dimension levels. Each level has to be combined with each level of all other dimensions. The values of the measures remain untouched. By using the reduce-function, the values of documents with the same key are aggregated. To get the average days-sales-outstanding and the total volume of all bookings, done in Bavaria on May 21st, 2013 all documents with key “2\_bavaria.1\_2013-05-21” are reduced to a single data set.

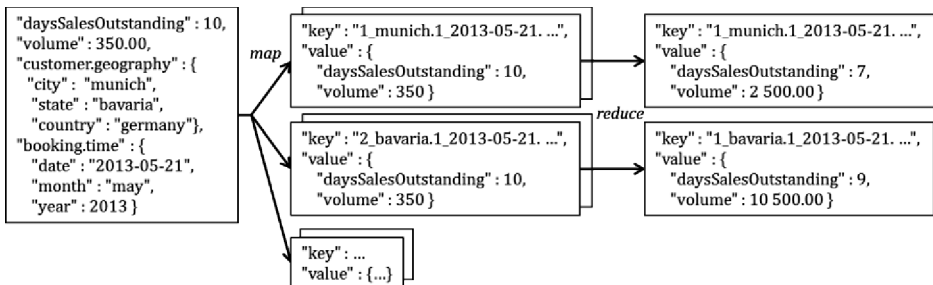


Figure 6: Exemplary application of Map/Reduce on the case study on the procurement of flights.

## 6 Discussion and Future Work

A new generation of non-relational data stores (NoSQL) has been emerging in recent years. The objective of this paper is to show how these Web 2.0-driven data stores can be used in enterprise application systems. For this, we proposed an architecture that enables the use of non-relational data stores in enterprise applications with respect to the requirements of polyglot persistence and of integrating OLTP- and operational right-time OLAP-scenarios. In the first step of research, the focus is restricted to a single business process. The architecture enables the deployment of the appropriate data model for a certain business entity with respect to its technical and structural characteristics. Furthermore, we propose the concept of operational data components, one for each business entity, to manage the heterogeneous data stores and to integrate their functionality in higher transactional workflows. These components also enrich their business data with multidimensional information. The multidimensional information of all components is integrated and can be analyzed by using the Map/Reduce-paradigm in an analytic data component. We have shown the application of the architectural concept and its analytic capabilities by using a case study on the procurement of flights.

The presented architecture is complementary to traditional data warehouse systems. It can be used in order to deal with the shortcomings because of periodic (long term) data load to support short term tactical and operational business process analysis. By using multidimensional data structures, historical and aggregating data analysis is enabled. Monitoring components track data in order to integrate run-time information in the business process analysis. To conclude, the hybrid architecture offers analytic capabilities within the boundaries of a single business process.

Figure 7 shows the time-value-curve of [Hac04] (adjusted by [Rus11]). Decision situations sharing the characteristics of the curve have the following in common: the more time that elapses following a business event, the more value they lose. The hybrid component architecture can contribute to shortening capture and analysis latency. Of course this is not the case if real-time reaction (within milliseconds) is required. But it is well suited for “right”-time problems. This means, data is delivered when a decision maker needs up-to-date information. We offer the possibility of defining that right point in time by specifying temporal, business, run-time and consumer events in operational and analytic data components. These events trigger the analysis of business process data. Capture latency is mainly reduced by short running, non periodically analytic data integration processes; analysis latency is merely influenced by Map/Reduce-strategies and hardware settings of the Map/Reduce-engine.

For future work we have to refine the architectural concept and enhance it in scenarios spanning more than one business process. We made initial propositions regarding which functionality which type of interface has to offer. Future work will need to address more detail in the design of all interface types. In the case of transactions it has to be discussed whether weaker concepts compared to ACID-transactions are sufficient in some scenarios and how they can be determined and realized. Further, how to realize data consistency in and between data components is a vital area for investigation. Persisting data denormalized and aligned to a single application system may result in redundancies. In order to prevent anomalies and inconsistency, concepts and strategies for data components and workflows have to be developed.

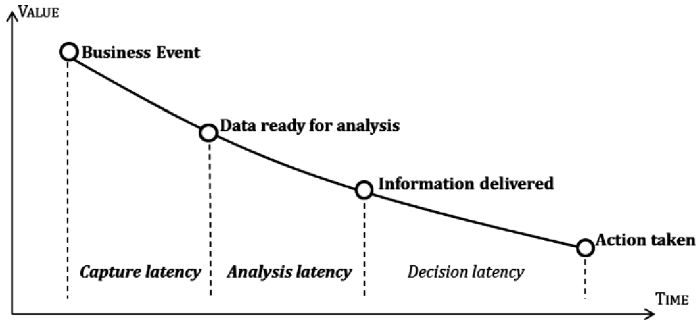


Figure 7: The time-value-curve according to [Hac04](adjusted by [Rus11]).

Parameterizing operational and analytic components, as well as realizing the integration process, are further topics of research. In order to design application systems on the proposed architecture, there has to be a design method, starting at a conceptual business process level and offering integrated design concepts (e.g. meta-models, transformation rules) to derive an application specification that is as complete as possible. It should cover analytic and operational requirements, non-functional requirements to identify data store and transaction concepts, identify data exchange patterns and formats as well as business entities and their managing data components with interfaces and realization. As mentioned earlier, the authors of [Rus11] suggest having bi-directional integration between business and analytic processes. We have to enhance the hybrid architecture to be able to react automatically on certain business events, and therefore shortening the decision latency. Finally, the proposed architecture needs to be evaluated in a more complex case study, including the realization based on a certain technology stack. This should lead to further research topics and substantiate the presented blueprint.

## References

- [ACN05] B. Azvine, Z. Cui, and D. Nauck. Towards real-time business intelligence. *BT Technology Journal*, 23(3): 214-225, 2005.
- [Bre00] Eric A. Brewer. Towards robust distributed systems. (Invited Talk) Principles of Distributed Computing, July 2000.
- [CCS93] Edgar F. Codd, Sharon B. Codd, and Clynh T. Salley. Providing OLAP (On-Line Analytical Processing) to User-Analysis: An IT Mandate, 1993.
- [Cod70] Edgar F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377-387, 1970.
- [Cro06] Douglas Crockford. Introducing JSON. <http://tools.ietf.org/html/rfc4627>, 2006. Online accessed May, 21st, 2013.
- [DG08] Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, 51(1):107-113, 2008.
- [Fie00] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. PhD thesis, University of California, Irvine, 2000.
- [Gho10] Debasish Ghosh. Multiparadigm Data Storage for Enterprise Applications. *IEEE Software*, 27(5):57-60, 2010.
- [GL02] Seth Gilbert and Nancy Lynch. Brewer’s conjecture and the feasibility of consistent, available, partition-tolerant web services. *ACM SIGACT News*, 33(2):51, 2002.

- [GR93] Jim Gray and A. Reuter. *Transaction processing: Concepts and techniques*. The Morgan Kaufmann series in data management systems. Morgan Kaufmann Publishers, San Mateo and Calif, 1993.
- [Hac04] Richard Hackathorn. The BI Watch: Real-Time to Real-Value. <http://bolder.com/pubs/DMR200401-Real-Time%20to%20Real-Value.pdf>, 2004. Online accessed May, 21st, 2013.
- [Hec11] Robin Hecht. NoSQL evaluation: A use case oriented survey. In Stefan Jablonski, editor, *Proceedings of the International Conference on Cloud and Service Computing*, pages 336–341, 2011.
- [Ind12] Maria Indrawan-Santiago. Database Research: Are We at a Crossroad? Reflection on NoSQL. *15th International Conference on Network-Based Information Systems*, pages 45–51, 2012.
- [KHB<sup>+</sup>11] Jens Krüger, Florian Hübner, Martin Boissier, Johannes Wust, Alexander Zeier, and Hasso Plattner. Main Memory Databases for Enterprise Applications. In *IEEE IE&EM*, 2011.
- [KN11] Alfons Kemper and Thomas Neumann. One Size Fits all, Again! The Architecture of the Hybrid OLTP&OLAP Database Management System HyPer. In Wil Aalst, John Mylopoulos, Michael Rosemann, Michael J. Shaw, Clemens Szyperski, Malu Castellanos, Umeshwar Dayal, and Volker Markl, editors, *Enabling Real-Time Business Intelligence*, volume 84 of Lecture Notes in Business Information Processing, pages 7–23. Springer Berlin Heidelberg, Berlin and Heidelberg, 2011.
- [OMG11] OMG. Business Process Model and Notation (BPMN): Version 2.0. <http://www.omg.org/spec/BPMN/2.0/PDF>, 2011. Online accessed May, 21st, 2013.
- [Pla09] Hasso Plattner. A common database approach for OLTP and OLAP using an in-memory column database. In *Proceedings of the 35th SIGMOD international conference on Management of data*, pages 1–2. ACM, Providence and Rhode Island and USA, 2009.
- [Rus11] Philip Russom. TDWI Best Practices Report: Operational Data Warehousing: The Integration of Operational Applications and Data Warehouses. available as PDF on <http://tdwi.org/research/2010/10/bpr-q4-operational-data-warehousing.aspx?tc=page0>, 2011. Online accessed May, 21st, 2013.
- [SF12] Pramod J. Sadalage and Martin Fowler. *NoSQL distilled: A brief guide to the emerging world of polyglot persistence*. Addison-Wesley, Upper Saddle River and NJ, 2012.
- [Som11] Ian Sommerville. *Software engineering*. Pearson, Boston and Mass and London, 9 edition, 2011.

## **Methods**



# Cognitive Efficient Modelling Using Tablets

Jeannette Stark, Martin Burwitz, Richard Braun, Werner Esswein

Technische Universität Dresden  
01062 Dresden

{ jeannette.stark | martin.burwitz | richard.braun | werner.esswein }@tu-dresden.de

**Abstract:** Recently, conceptual modellers have to choose between modeling tools running on tablets or in a common PC environment. This choice has some implications on how cognitive efficient the modeling process is. While using touch input (given on a tablet) eyes and hand are directly coordinated, which supports human gaze behaviour. On the contrary, modelling with a mouse (given in a common PC environment) does not allow direct eye-hand coordination and therefore needs additional effort to coordinate eye and hand. This effect is investigated within a two-group laboratory experiment using BPMN as a modeling language and thus assesses if modeling on tablet is cognitive efficient. Furthermore, we derive assumptions how improve cognitive efficiency of modeling tools.

## 1 Introduction

Interacting with an application by directly touching its graphical elements seems more natural than working with a mouse or another pointing device [For07]. This argument is valid also for conceptual modeling, which is why some modeling tools have been developed for tablet-use, recently (e.g. Cubetto BPMN and Mobile Process Designer). But why should modeling on tablets be more natural? And what does naturalness mean?

Several experiments have shown that conceptual modeling can be supported by adjusting the process of modeling [AST96] as well as the modeling grammar [Par02] in such a way that fits human cognition [LS87]. Depending on what has been adjusted to human cognition, the modeling process results more efficient or understanding of the resulting model can be improved. In these cases naturalness has been achieved by distinguishing between entities and properties [Web96], by adjusting the problem representation to the task [AST96] or by allowing conceptual and perceptual integration of different conceptual models [Moo03].

Stark and Esswein have conducted a review of experiments within Conceptual Modeling that base on Cognition. They could neither identify a study that has focused on input devices in particular nor did we find any study with a focus on how a modeling tool can be adjusted to fit human cognition in general [SE12].

Yet, within the field of Human Computer Interaction (HCI) advantages and disadvantages of different input possibilities, such as mouse-input for PC-use or touch-input for tablet-use, have been discussed and investigated. Tablets are easy to use, do not require any additional workspace and do not have moving parts. But they lack precision [BS78], [Pic86]. It is difficult to implement precise selection mechanisms as the finger as input medium is usually larger than the area which needs to be selected [HA11]. On the contrary, the mouse is the standard input medium and tool designers can suppose that conceptual modellers are familiar with it. Nevertheless, using a mouse requires an extensive eye-hand coordination [GA88], since hand-positioning cannot be supported by the modellers gaze [CMM04], [Des98].

Prior research of HCI has tested efficiency and effectiveness of touch and mouse input. AHLSTRÖM ET AL., KARAT ET AL. and MEYER ET AL. have investigated errors generated by performing pointing tasks [AL87], [KMA86] and drawing tasks [MCN94]. Yet, these investigations were performed over 20 years ago, which is why their findings may not represent results obtained with current hardware [HA11]. Two experiments were conducted recently. FORLINES ET AL. have investigated touch and mouse input for pointing and dragging tasks using one and two hands. They used a horizontal 107 cm diagonal MERL DiamondTouch touch sensitive table displaying 1024x768 pixels with each pixel approximately 1.2mm square. For one-hand input tasks touch and mouse input showed similar performance. More specifically, touch input was slightly but not significantly outperforming mouse input. For two-hand tasks using touch input significantly outperformed using mouse input [For07]. HOOTEN AND ADAMS investigated mouse and touch input for drawing tasks. They used a DELL XT2 tablet laptop that allowed touch-based interaction and found out that for drawing tasks touch-based interaction is significantly faster than mouse-based interaction [HA11]. Both experiments cannot simply be projected to conceptual modeling on tablets for two reasons: First, there is a difference in hardware requirements. Conceptual modeling tools for tablets have so far been developed for ipad use, which is currently sold with a size of 24,6 cm (width) and 9.7 inch displaying 2.048x1.536 pixels (1024x768 ipad1). Thus, the size of the display is significantly smaller than in [For07] and slightly smaller than in [HA11]. Furthermore the pixel-size on ipads is significantly different from that used in [For07]. Second, so far conceptual modeling on ipad requires only pointing and dragging tasks but no drawing task. Elements can be selected by pointing on the screen. Dragging is used to relate different elements to each other. Consequently, results of [HA11], who investigated drawing tasks, cannot be applied for modeling on tablets, so far.

This study aims on empirically investigating if touch-input for modeling is “more natural” than mouse-input. We claim that a direct eye-hand coordination (given with touch-input) is more cognitive efficient than an indirect eye-hand coordination (given with mouse-input). We describe a laboratory experiment, which evaluates the effect of eye-hand coordination for conceptual modeling. This research contributes to research and practice. For practice, we identify the effect touch- and mouse-input has on the modeling process and thus, give developers of conceptual modeling tools an idea of how to make their tool cognitive efficient for the input variable. Furthermore, we derive propositions how to develop cognitive efficient modeling tools.

For research we integrate our results within the framework of rules for conceptual modeling from cognition [SE12] that summarizes independent and dependent variables from empirical investigations within conceptual modeling that base on cognition. Furthermore, the authors have discussed future research possibilities to test cognitive efficiency of the grammar, use of grammar, user and task characteristics as well as medium of content delivery. We add future research possibilities for the modeling tool. The paper is structured as follows: In section 2 we describe prior research of eye-hand coordination. In section 3 the method is described. Results are presented in section 4 and discussed in section 5. The paper concludes with a summary in section 6.

## **2 Theoretical Background and Hypotheses**

Conceptual modeling using touch-input differs from modeling using mouse-input in eye-hand coordination. In this section these differences are discussed and hypotheses are presented.

### **2.1 Eye-Hand Coordination**

When pointing to a single target, people direct their gaze to the target while they initiate their hand movement. They maintain their gaze on the target until the hand arrives [GA88], [Des98]. This behaviour improves reach accuracy [BJF09]. Looking at the target allows an effective use of visual feedback of the hand position while guiding the hand to the target [Lan99], [Pai96]. This gaze behaviour normally happens when performing simple tasks such as pointing tasks [NB00], drawing tasks [RS03] as well as more complicated tasks such as manipulating tasks [Joh01].

If the pointer directed by the mouse arrives the target instead of the hand, eye and hand are still coordinated. Efferent and afferent signals related to gaze position can be used to guide the hand when the hand is not visible [PDG03]. This happens if the pointer of the mouse moves towards or away from the target. In this situation an indirect eye-hand coordination occurs.

Conceptual modeling using paper requires drawing tasks, which are supported by direct eye-hand coordination. However, modeling on paper reveals some important disadvantages as the result cannot easily be shared and worked on within physically separated groups [Her07]. Furthermore creating versions and variants of the model requires reconstructing the model on a new paper.

Using modeling tools bridges these gaps and allows to save, to change and to share the results. On traditional modeling tools the input of model elements and labels requires the use of a mouse while the output is usually displayed on a screen. Creating a conceptual model using a mouse usually requires both pointing and dragging tasks. [MSB91] have shown that dragging is a variation of pointing. While performing pointing tasks a typical gaze behaviour is focusing the target until the hand arrives [LBS09]. Since a cursor, directed by a mouse instead of the hand, arrives the target, additional effort is required to

coordinate eye and hand. In addition, the movement amplitude of the mouse and the precision of the hand movement required influence the effort required for eye-hand coordination [SJ05].

Traditional modeling tools usually offer a wide range of functionalities such as configuration management, application of reference models or creation of modeling languages. Yet, these tools are generally not applicable for modeling on spot. This is where tablet-use is advantageous.

Only recently tablets have been discussed for conceptual modeling. Especially mobile process modeling might be relevant for use cases such as modeling procedure models in health care (model-based clinical treatment processes) or for the adaptation of process models in logistical processes. However, there are only a few tools for modeling on tablets: Some process modeling tools for the iPad (e.g. Cubetto BPMN and Mobile Process Designer) and some tools for modeling with UML (e.g. astah\* UML Pad). In addition, there are a few web based modeling tools that can be used conveniently on tablets (e.g. Signavio). The most evident advantage of using tablets for conceptual modeling is that the input is also the output device, which is why there is a direct eye-hand coordination as well as a direct relationship between the modeller's input and the output on the tablet [GA88]. Creating a model on a tablet normally requires pointing and dragging tasks that belong to those tasks that benefit from a direct eye-hand coordination. Eye-hand coordination of modeling on paper, with traditional modeling tools or on tablets is summarized in Fig. 1.

|                       | Modeling on Paper                                                                                                          | Modeling with mouse input                                                                                                    | Modeling with touch input                                                                                                   |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Eye-hand coordination | Direct coordination of eye and hand<br> | Indirect coordination of eye and hand<br> | Direct coordination of eye and hand<br> |
| Task                  | Drawing                                                                                                                    | Dragging and pointing                                                                                                        | Dragging and pointing<br>(drawing so far not implemented)                                                                   |

Figure 1: Eye-Hand Coordination for Modeling on paper, with a mouse or touch input

## 2.2 Hypotheses

As modeling on paper became uncommon with the advances of information technology and its subsequent development of modeling tools we have focused on modeling using tablets (touch-input) and PC (mouse-input). We argue that direct eye-hand coordination is cognitive efficient and thus, leads to a higher modeling efficiency.

We have operationalized this hypothesis in a two-group laboratory experiment, one group working on a tablet and one group working with a PC. As a measurement for modeling efficiency we use the time required to complete the task. We argue that:

**H1:** (time) *The tablet-group will need less time to complete the modeling task than the PC-group.*

Performance-based measurements, such as required time to complete the model generally provide more convincing evidence than perception-based measurements, such as perceived Ease of Use [Moo03]. Yet, for decisions to use a tablet-application or a traditional modeling tool it is more important how efficient the modeller thinks the tool is. According to the work of MOODY we include three perception-based measurements: Perceived Ease of Use, Perceived Usefulness, Intention to Use. The method evaluation model is presented in Fig. 2.

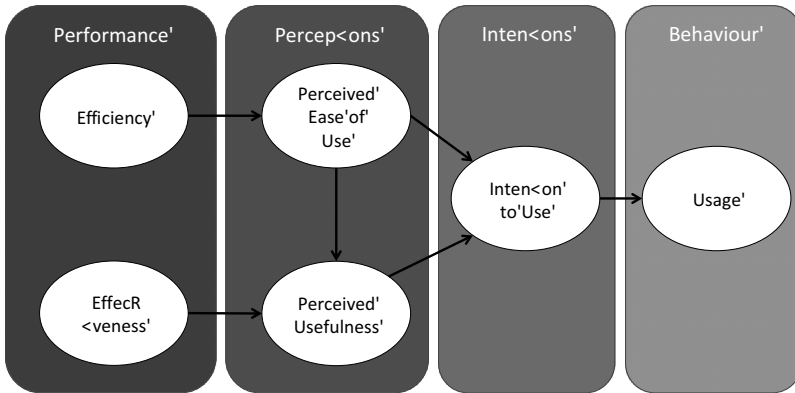


Figure 2: Method Evaluation Model [Moo03]

Based on the Method Evaluation Model we argue:

**H2:** (Perceived Ease of Use) *The tablet-group will believe to a higher degree that conceptual modeling would be free of effort than the PC-group.*

**H3:** (Perceived Usefulness) *The tablet-group will believe to a higher degree that using the modeling software will be effective to achieve their objectives than the PC-group.*

**H4:** (Intention to Use) *The tablet-group will intend to a higher degree to use the modeling software than the PC-group.*

### 3 Research Methodology

We have investigated the above-presented hypotheses in a two-treatment group laboratory experiment including pre- and post-test. The experimental design is shown in Fig. 3.

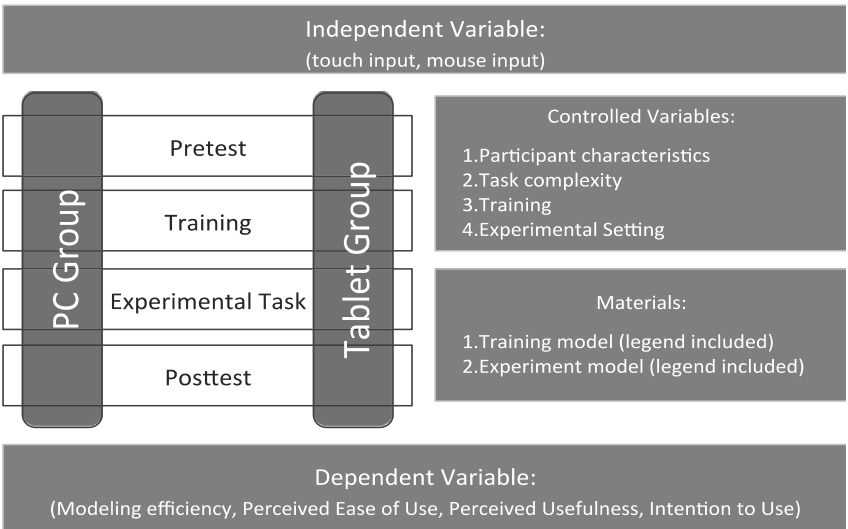


Figure 3: Experimental Design of the study

#### 3.1 Task Setting

We have operationalized modeling efficiency by measuring the time it requires participants to rebuild a conceptual model on a tablet (Treatment Group 1) or on PC (Treatment Group 2). On tablet, participants worked with the ipad application Cubetto BPMN [Grei10]. We have selected the Business Process Modeling Notation (BPMN) as modeling language for the reason that we tried to reduce affecting variables such as different layout and different handling of the program. This is why we have selected a program available on tablet and on PC that requires the same handling. We only found Cubetto BPMN on tablet and a simulator program of Cubetto BPMN on PC. These two programs only differ in their input-mediums (mouse and keyboard on PC and touchscreen on tablet). Furthermore, we have reduced keyboard-use to a minimum, which is why labels of elements only consisted of a few letters. The experiment model is shown in Fig. 4. We expected that participants did not have any previous knowledge in modeling with BPMN. To enable them using the modeling environment on tablet or on PC each experimental group was given a 45 minutes training session. Part of the session was explaining the modeling environment and together rebuilding a model that contains all modeling elements that they needed to complete the experimental task. The training model is presented in Fig. 4, too.

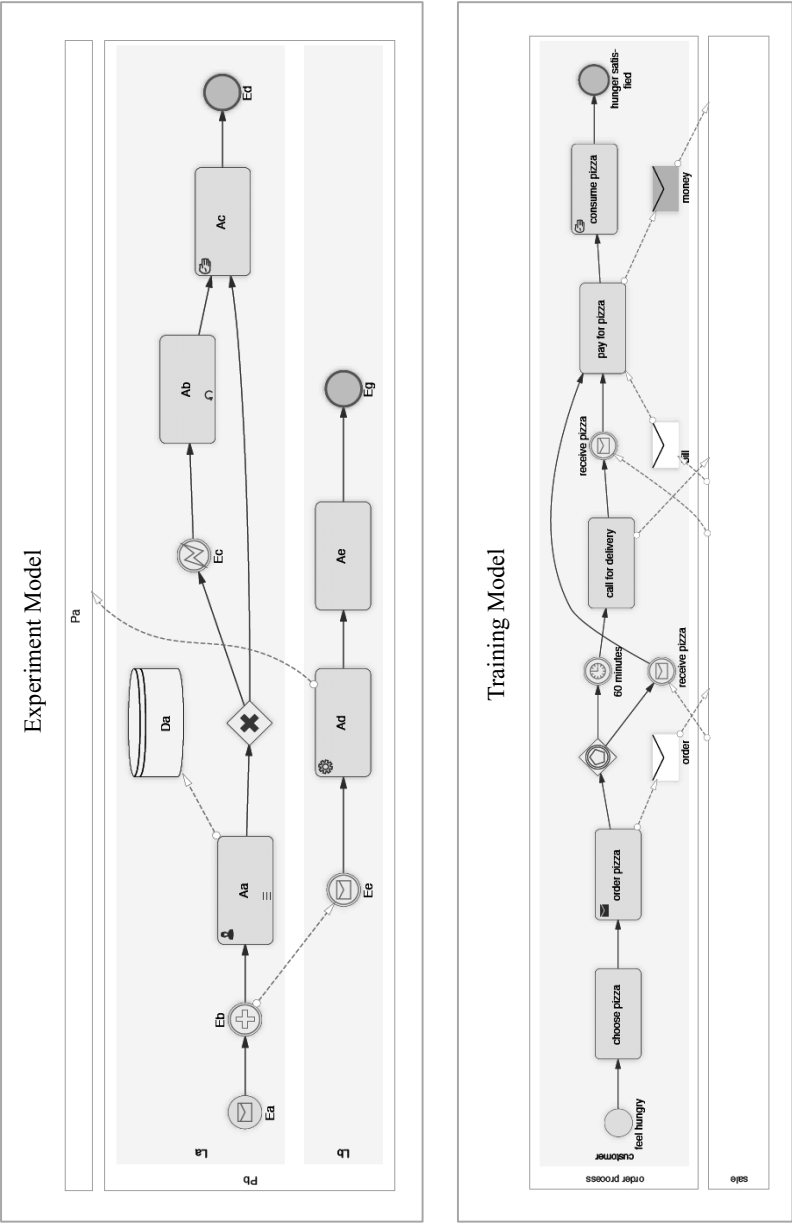


Figure 4: Experiment and Training model for PC- and tablet-group

After concluding the experimental task, students were given a post-test consisting of 13 questions for operationalizing Perceived Ease of Use, Perceived Usefulness and Intention to Use as well as to control possible affective variables such as if a participant felt to be set in a very stressful situation. For the post-test a 0-7 scale was used developed by [MB91].

### **3.2 Participants**

We invited 40 business and information systems students participating in a General Qualification Course for undergraduate Students to take part in the experiment. The students were randomly assigned to the tablet-group (treatment group 1) and to the PC-group (treatment group 2). We used a pre-test to assure that none of the students had previous knowledge using BPMN or had previously used the modeling tool. Students could reduce their expenditure of work by taking part in the experiment. Apart from three all students decided to take part in the experiment.

### **3.3 Pilot Study**

We had conducted a pilot study with undergraduate students not taking part in the General Qualification Course. This study helped to eliminate ambiguity in pre- and post-test questions as well as finding a number of constructs for training and experimental model that students would not be overburdened and thus, would be set in a stressful experimental atmosphere.

## **4 Results**

### **4.1 Performance-based Measurements: Modeling Efficiency (Time)**

PC- as well as tablet users have exactly rebuilt the paper model on tablet or on PC and did not leave out constructs or choose to take another design within the time taken. The mean time required to rebuild the model is 7 minutes and 21 seconds for the tablet-group seconds 8 minutes and 32 seconds for the PC-group. The times for modeling on tablets and on PC do not follow a normal distribution (Shapiro-Wilk-Test,  $\alpha < 0.02$ ). Since the times for both groups show a similar distribution (Kolmogorow-Smirnow-Test,  $\alpha < 0.02$ ) we decided to apply a Mann-Whitney U test to assess if the means differ significantly. Table 1 shows the means, standard deviation and significance for both groups.

Table 1: Times for tablet- and PC-group

| Treatment Group: | Mean ( $\mu$ )<br>in mm:ss | STDEV ( $\delta$ )<br>in mm:ss | Significance |
|------------------|----------------------------|--------------------------------|--------------|
| Tablet Group     | 07:21                      | 01:39                          | 0,02         |
| PC Group         | 08:32                      | 01:48                          |              |

A significant difference between the tablet- and the PC-group was found for  $\alpha < 0.02$  which is why we can assume that for  $\alpha < 0.02$  the means differs significantly. Thus, H1 is strongly confirmed.

## 4.2 Perception-based Measurements

For Perception-based measurements we have used the Shapiro-Wilk-Test to show that for every measurement (Perceived Ease of Use, Perceived Usefulness and Intention to Use; both groups) data collected does not follow a normal distribution ( $\alpha < 0.02$ ), which is why we could neither use an analysis of variance nor a t-test. Since data is equally distributed (according to a Kolmogorow-Smirnow-Test) we could apply the Mann-Whitney U Test. Results of the tests are presented below in Table 2.

Table 2: Perception-based measurements for tablet- and PC-group

| Dependent Variable    | Treatment Group: | Mean ( $\mu$ )<br>(0-7 scale) | STDEV ( $\delta$ )<br>(0-7 scale) | Significance |
|-----------------------|------------------|-------------------------------|-----------------------------------|--------------|
| Perceived Ease of Use | Tablet Group     | 6,35                          | 0,76                              | -            |
|                       | PC Group         | 5,95                          | 1,89                              |              |
| Perceived Usefulness  | Tablet Group     | 6                             | 0,74                              | -            |
|                       | PC Group         | 5,85                          | 1,06                              |              |
| Intention to Use      | Tablet Group     | 5,95                          | 0,91                              | -            |
|                       | PC Group         | 5,55                          | 1,43                              |              |

According to Table 2 the means of Perceived Ease of Use, Perceived Usefulness and Intention to Use of the tablet-Group outperform slightly those of the PC-group. Yet, we could not identify a significant difference between those two groups ( $\alpha < 0.02$ ). Hence, H2, H3 and H4 are not supported.

# 5 Discussion and Future Research

## 5.1 Discussion of the Results

Out of four hypotheses one was supported while the others lack any support. The mean time required to complete the modeling task was found to be significantly shorter for the tablet-group than for the PC-group, which is why we can assume that with a direct eye-hand coordination modeling efficiency can be enhanced. Thus, a direct eye-hand coordination is cognitive efficient for conceptual modeling.

For perception-based measurements the tablet-Group also slightly outperforms the PC-Group but we could not find any significant difference between the two groups. We might explain the loss of significance with the rating behaviour of the participants. On a 0-7 point scale the means range between 5,55 and 6,35. These means show that the students have ranked Perceived Ease of Use, Perceived Usefulness and Intention to Use for both tablet and PC relatively high, which is why discrimination for the two groups was hardly possible. A further reason might be that students mainly work with the mouse as input medium and hence, are more used to it. This behaviour might effect that they do not feel more comfortable using touch input while completing the modeling task.

## 5.2 Strength and Limitations of the Study

To secure internal validity the variables such as participants' characteristics, tool characteristics, experimental setting, task complexity and training have been controlled (see Table 3). We could not control the input medium for text. The PC-group could use an external keyboard while the tablet-group had to use an integrated keyboard on the tablet, which is smaller than the PC-keyboard. To reduce the effect of this variable we have only used very short names for the elements.

The type of participants we have chosen influences external Validity. As many other laboratory experiments in the field of conceptual modeling we have asked students to participate. Their knowledge of conceptual modeling is relatively low in contrast to that of practitioners. However, the task was designed to test modeling efficiency of rebuilding a model that participants are given on paper. Provided with a training how to use the modeling tool, students should not be disadvantaged while performing this task.

Table 3: Controlled Variables

| Variable                     | Control mechanisms                                                                                                   |
|------------------------------|----------------------------------------------------------------------------------------------------------------------|
| Participants' characteristic | Participants stem from the same course. Students only took part in the experiment if they had any tablet experience. |
| Tool                         | We have used the same tool running on tablet and on PC. The                                                          |

|                      |                                                                                                                                                                                                                                            |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| characteristics      | only difference is the input-medium.                                                                                                                                                                                                       |
| Experimental setting | We have run several experimental sessions, as we did not have enough tablets. In each session we have integrated participants of both experimental groups. That way conditions for tablet-group-members and PC-group-members were similar. |
| Task complexity      | Since the modeling tools were identical on tablet and on PC apart from the input medium, we could use the same task for both experimental groups.                                                                                          |
| Training             | Since participants did not have any previous knowledge in conceptual modeling, members of both groups have been trained before. The training for both groups was the same.                                                                 |

### 5.3 Future Research

STARK & ESSWEIN have reviewed independent and dependent variables for conceptual modeling based on cognition and integrated these variables into a framework. They have found independent variables that concern Medium of content delivery (e.g. narrated instead of written explanation [Gem04]), User characteristics (e.g. modeling language experience [RD07]), Task characteristics (e.g. task matches languages [AST96]), Grammar's constructs (e.g. distinction between entities and properties [Web96]) and Use of Grammar (e.g. classification rules [PW08]). They did not find any variable concerning the modeling tool. With this study we extend the framework for the modeling tool (see Fig. 5). We have investigated the effect of eye-hand coordination (given with the touch input) on the dependent variables Modeling Efficiency, Perceived Ease of Use, Perceived Usefulness and Intention and have found a significant influence of eye-hand coordination on modeling efficiency.

Based on the experiment we have received comments from participants. Some students found the layout algorithm we used to distribute model elements on the screen as helpful. Others perceived the same algorithm as disruptive. Based on these comments we asked a few students to complete the same experimental task for a traditional modeling tool not offering a layout algorithm. We found significant differences for the mean times. Using a layout algorithm for conceptual modeling might influence model understanding. On the one hand elements are equally distributed among the screen. That way the model is balanced. On the other hand the modeler might not be able to fit modeling elements that belong together into the same part of the screen. This influence of using a layout algorithm on modeling efficiency and model understanding still needs to be investigated.

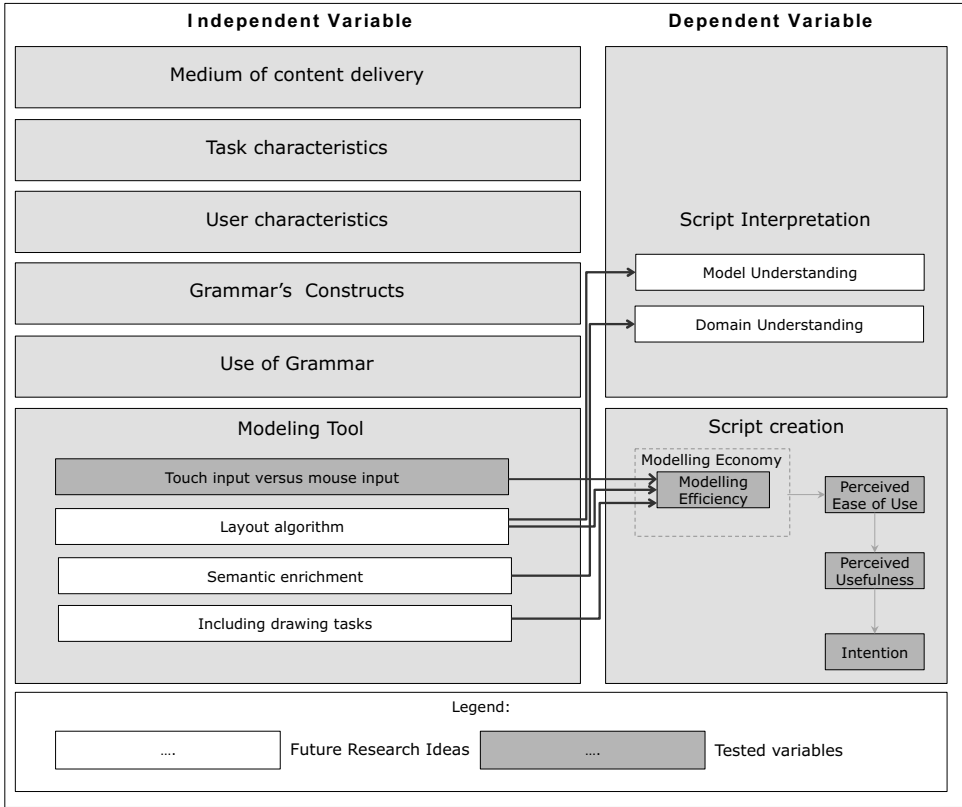


Figure 5: Variables tested and future research ideas

So far conceptual modeling tools require pointing and dragging tasks. [For07] have found significant advantages for modeling efficiency and effectiveness for drawing tasks using touch input. Besides drawing tasks might offer new possibilities for modeling language development as concrete syntax might be developed more easily.

A further starting point to make modeling tools cognitive efficient is integrating speech to obtain semantically enriched modeling elements. GEMINO has investigated the effect of narrated explanation of a model for domain understanding. Integrating speech within the modeling tool make explanations available for the model user without having to use annotations.

## 6 Summary

In this study we have investigated the effect of eye-hand coordination (given on a tablet) on modeling efficiency as well as on perception-based measurements. We found that modeling on a tablet has a significant effect on modeling efficiency and a slight but not significant effect on Perceived Ease of Use, Perceived Usefulness and Intention to Use.

For practitioners this might be a starting point to use the advantages a modeling tool for tablets offers not only for modeling efficiency but also for using the tool whenever it is needed.

For researchers we have started to investigate features of a cognitive efficient modeling tool and have furthermore derived future research ideas such as investigating the effect of a layout algorithm of modeling efficiency and model understanding.

## References

- [AL87] Ahlström, B., Lenman, S.: Touch screen, cursor keys and mouse interaction. In: Selected papers from the International Scientific Conference on Work with display units 86, 1987, p 831-837.
- [AST96] Agarwal, R., Sinha, A.P., Tanniru, M.: Cognitive fit in requirements modeling: A Study of object and process methodologies. In: Journal of Management Information Systems 13(2), 1996 , p 137-162.
- [BJF09] Bowman, M.C., Johannson, R.S., Flanagan, J.R.: Eye-hand coordination in a sequential target. In: Experimental brain research 195(2), 2009, p 273-283.
- [BS78] Brosey, M., Shneiderman, B.: Two experimental comparisons of relational and hierarchical database models. In: International Journal of Man-Machine Studies 10(6), p 625-637.
- [CMM04] Crawford, J.D., Medendorp, W.P. & Marotta, J.J.: Spatial transformations for eye-hand coordination. In: Journal of Neurophysiology, 92(1), 2004, p 10-19.
- [Des98] From eye to hand: planning goal-directed movements. In: Neuroscience & Biobehavioral Reviews, 1998, 22(6), p 761-788.
- [For07] Forlines, C et. al.: Direct-touch vs. mouse input for tabletop displays. In Proceedings of the SIGCHI conference on Human factors in computing systems, 2007, p 647-656.
- [GA88] Greenstein, J.S., Arnaut, L.Y.: Input devices. In Handbook of human-computer interaction, 1988, p. 495-519.
- [Gem04] Gemino, A.: Empirical comparisons of animation and narration in requirements validation. In: Requirements Engineering, 9(3), 2004, p 153-168.
- [Gre10] Greiffenberg, S.: *Cubetto toolset*®, Semture GmbH. Available at: <http://wise.wiwi.tu-dresden.de/cubetto>, 2010.
- [HA11] Hooten, E.R., Adams, J.A.: Comparing input error for mouse and touch input. In Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, 2011, p 2853-2858.
- [Her07] Herbsleb, J.D.: Global software engineering: The future of socio-technical coordination. In *Future of Software Engineering*, 2007, S. 188-198.
- [Joh01] Johansson, R.S. et al.: Eye-hand coordination in object manipulation. In: the Journal of Neuroscience, 21(17), 2001, p 6917-6932.
- [KMA86] Karat, J., McDonald, J., Anderson, M.: A comparison of selection techniques: touch panel, mouse, keyboard. In: International Studies of Man-Machine Studies, 1, 1986, p. 73-92.
- [Lan99] Land, M. et al.: The roles of vision and eye movements in the control of activities of daily living. In: PERCEPTION-LONDON 28(11), 1999, p. 1311-1328.
- [LBS09] Liesker, H., Brenner, E., Smeets, J.B.: Combining eye and hand in search is suboptimal. In: Experimental brain research 197(4), 2009, p 395-401.
- [LS87] Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. In: Cognitive science, 11(1), 1987, p 65-100.
- [MB91] Moore, G.C. & Benbasat, I.: Development of an instrument to measure the perceptions

- of adopting an information technology innovation. In: *Information systems research*, 2(3), 1991, p 192–222.
- [MCN94] Meyer, S., Cohen, O. & Nilsen, E.: Device comparisons for goal-directed drawing tasks. In *Conference companion on Human factors in computing systems, 1994*, p 251–252.
- [Moo03] Moody, D.: Comparative evaluation of large data model representation methods: The analyst's perspective. In: *Conceptual Modeling - ER 2002*, p 214–231.
- [MSB91] MacKenzie, I.S., Sellen, A. & Buxton, W.A.: A comparison of input devices in element pointing and dragging tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, 1991*, p 161–166.
- [NB00] Neggers, S.F., Bekkering, H.: Ocular gaze is anchored to the target of an ongoing pointing movement. In: *Journal of Neurophysiology*, 83(2), 2000, p 639–651.
- [Pai96] Paillard, J. et al.: Fast and slow feedback loops for the visual correction of spatial errors in a pointing task: a reappraisal. In: *Canadian journal of physiology and pharmacology*, 74(4), 1996, p 401.
- [Par02] Parsons, J.: Effects of local versus global schema diagrams on verification and communication in conceptual data modeling. In: *Journal of Management Information Systems*, 19(3), 2002, p 155–183.
- [PDG03] Prablanc, C., Desmurget, M. & Gréa, H.: Neural control of on-line guidance of hand reaching movements. In: *Progress in brain research*, 142, 2003, p 155–170.
- [Pic86] Pickering, J.A., 1986. Touch-sensitive screens: the technologies and their application. In: *International Journal of Man-Machine Studies*, 25(3), 1986, p 249–269.
- [PW08] Parsons, J. & Wand, Y.: Using cognitive principles to guide classification in information systems modeling. In: *MIS Quarterly*, 32(4), 2008, p.839–868.
- [RD07] Recker, J.C., Dreiling, A.: Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education, 2007.
- [RS03] Reina, G.A., Schwartz, A.B.: Eye–hand coupling during closed-loop drawing: Evidence of shared motor planning? In: *Human movement science*, 22(2), 2003, p 137–152.
- [SE12] Stark, J. & Esswein, W.: Rules from Cognition for Conceptual Modeling. In *Proceedings of ER 2012*. Florence: Springer.
- [SJ05] Sandfeld, J., Jensen, B.R.: Effect of computer mouse gain and visual demand on mouse clicking performance and muscle activation in a young and elderly group of experienced computer users. In: *Applied Ergonomics*, 36(5), 2005, p 547–555.
- [Web96] Weber, R.: Are attributes entities? A study of database designers' memory structures. In: *Information Systems Research*, 7(2), 1996, p 137–162.

# SecSy: Synthesizing Process Event Logs

Thomas Stocker and Rafael Accorsi  
University of Freiburg, Germany  
{ stocker | accorsi }@iig.uni-freiburg.de

**Abstract:** One difficulty at developing mechanisms for business process security monitoring and auditing is the lack of representative, controllably generated test runs to serve as an evaluation basis. This paper presents an approach and the corresponding tool support for event log synthesis. The novelty is that it considers the activity of an “attacker” able to purposefully infringe security and compliance requirements or simply manipulate the process’ control and data flow, thereby creating deviations of the intended process model. The resulting logs can be readily replayed on a reference monitor, or serve as input for auditing tools based upon, e.g., process mining.

## 1 Introduction

Research into business process security and compliance is concerned with the requirements formalization of security (e.g. secrecy, binding of duties) and compliance requirements (e.g. obligations, interdependencies between activities) [BAS09] and the development of well-founded techniques and tools for analyzing, monitoring and auditing these requirements in business process specifications [Acc13]. Approaches for this may include, for instance, program analysis [AL12, AW10b, AW11], usage control monitoring [ASK08] and various types of a-posteriori process analysis [AS08, AW10a, vdA11].

Here, a particular challenge arises when it comes to testing the effectiveness of monitoring and auditing techniques and corresponding tools [AS12, AS13]. Specifically, to test these tools one needs controllably generated event logs that contain process executions comprising process flexibility and variability on the one hand, and process non-compliance on the other, thereby mimicking structural vulnerabilities, process dynamics, intentional attacks and user errors [LA11]. Such event logs can serve as input for monitoring, auditing and mining tools, thereby allowing developers to assess their *kill-rate*, i.e. the precision to identify the violation of the designated security and compliance properties or deviations from an original process model.

Business process simulation techniques can be used to generate these runs and, thus, to synthesize event logs. However, the current state of the art (e.g. [CKH98, Tum96, vdA10a]) and tool support (e.g. [Ala, BSB98, BS10, CPN], see [JVN06] for survey) do not take into account the controlled generation of event logs considering flexibility and non-compliance for a specific process.

In computer security, the “attacker model” describes activities that deviate from the prescribed behavior of programs, systems and protocols [Cer01]. This abstraction serves

not to solely model an “attacker”, but rather unexpected fault situations that interfere with the usual operation of a program or protocol. In security, if a program (provably) withstands a certain “attacker model”, then it is considered “secure” against such an attacker. This paper employs the “attacker model” abstraction as a building block of event log generation. Specifically, the paper presents a controllable simulation approach and corresponding tool support that allows users to synthesize event logs of business processes comprising non-compliance and business process flexibility wrt. their control flow.

The contributions of this paper are threefold: (1) the definition of model transformation operators to derive and encode variants of process models, e.g. skipping and swapping activities; (2) the definition of trace transformations which transform valid process traces by enforcing or violating security-related properties; (3) considering traditional security and organizational properties for business processes [KR01], the approach allows for a controllable generation of event logs in which these requirements are enforced or violated; and (4) the presentation and evaluation of `SecSy`, a tool that allows the configurable generation of event logs in several output formats. Due to space constraints, this paper will present only the main intuition behind the approach and key features of `SecSy`. (`SecSy` can be downloaded at <http://bit.ly/SecSy>). A follow-up paper will present the whole set of operators and a thorough description and evaluation of `SecSy`.

To the best of our knowledge, the insights behind the approach we present in this paper are novel. `SecSy` has been already used in the generation of several case studies and is capable of producing log files of industrial complexity that can be readily used for testing, e.g., process mining and monitoring approaches. Overall, by allowing users to configure the executions according to transformations on model and trace level, it is possible to capture the evolution (flexibility) of business process, as well as intentionally building non-compliance into the log. However, the approach does not guarantee that all possible ways of violating specific security properties (i.e. changing the shape of a trace accordingly) are reflected by the synthesized log. Ongoing work investigates how to apply testing methods – especially mutation testing [JH11] – as a foundation for synthesis.

*Paper structure.* Section 2 introduces the overall approach, whereas the main building blocks are presented in detail in Section 3. Section 4 presents a subset of operators for model transformation and property enforcement/violation, thereby defining the aforementioned “attacker model”. The tool `SecSy` is described and high-level evaluated in Section 5.

## 2 Simulation Approach

The overall approach is depicted in Fig. 1. It takes a series of business process specifications as input and generates a process log that contains traces of these specifications. Based upon the security and compliance requirements, deviations from

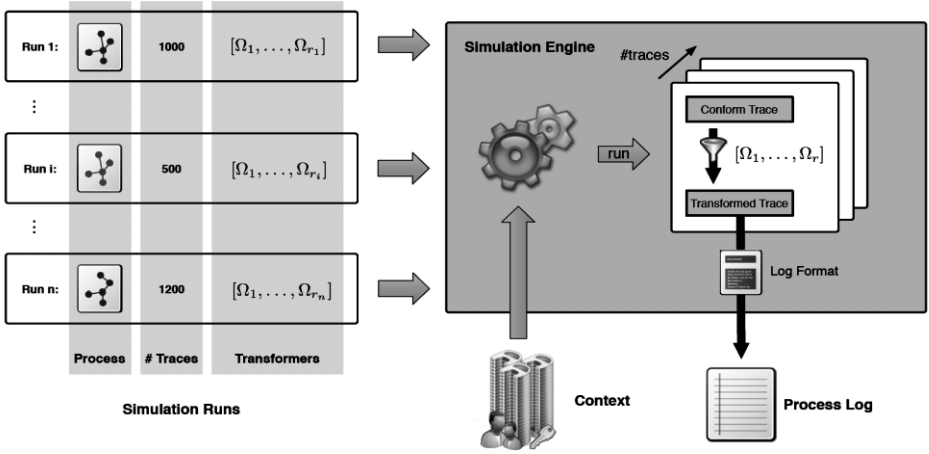


Figure 1: Overview of the approach

the defined control flow are generated with the help of *transformers* which encode specific trace properties that are either enforced or violated on a random basis. Together with the ability to subsequently generate traces of a series of different business process models into a single log, the approach provides for a definition of an attacker model, which can be configured in a detailed way for a particular test case. Generated logs are outputted in MXML (a format for process mining), as well as plain text. Ongoing work adds support to Extended Event Streams format ( XES ).

Processes subject to simulation are considered to be executed within a *context*. While the control flow of a process defines possible execution traces, subjects authorized to execute process activities and objects used by activities are defined by contexts. Our approach allows to define *simulation runs*, each relating to one process model which are processed one after another according to the number of desired traces for each run. This way, the engine is capable of simulating situations in which there is an initial model for planned process behavior, but a variant of this model was executed for some time, possibly due to the activity of an attacker or process variation/flexibility. To support the user in defining such variants, we define model transformation operators that replace specific control flow patterns of a process according to Weber et al. [RW12] (see Section 4 below). Additionally, each simulation run can relate to a set of *trace transformers* which operate on trace level.

During the processing of a simulation run, the engine generates valid log traces according to the control flow of the corresponding process and context and then passes them through the trace transformers which apply transformations in a post-processing manner. Trace transformers can remove or add process activities (simulating skipped activities or incomplete logging), as well as change information within traces in a way business related properties like *separation/binding of duties* or authorization constraints are enforced or violated.

### 3 Process Specifications and Execution Contexts

The main building blocks are, firstly, the process specification which defines activities and the order in which they can be executed and, secondly, the execution context which defines the organizational environment for process execution in terms of subjects authorized to execute process activities and objects, used by activities. For simulation purposes, a process specification and a compatible context are paired together to an execution system.

**Process specification** The process specification defines process activities and the order in which they are executed. To focus on the simulation procedure, we abstract from concrete specification meta-models at this point. The only prerequisite is that processes are well-structured [VVK09]: a *well-structured* process model is a model composed of blocks, which may be nested but must not overlap. A block can be a single activity, a sequence, a parallel/conditional branching and a loop block.

A process is defined as a tuple  $P = (A_P, T_P)$ , where  $A_P$  is a set of activities and  $T_P \subseteq A_P^*$  is the set of possible traces over  $A_P$ . It is assumed that the control flow of a process which defines  $T_P$  by establishing a successor-relation among activities is modeled with the help of OR-, XOR- or AND-gateways (implicitly or explicitly). These gateways are logical connectors modeling choice, combination and parallelism of activities in the conventional way. In case a process contains more than one gateway of the same kind, they are enumerated in a natural way (e.g.  $OR_1, OR_2$ ). Such process models (specifications) are instantiated when processes are executed. Each instance relates to a possible execution sequence of process activities (trace). The set of all processes is denoted as  $\mathbb{P}$ .

Activities are executed by subjects at a specific point in time and involve a number of objects. Subjects, objects and clearances (rights of subjects to execute activities) are defined within process contexts  $C = (A_C, S_C, O_C, \mathcal{A}_C, \mathcal{O}_C)$ , where  $A_C$  is a set of activities,  $S_C$  is a set of subjects and  $O_C$  is a set of objects. The function  $\mathcal{A}_C: S \rightarrow \mathcal{P}(A_C)$  defines the set of activities a subject is authorized to execute and the function  $\mathcal{O}_C: A \rightarrow \mathcal{P}(O_C)$  assigns the objects employed by activities.

The execution of process instances leads to the occurrence of events, each reporting the execution of a process activity. Formally, an event is a triple  $E = (Z_E, A_E, S_E)$ , where  $Z_E \in \mathbb{N}$  is a unique timestamp,  $A_E \in A_P$  is an activity and  $S_E \in S_C$  is the executing subject. For convenience, we use  $e.time$  to access the timestamp of an event,  $e.activity$  and  $e.subject$  respectively. In contrast to process traces that just consider activity orders, log traces denote the sum of all events relating to the execution of a process instance. A log trace is defined as sequence of events  $t = \langle e_1, \dots, e_n \rangle$ , where the sequence order is defined according to the event timestamps  $e_i.time$ . The set representation of a log trace is defined as  $t_E$ . Log trace  $t$  is a sub-trace of another log trace  $t' = \langle e'_1, \dots, e'_n \rangle$ , denoted  $t \sqsubseteq t'$ , iff  $t_E \subseteq t'_E$ . We use the notation  $e^{\leftarrow t}$  for the prefix of  $t$  before the occurrence of  $e$  and  $e^{\rightarrow t}$  for the suffix after the occurrence of  $e$ , whereas  $e^{\preceq t}$  denotes the direct predecessor of  $e$  within  $t$ ,  $e^{\succeq t}$  the direct successor. The set of all log traces that can be formed based on activities  $A_T$  and subjects  $S_T$  is denoted

as  $\mathbb{T}_{A_T, S_T}$ , for events  $\mathbb{E}_{A_T, S_T}$  respectively. We use the notation  $t[i]$  to access the  $i$ -th event of a log trace  $t$ . The activity reduction of a log trace is defined as  $t|_A = \langle a_1, \dots, a_n \rangle, a_i = e_i.activity$ . The function  $\theta: \mathcal{P}(\mathbb{E}_{A_T, S_T}) \times A_T \rightarrow \mathcal{P}(\mathbb{E}_{A_T, S_T})$  determines the reduction of a set of events to those with a specific activity, i.e.  $\theta(\{e_1, \dots, e_n\}, a) = \{e_i | e_i.activity = a\}$ .

**Execution System** An execution system is a pair  $(P, C)$ , where  $P = (A_P, T_P)$  is a process and  $C = (A_C \supseteq A_P, S_C, O_C, \mathcal{A}_C, \mathcal{O}_C)$  is a process context. A possible trace of an execution system is denoted  $t_{(P, C)} \in \mathbb{T}_{A_P, S_C} \subseteq \mathbb{T}_{A_C, S_C}$  and contains only events relating to activities in  $P$  and subjects in  $C$ , i.e.  $\forall (z, a, s) \in t_{(P, C)}: a \in A_P \wedge s \in S_C$ . We use  $\mathbb{T}_{(P, C)}$  for the set of all possible traces of an execution system. A possible trace  $t'$  of an execution system is *conform*, if it is a valid trace of the corresponding process and subjects are assigned according to the authorization function  $\mathcal{A}_C$ , i.e.  $t'|_A \in T_P \wedge t'.activity \in \mathcal{A}_C(t'.subject)$ . The set of all conform traces of an execution system is denoted as  $\mathbb{T}_{(P, C)}^C \subseteq \mathbb{T}_{(P, C)}$ . A process log resulting from executing activities of process instances with respect to an execution system is defined as a sequence of traces  $L_{(P, C)} = \langle t_0, t_1, \dots, t_n \rangle, t_i \in \mathbb{T}_{(P, C)}$ , where the ordering of sequence elements is defined by case starting times, i.e.  $t_i[1].time$ .

## 4 Model- and Trace-Transformations

Model transformations regard only the control flow (structure) of a process and change the order in which activities can be executed and are determined with the help of transformation functions  $\Theta_T: \mathbb{P} \rightarrow \mathbb{P}$ . Currently, the following transformation operators at gateway-level are considered:

- $\Theta_{\text{And2Xor}}$ : This transformation provides a variation of the input process in which a specific AND-gateway is rewritten to an XOR-gateway. This means: while the original process requires traversing both parts of the AND-gate, the variant assumes, that only one (randomly chosen) path is traversed.
- $\Theta_{\text{Xor2And}}$ : The opposite of  $\Theta_{\text{And2Xor}}$ , i.e. a specific XOR-gateway is rewritten to an AND-gateway.
- $\Theta_{\text{Swap}}$ : This transformation provides for variants in which the order of a pair of activities is swapped.

In contrast to changing the behavior of a process persistently by modifying its structure, trace transformers are used to transform traces in a way specific trace properties are enforced or violated. Transformers operate on valid traces generated during the simulation process in a post-processing manner, before they are added to the output log file. Formally, a transformer wrt an execution system  $(P, C) = ((A_P, T_P), (A_C, S_C, O_C, \mathcal{A}_C, \mathcal{O}_C))$  is a function  $\Omega_F: \mathbb{T}_{(P, C)}^C \times I_1 \times \dots \times I_k \rightarrow T_F \subseteq \mathbb{T}_{A_P, S_C}$  which takes a valid process trace as input and generates a modified version of this trace according to a number of parameters  $I$ , which is not necessarily valid. The set of possible events to apply the transformer is denoted  $E^S$ . The set of events, for which the

transformer is actually applied is denoted  $E_\pi^S \subseteq E^S$ . The modified events resulting from applying the transformer is denoted  $E'^S_\pi$ . Below we define a number of transformers that can be applied while simulating processes. It should be noted that in situations where more than one transformer is applied on a trace, the fields modified by the transformers have to be “locked”. This is to prevent further transformers from corrupting the already enforced properties coming from previous applications of transformers.

Currently, the approach comprises the following six trace transformers: (a) *delay* inserts a delay in the process execution; (b) *skip* generates executions in which some activities are skipped; (c) *silent* captures a situation in which a particular activity has not been logged onto the file; (d) *authentication* mimics an access control policy and its violation; (e) *binding of duty* (BOD) mimics the compliance with or the violation of a binding of duty requirement; and (f) *separation of duties* can be seen as the opposite of BOD: it states that a particular activity must be performed by a *different* subject or role. Due to space constraints, below we present only three transformers – namely, delay, authentication and separation of duties. The definition of the other transformers occur analogously. It is important to emphasize, though, that this list of transformers can be extended with or refined to other domain-specific transformers and, subsequently, added to SecSy.

**Delay transformer.** The delay transformer inserts a delay in the process execution. Formally, the transformer is defined as  $\Omega_{\text{delay}}: \mathbb{T}_{(P,C)}^C \times A \subseteq A_P \times \mathbb{N} \times \mathbb{R} \times \mathbb{N} \rightarrow \mathbb{T}_{(P,C)}^C$ . The transformer  $\Omega_{\text{delay}}(t, a, \delta, \lambda, \omega)$  transforms an input trace  $t = \langle e_1, \dots, e_n \rangle$  into an output trace  $t' = \langle e'_1, \dots, e'_n \rangle$  by adding a random delay  $\varepsilon \in [\delta - \lambda; \delta + \lambda]$  to events relating to activity  $a$ . The number of events which are delayed is bounded by the parameter  $\omega$ , whereas  $E^S = \{e \in t \mid e.\text{activity} = a\}$  and  $0 \leq |E_\pi^S| \leq \min(\omega, |E^S|)$ . Note that  $E^S$  can be empty. All events  $e'_i$  of the generated output trace fulfill the following properties:

- $e'_i.\text{activity} = e_i.\text{activity}$
- $e'_i.\text{subject} = e_i.\text{subject}$
- $e'_i.\text{time} = \begin{cases} e_i.\text{time} + \sum_{e_j \in e_i^{\leftarrow t}} (e'_j.\text{time} - e_j.\text{time}) + \varepsilon \text{in}[\delta - \lambda; \delta + \lambda] & \text{iff } e_i \in E_\pi^S \\ e_i.\text{time} + \sum_{e_j \in e_i^{\leftarrow t}} (e'_j.\text{time} - e_j.\text{time}) & \text{otherwise} \end{cases}$

**Authentication transformer.** The authentication transformer mimics an access control policy and its violation. Formally, the transformer is defined as  $\Omega_{\text{auth}}: \mathbb{T}_{(P,C)}^C \times \mathbb{N} \rightarrow \mathbb{T}_{(P,C)}$ . The transformer  $\Omega_{\text{auth}}(t, \omega)$  transforms an input trace  $t = \langle e_1, \dots, e_n \rangle$  into an output trace  $t' = \langle e'_1, \dots, e'_n \rangle$  by randomly assigning subjects to events, which are not authorized to execute the event activity. The number of events for which the transformer is applied is bounded by the parameter  $\omega$ , whereas  $E^S = t_E$  and  $0 \leq |E_\pi^S| \leq \min(\omega, |E^S|)$ . Note that the transformer fails when  $E_\pi^S$  is empty and subjects cannot be assigned accordingly (e.g. when all subjects are allowed to execute all activities). All events  $e'_i$  of the generated output trace fulfill the following properties:

- $e'_i.activity = e_i.activity$
- $e'_i.time = e_i.time$
- $e'_i.subject \in S_C \setminus \mathcal{A}_C(e_i.activity)$  iff  $e' \in E_{\pi'}^S$ ,
- $e'_i.subject = e_i.subject$ , otherwise

**Separation-of-duties transformer.** The separation-of-duties ( SOD ) can be seen as the opposite of the aforementioned BOD transformer: it states that each activity in an activity set must be performed by a *different* subject or role. The transformer is formally defined as  $\Omega_{\text{SOD}}: \mathbb{T}_{(P,C)}^C \times \mathcal{P}(A_P) \times \mathbb{N} \times \{\text{enforce}, \text{violate}\} \rightarrow \mathbb{T}_{(P,C)}^C$ . The transformer  $\Omega_{\text{SOD}}(t, A, \lambda)$  transforms an input trace  $t = \langle e_1, \dots, e_n \rangle$  into an output trace  $t' = \langle e'_1, \dots, e'_n \rangle$  by enforcing or violating the *separation of duties* security property on trace  $t$ . Formally, the SOD property on a set of activities  $M$ , whereas  $S: M \rightarrow S_C$  denotes the executing subject is defined as  $|\bigcup_{m \in M} S(m)| = |M|$ . In the *enforcement* case, applied on a given trace  $t$  and a set of *bindings*  $A = \{a_1, \dots, a_n\}$ , this translates to:

- $E^S = \{e \in t \mid e.activity \in A\}$
- $|E_{\pi'}^S| \stackrel{!}{=} |E^S|$
- $A_{\phi} = \{(s_1, \dots, s_n) \mid s_i \in \bigcap_{e \in \theta(E_{\pi'}^S, a_i)} \mathcal{A}_C(e)\}, s_i \neq s_j$
- $A_{\phi} \neq \emptyset$
- $\gamma = (\gamma_1, \dots, \gamma_n) \in A_{\phi}$  (conform subject configuration)
- $\forall a_i \in A \ \forall e' \in \theta(E_{\pi'}^S, a_i): e'.subject = \gamma_i$
- $\psi: A \leftrightarrow S_C, \ \psi(a_i) \mapsto \gamma_i$

The enforcement fails, if there is no conform subject assignment which enforces the property, i.e.  $A_{\phi} = \emptyset$ . All events  $e'_i$  of the generated output trace fulfill the following:

- $e'_i.activity = e_i.activity$
- $e'_i.time = e_i.time$
- $e'_i.subject = \begin{cases} \psi(e'.activity) & , e' \in E_{\pi'}^S \\ e_i.subject & , \text{otherwise} \end{cases}$

In the *violation* case, the transformer ensures that there are events with activities in  $A$  with the same subject: it first enforces the SOD property as stated above and then randomly chooses a set of events  $V = \{v_1, \dots, v_u\} \subseteq E_{\pi'}^S$  for which it assigns the same authorized subject, i.e.  $v_1.subject = \dots = v_u.subject$ . The violation fails if there is no conform subject assignment which violates the property. This procedure allows for a more realistic simulation of SOD violations than a simple random choice of different authorized subjects.

## 5 SecSy: Prototypical Realization and Initial Evaluation

The event log synthesis approach based upon the framework presented in Section 3-4 has been implemented as a standalone, extensible Java application, called SecSy that can be downloaded at <http://bit.ly/SecSy>). Figure 2 depicts the configuration panel of the application. The simulation settings require the path for the generated log file and its name, as well as the simulation type ( SIMPLE for traces containing only timestamps and activity names and EXTENDED for traces containing also information about executing subjects and data items) and the output log format (plain text or MXML ). The latter is a standard input for process mining technologies, thereby serving as a basis for generating logs to test with process mining tools.

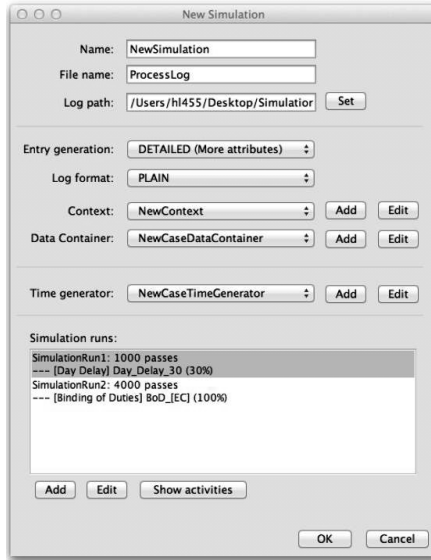


Figure 2: Screenshot of the configuration panel

The tool allows for flexible configuration of all required parameters and the creation and editing of corresponding components (Fig. 3). The time generator component contains all timing related simulation properties, including the start time for the simulation (date for first trace), the number of cases per day, office-days and -hours, as well as individual activity durations and delays between succeeding activities. All properties can be randomized by adding deviation bounds. The possibility to add several simulation runs is useful for simulating the execution of different process models or different property enforcements along time. To model the requirements that should (not) hold for single simulation runs (thereby specifying the policy applicable to the process), the user can assign different transformer configurations.

This way, the frequency and kind of security violations can be thoroughly configured. In the case of simulation type `EXTENDED`, the user has to specify a context and a data container. Data containers generate values for data items used during process execution (e.g. credit amount) and store the values until a trace is completed. This way, the consistent usage of attribute values along a complete trace is ensured. A context holds subjects and their permissions to execute activities and access data elements together with activity data usage (attributes used by activities) information. To specify subject permissions, the user can choose between an access control list or a role based concept, which is particularly helpful for large contexts. Additionally, a context allows to specify constraints on attribute values that can be added to process activities. This influences the generation of process traces in a sense, that only those activities are executed whose constraints are fulfilled. On Petri net level (which is the used meta-model for process models), the set of enabled transitions is reduced according to existing constraints, when the simulation tool decides about the next transition to fire.

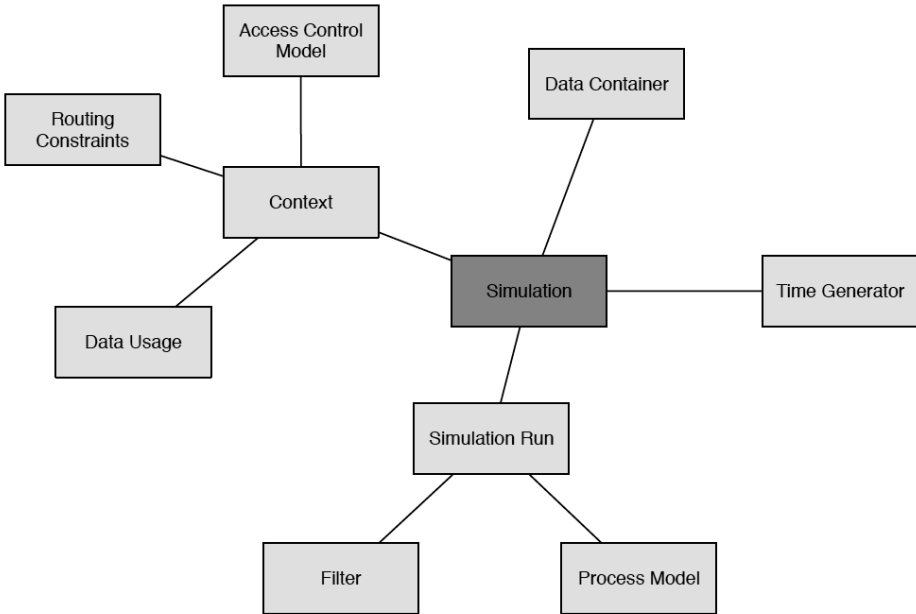


Figure 3: Configuration of simulation runs

**Evaluation.** We have carried out both a *quantitative* and *qualitative* evaluation of SecSy. The quantitative evaluation regards the overall performance of SecSy, which is acceptable for synthesizing large event logs. As an example, a process log containing 100K traces of a non-trivial loan application process with 19 activities incorporating 50 subjects organized in 4 different roles takes 24 seconds including the enforcement of a binding of duties constraint on a pair of process activities (3GHz Intel Core i7, 8GB DDR3 memory under OSX 10.8). For comparison, the generation of 10K traces in the same configuration takes 3 seconds. Future work will further qualitatively evaluate

SecSy in other dimensions. For instance, the interplay between the output format chosen (plain text or MXML ) and the performance and limits to the generation of event log files has to be clarified. Such an evidence is necessary for the generation of “big log files”. Further, detailed sensitivity analysis of different context parameters (e.g. number of subjects and roles) have to be carried out.

The qualitative evaluation aims at showing that the logs produced with SecSy in fact correspond to the original configuration setup for model- and trace-transformations. Put another way, checking whether the transformations are effective. For this, we generate log files employing, for each, the same context configuration but different transformers. Further, we use PROM [vdA11, Chap.~10] to check whether these parameters are fulfilled or not. For example, to validate the effectiveness of the SOD transformer, we first configure the transformer so that, say, 20 percent of the synthesized process runs violate this requirement: subject *S* performs both activities *X* and *Y*. Subsequently, we employ the LTL -checker plug-in to isolate the cases in which this requirement is violated. For this, it is sufficient to use the predefined formula “*exists person doing task X and Y*”. Similar checks employing other techniques – e.g. conformance checking and process discovery – were carried out for all the transformations, so that we could validate their effectiveness. Clearly, this is an ad-hoc, empirical validation style based upon the assumption that process models respect the designated structural constraints. Further work aims at obtaining well-founded evidence that transformations are effective.

## 6 Related Work

Business process simulation is an important tool in business process management to assessing the runtime operation of a business process. Simulation can also be used as a means to synthesize traces to test analysis techniques and tools, e.g. conformance checking and process mining. In both cases, it is important to insert deviations from the prescribed behavior in order to achieve variability (c.f. [vdA10a, vdA10b] and [DRMR13, Chap. 7.4]).

However, the current state of the art consider either variations on the execution time and workflow of the process or control flow deviations mimicking process flexibility and dynamics. Nakatumba et al. [NWvdA12] present an approach to generating event logs with workload-dependent speeds from simulation models. Mans et al. [MRvdA+10] employ simulation to analyze the impact of a schedule-aware workflow management system. Schonenberg et al. [SJSvdA10] devise a simulation approach for the analysis of business trends. Rozinat et al. [RMSvdA09] proposes an approach to discovering simulation models. These works consider, taking our transformers, delays and workload changes, but they do not consider security and compliance policies. Other works propose business process simulation frameworks (e.g. [CKH98, Tum96]) without actually testing them. Weber et al. [WPZW11, Ala] propose an simulation approach and tool support that considers flexible business processes. However, also here one cannot configure policy violations.

Several tools exist to simulate business processes (see [JVN06] for survey up to 2006). Bahrami [BSB98] present a special purpose tool developed at IBM. The focus of the tool

was on the enterprise architecture and not on processes. Burattin and Sperduti [BS10] present a framework for the generation of business process models and their execution logs. However, the tool does not offer any control over the generated models and event logs. The AristaFlow BPM Suite provides support for simulation, however it is unclear whether compliance and security rules can be considered [RW12, Chap.~15]. The CPNTools toolset provides for a highly configurable simulation environment. However, the toolset is not designed exclusively for this purpose, requiring the programming or definition of the aforementioned transformers. Our tool provides them in a natively and extensible manner.

Overall, the generation of “defect” data has been applied to software process improvement in general [RTVA12], but has never been seen in the BPM area. This paper is a first step in this direction. We firmly believe that the controlled, push-button generation of (large) test data is a promising research direction and application domain in business process testing and improvement. Not only for testing mechanisms in the context of security and compliance (as argued in this paper), but also for resilience [KS11, WL07], visualization [KRM12] and dependability [WLR+09] engineering for business process and process-aware information systems.

## 7 Summary and Further Work

This paper outlined a method for simulating business process flexibility and noncompliance, thereby synthesizing event logs. Overall, this approach is novel, both in the security, compliance and business process management communities. It allows for highly-controllable simulation output incorporating conform process behavior with respect to targeted security properties, as well as random failures that may lead to the violation of selected properties and incomplete/distorted process logs. The approach has been realized in a prototype and the results obtained with it are promising and useful in practice.

We firmly believe that the automated generation of such event logs is a necessary step toward the development of tools for business process analysis. This not only applies to standard process-aware information systems. More importantly, with the advent of big data and business process management “in the large”, there will be the need to test with such log data. Today, generating these data is very tedious. Our proposal should, in the future, be also used for this setting, thereby considering process architectures (instead of single models) and event-intensive simulation.

Besides the ongoing and future works mentioned above, there are several other issues that need to be added to obtain a fully-fledged log synthesizing tool. *Firstly*, the simultaneous execution of parallel process models, thereby allowing for the trace interleaving. *Secondly*, the completeness of generated logs has not been investigated, i.e. are all the possible cases and policy violations considered while generating the log. *Thirdly*, soundness should be guaranteed by construction, but will be investigated formally in the future. *Finally*, we plan to extend the set of policies. In particular, we will employ the policy patterns provided by Dwyer et al. [DAC99] to provide for policy templates and allow their specification before execution.

## References

- [Acc13] Rafael Accorsi. Sicherheit im Prozessmanagement. *digma Zeitschrift für Datenrecht und Informationssicherheit*, 2013.
- [AL12] Rafael Accorsi and Andreas Lehmann. Automatic Information Flow Analysis of Business Process Models. In *Conference on Business Process Management*, volume 7481 of *Lecture Notes in Computer Science*, pages 172–187. Springer, 2012.
- [Ala] Alaska Simulator. <http://www.alaskasimulator.org>, Last seen: June 2, 2013.
- [AS08] Rafael Accorsi and Thomas Stocker. Automated Privacy Audits Based on Pruning of Log Data. In *EDOC International Workshop on Security and Privacy in Enterprise Computing*. IEEE, 2008.
- [AS12] Rafael Accorsi and Thomas Stocker. On the Exploitation of Process Mining for Security Audits: The Conformance Checking Case. In *ACM Symposium on Applied Computing*, pages 1709–1716. ACM Press, 2012.
- [AS13] Rafael Accorsi and Thomas Stocker. On the Exploitation of Process Mining for Security Audits: The Process Discovery Case. In *ACM Symposium on Applied Computing*. ACM Press, 2013.
- [ASK08] Rafael Accorsi, Yoshinori Sato, and Satoshi Kai. Compliance Monitor for Early Warning Risk Determination. *Wirtschaftsinformatik*, 50(5):375–382, 2008.
- [AW10a] Rafael Accorsi and Claus Wonnemann. Auditing Workflow Executions against Dataflow Policies. In *Business Information Systems*, volume 47 of *Lecture Notes in Business Information Processing*, pages 207–217. Springer, 2010.
- [AW10b] Rafael Accorsi and Claus Wonnemann. Static Information Flow Analysis of Workflow Models. In *Conference on Business Process and Service Computing*, volume 147 of *Lecture Notes in Informatics*, pages 194–205. GI, 2010.
- [AW11] Rafael Accorsi and Claus Wonnemann. Strong Non-Leak Guarantees for Workflow Models. In *ACM Symposium on Applied Computing*, pages 308–314. ACM, 2011.
- [BAS09] Travis Breaux, Annie Antón, and Eugene Spafford. A distributed requirements management framework for legal compliance and accountability. *Computers & Security*, 28(1-2):8–17, 2009.
- [BS10] Andrea Burattin and Alessandro Sperduti. PLG: A Framework for the Generation of Business Process Models and Their Execution Logs. In *Business Process Management Workshops*, volume 66 of *Lecture Notes in Business Information Processing*, pages 214–219. Springer, 2010.
- [BSB98] Ali Bahrami, Deborah A. Sadowski, and Soheila Bahrami. Enterprise Architecture for Business Process Simulation. In *Winter Simulation Conference*, pages 1409–1414, 1998.
- [Cer01] Iliano Cervesato. The Dolev-Yao Intruder is the Most Powerful Attacker. In *Annual Symposium on Logic in Computer Science*. IEEE Computer Society Press, 2001.
- [CKH98] Yoon Ho Cho, Je Keyong Kim, and Kim Soung Hie. Role-based approach to business process simulation modeling and analysis. *Computers & Industrial Engineering*, 35(1/2):343–346, 1998.
- [CPN] CPNTools. <http://cpntools.org/>.
- [DAC99] Matthew B. Dwyer, George S. Avrunin, and James C. Corbett. Patterns in Property Specifications for Finite-State Verification. In *Conference on Software Engineering*, pages 411–420. ACM, 1999.
- [DRMR13] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Fundamentals of Business Process Management*. Springer, 2013.
- [JH11] Yue Jia and Mark Harman. An Analysis and Survey of the Development of Mutation Testing. *IEEE Transactions on Software Engineering*, 37(5):649–678, 2011.
- [JVN06] Monique Jansen-Vullers and Mariska Netjes. Business process simulation – A tool survey. In *Workshop and Tutorial on Practical Use of Coloured Petri Nets and the*

*CPN Tools*, 2006.

- [KR01] Konstantin Knorr and Susanne Röhrig. Security Requirements of E-Business Processes. In *IFIP Conference on E-Commerce, E-Business, E-Government*, volume 202 of *IFIP Conference Proceedings*, pages 73–86. Kluwer, 2001.
- [KRM12] Simone Kriglstein and Stefanie Rinderle-Ma. Change Visualizations in Business Processes - Requirements Analysis. In *Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*, pages 584–593. SciTePress, 2012.
- [KS11] Thomas Koslowski and Jens Strüker. ERP On Demand Platform - Complementary Effects Using the Example of a Sustainability Benchmarking Service. *Business & Information Systems Engineering*, 3(6):359–367, 2011.
- [LA11] Lutz Louis and Rafael Accorsi. Vulnerability Analysis in SOA-Based Business Processes. *IEEE Transactions on Service Computing*, 4(3):230–242, August 2011.
- [MRvdA+10] Ronny Mans, Nick C. Russell, Wil M. P. van der Aalst, Piet J. M. Bakker, and Arnold J. Moleman. Simulation to Analyze the Impact of a Schedule-aware Workflow Management System. *Simulation*, 86(8-9):519–541, 2010.
- [NWvdA12] Joyce Nakatumba, Michael Westergaard, and Wil M. P. van der Aalst. Generating Event Logs with Workload-Dependent Speeds from Simulation Models. In *CAiSE Workshops*, volume 112 of *Lecture Notes in Business Information Processing*, pages 383–397. Springer, 2012.
- [RMSvdA09] Anne Rozinat, R. S. Mans, Minseok Song, and Wil M. P. van der Aalst. Discovering simulation models. *Inf. Syst.*, 34(3):305–327, 2009.
- [RTVA12] Anu Raninen, Tanja Toroi, Hannu Vainio, and Jarmo J. Ahonen. Defect Data Analysis as Input for Software Process Improvement. In *Conference on Product-Focused Software Process Improvement*, volume 7343 of *Lecture Notes in Computer Science*, pages 3–16. Springer, 2012.
- [RW12] Manfred Reichert and Barbara Weber. *Enabling Flexibility in Process-Aware Information Systems - Challenges, Methods, Technologies*. Springer, 2012.
- [SJSvdA10] Helen Schonenberg, Jingxian Jian, Natalia Sidorova, and Wil M. P. van der Aalst. Business Trend Analysis by Simulation. In *Conference on Advanced Information Systems Engineering*, volume 6051 of *Lecture Notes in Computer Science*, pages 515–529. Springer, 2010.
- [Tum96] Kerim Tumay. Business Process Simulation. In *Winter Simulation Conference*, pages 93–98, 1996.
- [vdA10a] Wil M. P. van der Aalst. Business Process Simulation. In *Handbook on Business Process Management*, volume 1, pages 313–338. Springer, 2010.
- [vdA10b] Wil M. P. van der Aalst. Business Process Simulation Revisited. In Joseph Barjis, editor, *Enterprise and Organizational Modeling and Simulation*, volume 63 of *Lecture Notes in Business Information Processing*, pages 1–14. Springer, 2010.
- [vdA11] Wil M. P. van der Aalst. *Process Mining – Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.
- [VVK09] Jussi Vanhatalo, Hagen Völzer, and Jana Koehler. The refined process structure tree. *Data Knowl. Eng.*, 68(9):793–818, 2009.
- [WL07] Qihua Wang and Ninghui Li. Satisfiability and Resiliency in Workflow Systems. In *Proceedings of the European Symposium On Research In Computer Security*, volume 4734 of *Lecture Notes in Computer Science*, pages 90–105. Springer, 2007.
- [WLR+09] Branimir Wetzstein, Philipp Leitner, Florian Rosenberg, Ivona Brandic, Schahram Dustdar, and Frank Leymann. Monitoring and Analyzing Influential Factors of Business Process Performance. In *IEEE International Enterprise Distributed Object Computing Conference*, pages 141–150. IEEE Computer Society, 2009.
- [WPZW11] Barbara Weber, Jakob Pinggera, Stefan Zugal, and Werner Wild. Alaska Simulator Toolset for Conducting Controlled Experiments on Process Flexibility. In *Information*

*Systems Evolution*, volume 72 of *Lecture Notes in Business Information Processing*, pages 205–221. Springer, 2011.

# Towards a Modelling Method in Support of Evaluating Information Systems Integration

Heiko Kattenstroth, Ulrich Frank, David Heise

Information Systems and Enterprise Modelling Research Group  
Institute for Computer Science and Business Information Systems (ICB)  
University of Duisburg-Essen, Universitaetsstr. 9, 45141 Essen, Germany  
{ heiko.kattenstroth | ulrich.frank | david.heise }@uni-due.de

**Abstract:** IT organisations face a number of challenges when assessing, realising, and maintaining a level of integration—between IT resources and between IT and business—that is satisfactory. To cope with these challenges, methods are required that support the evaluation of information systems integration by reducing the complexity inherent to the IT of today’s enterprises, by facilitating communication about integration matters among groups of stakeholders with differing perspectives, and by accounting for potential ambivalent effects of integration. In this paper, we investigate the potentials of an enterprise modelling-based method for IS integration evaluation and present initial conceptualisations—with a specific focus on data integration.

## 1 Introduction

Hardly any other characteristic of information systems (IS) is regarded as important as integration. The benefits of integration are undisputed: Integration helps with reducing redundancy and, hence, protecting system integrity; integrated information systems contribute to reuse, e.g., of data or system functions, and thus to a more efficient maintenance; and it is a prerequisite for aligning IT and business by providing information systems supporting a seamless support of the business processes [LZ11]. However, despite these prospects the integration of IS also comes with potential drawbacks: At first, there are the costs and risks that are related to the implementation of integrated systems. Since respective projects will usually require extensive modifications of an information system, many IT managers are reluctant to support them. Second, there are possibly negative effects of integration. Components that were designed for integrating well with a particular system may resist against reuse in another context. Tightly integrated components may compromise the adaptability of the IS and, thus, hamper the flexibility of the company’s IT landscape. It seems thus reasonable to assume that the *quality of information systems integration* is of pivotal relevance for productivity and quality of organisational work.

The challenges of integration can be purposefully addressed *ex ante* during the development phase of an information system, for instance, by accounting for integration on a conceptual level (e.g., by creating conceptual models) and on the implementation level, for instance, by using respective technologies such as DBMS or WFMS. However,

in many companies developing their information systems from scratch is no option nowadays. Instead, IT managers often have to deal with existing IT landscapes and (legacy) systems requiring ex post integration —potentially coming with remarkable costs for integration projects. High costs and the ambivalent effects of integration recommend a thorough course of action. As a result, IT management needs methods that support the analysis and assessment of IS integration in a differentiated manner; especially it is advisable to first evaluate the level (or quality) of integration featured by the existing systems.

In this context, various authors recognise the potential of using conceptual models of the enterprise—enterprise models—for integration matters. Concrete instruments are suggested, for instance, by Johannesson and Perjons, who propose design principles for the design and use of process models and associated data models to facilitate enterprise application integration [JP01]. An enterprise modelling method for supply chain integration based on CIMOSA is proposed by de la Fuente et al. [dlFRO10]. However, existing approaches mainly focus on selected aspects of integration only and lack a comprehensive conception of integration. As a consequence, they do not guide the assessment of integration or provide support for dealing with the ambivalent effects of integration [CLT12].

In this respect, the motivation of the paper is threefold: (1) We propose an elaborate conceptualisation of integration that, among others, allows distinguishing between different levels of integration, (2) investigate the potentials of an enterprise modelling-based method in support of evaluating information systems integration, and (3) present an outline of a corresponding method—in this paper with a particular focus on *data integration*—as a first step towards a comprehensive method for integration assessment and management.

Our research makes use of a research method configured for the epistemological particularities of research on modelling methods and applies a method for designing domain-specific modelling languages [Fra10]. Next, we present a conception of integration. Section 3 introduces requirements a method should satisfy. An outline of the solution and corresponding prospects are presented in Section 4. In Section 5, we discuss design decisions pertaining domain-specific modelling constructs and present a process model. Section 6 provides concluding remarks.

## 2 Conception of Integration

Realising integrated information systems requires an elaborate conception of integration that enables a differentiated assessment of its quality as well as of benefits and risks. Despite its long history in research on information systems, in particular in the German “Wirtschaftsinformatik” (see, e.g., [Ket09, Hei89, MH92, CLT12]), and its relevance in practice, there is still a noticeable lack of an agreed conceptualisation of integration and its various aspects that is differentiated and elaborate enough to allow for an assessment of IS integration. Integration is often used without a definition at all or by referring to the etymological interpretation based on the Latin *integrare* (“to make a whole”). While such a general understanding of integration as the result of constructing a whole from previously isolated parts makes sense, it is not sufficient for dealing with IS integration:

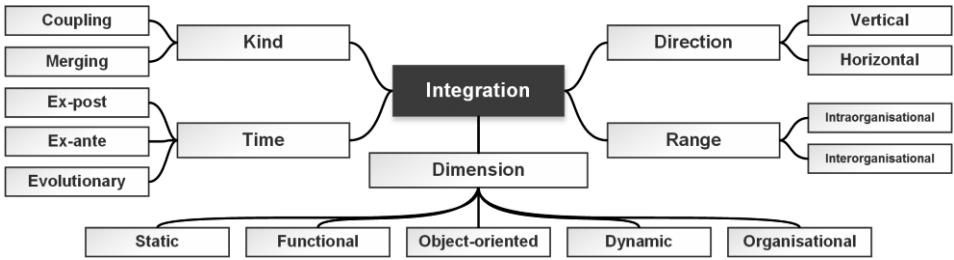


Figure1: Selected Aspects of Integration

First, it does not account for the specifics of information systems [Fra08]; second, it is not at all obvious what to consider the “whole” with regard to the integration of information systems, as this presupposes an unambiguous understanding and interpretation of the required integration; and, finally, such a conception does not allow to assess the quality of integration or to distinguish between different levels of integration. Only very few authors (e.g., [Jun06,Vog06]) present conceptions of integration that account for (selected) peculiarities of IS. However, their focus is different: The former focusses on architectures for data integration while the latter addresses integration based on “Enterprise Application Integration”.

Several proposals for the classification of integration-related phenomena exist (see, e.g., [CLT12] for an overview). Various criteria (e.g., time, direction, reach) allow for describing, for instance, different integration scenarios (e.g., M & A) and integration constructs (e.g., data integration). However, dimensions related to the quality of integration as well as corresponding risks and benefits are missing. Against this background we present our conception of integration, which consists of three parts that build upon one another: the conceptualization of integration itself, selected aspects of integration, and a proposal to distinguish between different levels of integration.

Information systems are linguistic artefacts. Therefore, integration is also mainly a *linguistic* conception, i.e., it is accomplished through language and communication respectively [Fra08]. Accordingly, the integration of two IT artefacts requires them to be able to communicate, either directly or through some kind of mediator. Communication in turn implies the existence of common concepts that define the semantics of the linguistic artefacts that are subject of communication relationships—in other words: Integration requires the existence of a *common semantic reference system*. Examples for such reference systems are data types or database schemas [Fra08].

We further elaborate our conception of integration in that it accounts for multiple aspects. Based on considerations from [Hei89,CLT12] we differentiate the concept of integration into *kind*, *time*, *direction*, *range*, and *dimension* of integration (cf. Fig.1). Although all aspects are relevant for describing IS integration, the *dimensions* are of particular importance—since they, among others, describe the use of a common semantic reference system: *Static* integration, also referred to as *data* integration [GWK92], is accomplished through shared static concepts. A typical example would be a common database schema used by a number of applications. *Functional* integration is

aimed at linking applications by using common functions. It requires static integration to allow for common interfaces. Common function libraries that are accessible throughout an entire—potentially distributed—platform are a typical example for functional integration. *Object-oriented* integration is a combination of static and functional integration. *Dynamic* integration is aimed at synchronising contributions of different applications to support a certain (business) process. It requires a common collaboration or process schema, which includes common event types and implies functional integration. Orthogonal to the dimensions discussed above and of particular relevance for information systems is *organisational* integration. The efficient use of information systems requires them to be integrated with the organisational action systems they are supposed to support—often referred to as “IT-Business Alignment”—which, in turn, requires common concepts. A method supporting an comprehensive and elaborate evaluation of IS integration should account for all of those dimensions and aspects. In this paper we particularly focus on data integration as a first step for development of such a method.

Integration is not a mere characteristic of information systems that exists or is missing. To assess the quality of integration we use the concept *level of integration* [Fra08]: The higher the level of semantics the common concepts include, the higher the level of integration they allow for. Note that this concept of semantics corresponds to the concept of information content<sup>1</sup>: The more possible interpretations are excluded by a concept, the higher its semantics. For example, the concept “Customer” is of higher semantics than the concept “String” (data type) [Fra08]. A high level of integration offers appealing benefits. The higher the level of semantics, the more focused and efficient information exchange will be. At the same time, a high level of semantics reduces the effort that is required for reconstructing the meaning of a message; hence, it also reduces the threat a message imposes to integrity [FS09]. However, integration also comes with possible negative effects: A high level of integration between two information systems may, for instance, compromise their adaptability. This, on the one hand, also impairs the potential range of reuse of the common concepts (i.e., they are too specific). On the other hand, a high frequency of changes of the common concepts increases the effort required for maintenance of the integrated systems. Accordingly, integration should not be treated as an end in itself but rather considered as an *ambivalent* concept [GWK92]. Further aspects to consider for assessing quality of integration are, for instance, how common concepts are accessed by the information systems (i.e., read-only access to redundant data reduces the risk for inconsistencies compared to write-access) and to what extent integration can improve action patterns in the organisational context (i.e., the higher the positive impact on the action patterns, the better the integration). Hence, assessing quality of integration specifically requires to account for the context of integration.

Based on this conception of integration we derive requirements for a method in support of evaluation IS integration in the next chapter.

---

<sup>1</sup> Note that the notion of semantics we use here is different from the one often used in Computer Science, where semantics denotes a feature of a formal system: Either there is semantics defined for it or not.

### 3 Requirements Analysis

The design of a modelling method requires the specification of corresponding language concepts, which, in turn, implies the need for analysing corresponding requirements.<sup>2</sup> We classify the requirements into three groups: (1) General requirements, (2) requirements related to the assessment of integration, which includes evaluating the quality of integration and accounting for its ambivalence, and (3) requirements addressing the realisation of integration. The underlying rationale is explained briefly along with each requirement.

**(1) General Requirements:** Integrating IT resources, e.g., networks, databases, applications, as well as integrating information systems with the surrounding action system confronts IT-organisations with the complexity of IT infrastructures of present day enterprises as well as of enterprises in general.

*Req. 1–Reduction of complexity:* The method should provide abstractions that allow for focussing on those concepts that are pivotal for analyses and application scenarios concerning integration.

Organisations that analyse information systems integration pursue different objectives and have differing settings (e.g., with regard to existing models of (parts of) the enterprise) as well as stakeholders who are involved.

*Req. 2–Adaptable process model:* The method should provide a process model that can be adapted for different application scenarios with regard to the level of detail and required effort as well as organisations with varying objectives.

Integration, for example realised in projects, requires the participation of stakeholders with different professional backgrounds and perspectives on integration matters including executives, process owners, and IT personnel, which may hamper communication and collaboration between these different groups of stakeholders.

*Req. 3–Support for multiple perspectives:* The method should provide perspectives specific to (groups of) stakeholders involved in integration projects. A perspective should, as far as possible, correspond with the abstractions, concepts and visual representations known and meaningful to the targeted stakeholders. All perspectives should be integrated with each other to foster cross-perspective communication and cooperation.

**(2) Integration Assessment:** Integration is often regarded as a core term, which does not need further explanations [Fra08]. However, such an arbitrary conception of integration does neither account for the peculiarities of information systems nor for the various phenomena related to integration.

*Req. 4–Conception of integration:* The method should provide an elaborate conception of integration that enables a differentiated assessment of benefits and risks. This also includes the various phenomena associated with integration and in particular the distinction between different dimensions of integration.

---

<sup>2</sup> Note that requirements for modelling IT landscapes in general are not discussed in this paper. Refer to, e.g., [Kir08] for more details.

As elucidated earlier, integration depends on communication, which can be realised through common concepts. These concepts can be part of a common semantic reference system.

*Req. 5–Common concepts:* Based on an elaborate conception of integration that supports a differentiated evaluation, the method should allow to represent the existence and lack of common concepts as well as to account for common semantic reference systems.

Evaluating the quality of integration, e.g., between two IS, is essential for finding room for improvement. The quality of integration depends on various factors; but first and foremost on the quality of the communication between the IS, i.e., the common concepts. Accordingly, the level of integration allows to assess the quality of integration (cf. Sec. 2).

*Req. 6–Level of integration:* To support the differentiated assessment of integration, the method should enable to distinguish different levels of integration and to identify related effects on, e.g., adaptability, economies, or data consistency and redundancy.

Frequent changes of structures shared between information systems compromise benefits realised by integration as it increases the effort required for maintenance of the integrated systems. If, for instance, several information systems are integrated and use the same database (i.e., share common concepts), changes of the database schema certainly lead to high effort adapting the depending information systems (i.e., a lower integration quality).

*Req. 7–Frequency of changes:* The method should include concepts that allow to express the (expected) frequency of changes of existing structures (i.e., their stability), for example of a database schema.

Integration assessment requires, on the one hand, a technical focus to analyse the IT infrastructure, on the other hand, assessing economic aspects of integration requires accounting for the context, i.e., the action system an information system is supposed to support.

*Req. 8–Organisational context:* The method should provide means to link integration initiatives to the surrounding organisational action system. This organisational context is provided by (at least) strategic goals, business processes and organisational structures.

Integration might have a direct (e.g., automation of a business process) and indirect (e.g., higher data quality) impact on the organisational context. However, as integration might also have negative effects, there is need to account for the ambivalence of integration.

*Req. 9–Ambivalence of integration:* The method should explicitly account for the ambivalence of integration: It should be possible to provide justifications for integration endeavours and the method should include concepts that support assessing the effect of integration (or the lack of it) on the business as the lack of integration will not always have the same economic impact. This includes to account for negative aspects of integration such as long-term costs and to reveal underlying assumptions to foster transparency. The method should also include concepts to express circumstances that inhibit integration.

**(3) Realising Integration:** Use of standards (e.g., from an industrial consortium) for realising integration seem to promising: They foster a wide range of reuse of proven (standardized) artefacts and particularly support cross-organisational integration. However, use of a standard might, at the same time, lack potential advantages provided by an enterprise-specific solution and might result in a lower level of integration and less benefits for the enterprise. Accordingly, standards promote and—at the same time—inhibit integration.

*Req. 10–Standards* The method should include concepts that allow to represent standards for integration. This should include, among others, the scope and purpose of the standards, their maturity and the requirements for using them. In the end, the method should allow for a critical assessment of the contribution of standards to integration.

Differentiating between the various technologies and approaches that are promoted in the context of integration is a challenge. They possess specific (dis-)advantages and should be selected carefully.

*Req. 11–Integration technologies:* The method should include concepts that allow to represent the various (types of) integration technologies and different integration architectures. On the one hand, this refers to defining enterprise-wide integration standards. On the other hand, there is need to the support project-specific selection and configuration of an integration approach.

The application of a method that addresses the identified requirements is outlined hereafter.

## 4 Outline of a Solution

With regard to the identified requirements, there is need for a conceptual foundation and corresponding methods. To satisfy this demand, it seems promising to build on an existing method for enterprise modelling for several reasons: (1) It serves to structure an enterprise by providing purposeful abstractions of IT and the surrounding action systems, arranged on different layers and corresponding to multiple perspectives (Req. 3) and, at the same time, reducing the complexity (Req. 1). (2) Accordingly, it provides stakeholders with specific and illustrative views on a company at various organisational levels, such as on value chains, business processes, or IT landscapes. (3) It makes use of domain specific modelling languages (DSMLs), which reconstruct a domain and its language as a *modelling language* and provide users with concepts they are familiar with. Accordingly, a DSML is intended to provide concepts on a higher level of domain-specific semantics. (4) The DSMLs are integrated and include concepts of the organisational and technical context, such as for goal, business process or organisational structure modelling as well as for modelling the IT organisation [Fra12]. This promises to account for the effects of integration on the organisational context, which is, among others, necessary to address its ambivalence as well as to identify strategies and goals that inhibit integration (Req. 8 & 9).

The prospects of extending an enterprise modelling method are illustrated in the application scenario in Figure **Fehler! Verweisquelle konnte nicht gefunden werden.**, which is based on the MEMO approach [Fra12]. Thus the present models are illustrated

presuming modelling languages and notations of MEMO, for instance, for strategic, organisational, and IT landscape modelling. It is, however, important to note that shown excerpts are not intended to predetermine a specific enterprise modelling approach; instead, they serve as an illustration of principle application and potentials of evaluating IS integration using enterprise models. The scenario shows:

- a *goal model* that represents selected goals of the enterprise (top left),
- various business process types as part of a *business process map* (top right),
- a *business process model* for the complaints management process,
- an excerpt of the *IT landscape* (third layer),
- and models of two selected *information systems* showing hardware and software used to realise the information systems (bottom layer).

The relationships between elements of the enterprise model are explicitly modelled using associations (e.g., between the CRM system and the processes it is used in). Additionally, the enterprise model is enriched by exemplary language concepts representing matters of integration: High-level *integration topics* describe and categorise the data managed by the information systems (e.g., customer data). Furthermore, *data exchange* relations as well as *similarities* and *redundancies* with respect to data the IS manage are displayed.

Given such an enterprise model exists, various analyses to improve integration can be supported. For this purpose, even further representations of the enterprise model such as a portfolio diagram or a matrix can be used. IT managers can analyse, for instance,

- which integration relations exist (by identifying data exchange relations);
- what is the current level of integration (by analysing the common concepts);
- where is potential for further integration and where does the lack of integration threatens data integrity (by analysing the associated business processes and the current data exchange relations as well as by identifying similar and/or redundant data);
- what is the (possibly negative) impact of integration on the organisational context (e.g., a goal) or—in other words—can the performance of the business processes and the achievement of the goals be improved by raising the level of (data) integration (by tracing the associations between the IS and the organisational context)?

In the application scenario (Figure 2), the analysis is aimed at improving the throughput of the “complaints management” process to account for the goal “increase throughput by 10%” (see (1) in Figure 2). For this purpose, existing integration relations between the information systems are identified (2). An assessment of the relation between the “Risk Management IS” and the Investment & Trader IS”, for instance, reveals a low level of integration due to the use of strings stored in flat files for data exchange (3). However, no direct effect of an improvement on the goals under considerations is expected. Integration topics associated with the CRM and the “Account and Deposit Mgmt. IS” (4) indicate that customer data is managed by both systems. This is supported by an analysis of the “use”-associations between the process model for the complains management process and the information systems. However, currently no integration exists. An assessment of integrating the customer data and the corresponding information

systems—based on the context they are used in—shows that it would have a mostly positive impact (5), which is indicated by the green “+” or red “-” next to the affected business processes and goals (6).

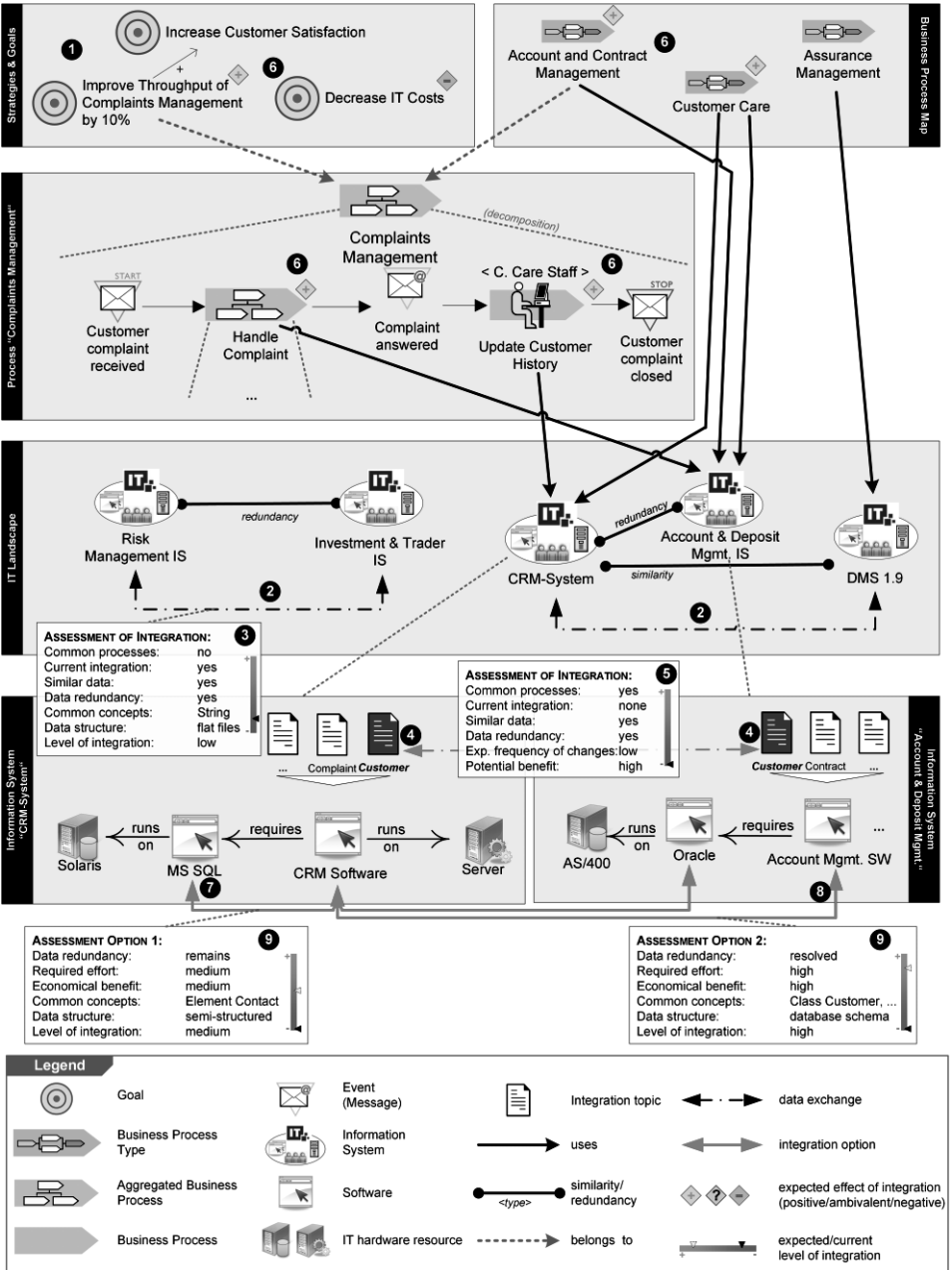


Figure 2: Exemplary enterprise model for integration analysis

Two general options are possible: Integration of the databases (7, e.g., periodic data exchange) or integration of the CRM and the account management IS based on a shared database (8). The latter option would allow for a higher level of integration: Using a common concept such as “Customer” in both systems is of higher semantics than exchanging fragmented data between the two databases. The former solution would be in most situations more flexible as it requires less effort to adapt it to new structures and requirements (9). However, to reduce hardware and software costs for the database server as well as to avoid data redundancy and potential inconsistencies, integration on software-level seems to be the more promising option—albeit initially a higher effort is required.

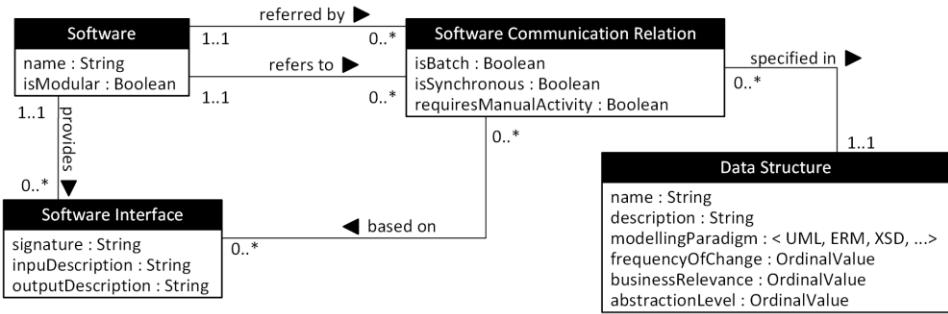


Figure 3: Representing communication relations (constraints & selected attributes omitted)

## 5 Elements of the Method

Based on the requirements analysis and the outlined potentials of an enterprise modelling-based approach, we present first considerations on language concepts and a process model.

### 5.1 Outline of a Metamodel—Selected Concepts

In this section, we present preliminary specifications of modelling constructs as meta model excerpts specified using the MEMO Meta Modelling Language [Fra11]. We assume a modelling infrastructure as described in Sec. 4. The reuse of modelling concepts from existing MEMO modelling languages is visualised by a coloured rectangle attached to the meta type header indicating the concept’s origin (as suggested in [Fra11]).

**Representation of communication relations:** A first basic modelling concept of the method is the representation of communication relations to depict, for example, dependencies between software and, in particular, how data is exchanged. For this purpose, *Software Communication Relation*, which can be based on *Software Interfaces*, allows to specify details related about communication relations. Of particular importance are details describing the need for manual activity (*requiresManualActivity*), whether the communication is (a)synchronous (*isSynchronous*), and if data is batched (*isBatch*) since these details might indicate room for improvement. Furthermore, *Data Structure* allows for a high-level description of the exchanged data (i.e., the common concepts, see Req.

5). It allows to express the *modellingParadigm* the data is specified with, the expected *frequencyOfChange* (cf. Req. 7), the *businessRelevance* of the data for the business as well as the *abstractionLevel* as an indicator for the level of integration (Req. 6) [Kir08].

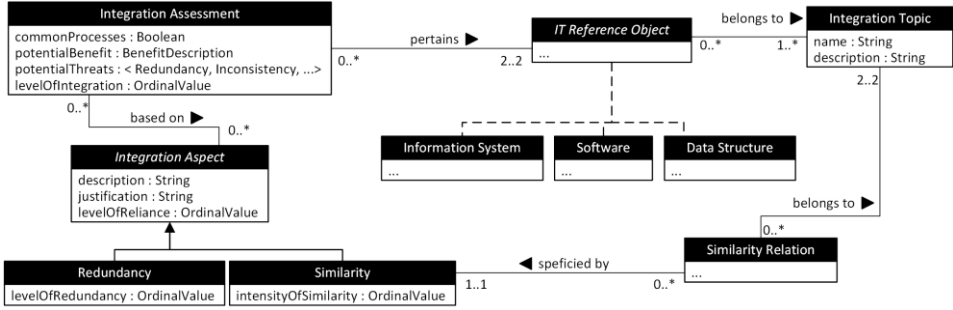


Figure 4: Representing integration assessment (constraints & selected attributes omitted)

**Assessment of integration:** At the core of the method for assessing IS integration are two metatypes: *Integration Assessment* and *Integration Topics*. *Integration Assessment* serves to describe the assessment of relations between two IT artefacts (represented in Fig. 1 as *IT Reference Object*, which acts as a surrogate, e.g., for software, information systems, and in particular data structures). The metatype allows to express whether the IT artefacts are used in identical processes (*commonProcesses*) and to assess the integration’s benefits and threats (*potentialBenefit* and *potentialThreats*). The *levelOfIntegration* (Req. 6) is based on identified similarities and redundancies of data between the artefacts, which allows to assess potential redundancies and inconsistencies. We propose to include corresponding concepts: *Similarity* allows to assess the intensity of similarity (i.e., related but not overlapping data), whereas *Redundancy* allows to describe the level of redundancy (i.e., data is stored redundantly). Furthermore, these concepts provide means for *justifications* and assessing the modeler’s *levelOfReliance* with respect to the modelled relation. If applicable, attributes of *data structure* (e.g., *businessRelevance* or *abstractionLevel*) provide further insights. The metatype *Integration Topic* allows for a high-level categorisation of IT artefacts—with regard to the focus of the paper based on the data managed by or exchanged between the IT reference objects (e.g., “Customer” or “Contract”). Accordingly, a new information system needs only be associated with corresponding integration topics and candidates for integration can be identified automatically based on identical or similar integration topics. Hence, integration topics should be stable over time. We propose a dedicated concept (*Similarity Relation*) for describing similarity relations between integration topics. The *Similarity Relation* in turn is specified by the aforementioned concept for *Similarity*.

**Accounting for context:** To account for the organisational context (Req. 8) and to differentiate between different types of utilisation (*utilizationType*) of data in business processes, we suggest a metatype *Utilization Relation* between *Process* and *Software*. The relation is further detailed by the *Data Structure* that serves to describe the data managed by the *Software*, affected by a *Process*, and related to an *Integration Topic*.

These metatypes present a first draft of basic language concepts as extension for an enterprise model-based approach in support of IS integration evaluation. In the next section, we briefly outline a process model that guides the application of these concepts.

## 5.2 Outline of a Process Model

The purpose of our approach is to systematically support the assessment of integration—with a particular focus on data integration. Creating and maintaining an elaborate enterprise model that covers, among others, strategies, goals, business processes, and the IT landscape might require too much effort or is out of scope (see Req. 2). Therefore, we promote an adaptable approach: Integration analysis can focus on the business processes that depend on the information systems (top-down) and/or the information systems themselves (bottom-up) using integration topics to describe data exchanged and managed by the IS. Accordingly, step 1 and 2 are optional and the level of detail can be adapted (Req. 2). Note, that the process model does not provide a “universal solution” for integration analysis, instead enterprise specific adaptations are necessary to account for stakeholders with, for instance, different qualifications as well as the available resources.

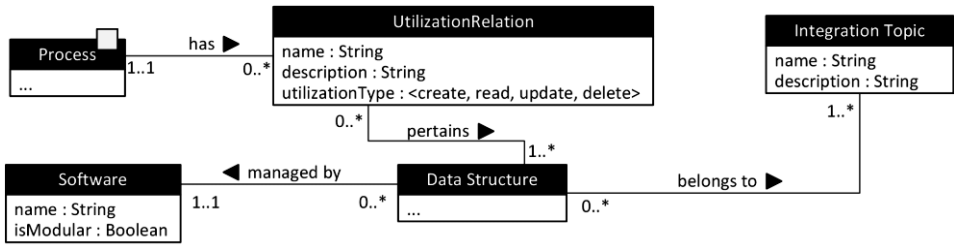


Figure 5 : Representing organisational context with regard to data utilisation (constraints & selected attributes omitted, *Process* is specified in MEMO-ORGML [Fra12])

**(1) Clarification of IT-strategy (optional):** For a long-term orientation, the (IT-) strategy needs to be clarified. This applies especially to aspects that might inhibit integration (e.g., a proposed carve-out) or demand integration (e.g., a new business process).

**(2) Modelling of relevant business processes (optional):** Information systems are predominantly intended to support business processes. Hence, requirements resulting from the business processes determine the required integration of data. To identify these integration requirements the (relevant) business process are modelled. For a first analysis, aggregated processes and a high level of abstraction respectively are sufficient.

**(3) Modelling of involved information systems:** The relevant parts of the application landscape are modelled, for instance, applications, databases, or data exchange connections. The models can either focus on information systems and abstract from the underlying details or include them, too. To allow for better analysis, descriptive relationships between the IT artefacts (e.g., similarities or commonalities) as well as relations between information systems and business processes or integration topics respectively are modelled using the corresponding metatypes.

**(4) Assessment of current state of integration:** The integration topics or associated business processes help to identify information systems that share commonalities or are related for further investigations. Therefore, the current state of integration between these information systems is assessed, among others with regard to the level of static integration, using the metatype *Integration Assessment*. For a more detailed analysis, similarities and commonalities as well as the pertained data structures and the level of integration they support can be analysed. Based on this analysis, the organisational context is considered to identify room for improvement. On the one hand, applications, for instance, that handle similar data (e.g., customer data) or are used in related business processes should potentially be integrated. Or, on the other hand, strategic considerations might require the integration of information systems to provide support for new processes.

**(5) Identification & assessment of integration options:** Options to realise the identified potentials for integration are assessed. This includes further aspects such as economic pressure for or against integration, strategic relevance of integration, required investments, related risks, life cycle of involved information systems, architecture standards of the enterprise, and regulatory restrictions that might prohibit integration of information systems. A survey and an assessment of integration technologies completes the analysis of the identified integration options. Based on the assessment of the identified integration options, projects are planned for an evolutionary path for changing the enterprise.

**(6) Project realisation:** The project is realised using the enterprise model as the central coordination mechanism.

## 6 Conclusions

This paper investigates the potentials of extending an enterprise modelling approach to support evaluation of IS integration—with a particular focus on data integration. The approach is based on the conclusion that enterprise models provide a substantial foundation for integration analyses in that they, e.g., represent both information technology and the organisational context (Req. 8), reduce complexity (Req. 1), and support multiple perspectives (Req. 3). Our contribution in this paper is threefold: First, we provide an elaborate conception for integration. Second, we provide results of a requirements analysis. Third, we illustrate the potentials of the intended method and discuss initial considerations pertaining modelling concepts and process model. The results presented in this paper are part of a larger research project and intended as a first foundation for discussion with and discursive evaluation by peers and domain experts. The scope of this paper was, therefore, limited to selected aspects of a method for IS integration evaluation: Application scenario and method elements (meta model and process model) focused on data integration and on the key basic language concepts; various requirements (e.g., Req. 10 & 11) remain, thus, unaddressed and will be topic of our next research activities. Accordingly, the paper provides first steps towards evaluation of IS integration. However, additional factors might be needed for a more comprehensive quality assessment of integration. This might, in the end, lead to revised language concepts and process model. Besides the discussed prospects an enterprise model-based approach for integration evaluation promises for further, more advanced

application scenarios, e.g.: First, enterprise models enriched with details about integration may be used to define “integration patterns” that provide proven solutions for recurring integration scenarios. Second, enterprise models covering integration aspects can be used as the foundation for advanced “dashboards” to support IT integration management. Third, information systems could describe themselves (e.g., need and potentials for integration, data they manage) by referring to IT models extended with business models—hence, enabling self-referential information systems [FS09].

## References

- [CLT12] Maximilian Chowanetz, Christine Legner, and Frederic Thiesse. Integration: An omitted Variable in Information Systems Research. In *Proceedings of the 20th European Conference on Information Systems (ECIS 2012)*, page 227, 2012.
- [dlFRO10] M. Victoria de la Fuente, Lorenzo Ros, and Angel Ortiz. Enterprise modelling methodology for forward and reverse supply chain flows integration. *Computers in Industry*, 61(7):702–710, 2010.
- [Fra08] Ulrich Frank. Integration - Reflections on a Pivotal Concept for Designing and Evaluating Information Systems. In *Information Systems and e-Business Technologies. 2nd Int. United Information Systems Conference (UNISCON 2008)*, pages 111–122. 2008.
- [Fra10] Ulrich Frank. Outline of a Method for Designing Domain-Specific Modelling Languages. ICB Research Report No 42, Universität Duisburg-Essen, Essen, 2010.
- [Fra11] Ulrich Frank. The MEMO Meta Modelling Language (MML) and Language Architecture. ICB Research Report No 43, Universität Duisburg-Essen, Essen, 2011.
- [Fra12] Ulrich Frank. Multi-perspective enterprise modeling: foundational concepts, prospects and future research challenges. *Software & Systems Modeling*, pages 1–22, 2012.
- [FS09] Ulrich Frank and Stefan Strecker. Beyond ERP Systems: An Outline of Self-Referential Enterprise Systems - Requirements, Conceptual Foundation and Design Options. ICB Research Report No 31, Universität Duisburg-Essen, Essen, 2009.
- [GWK92] Dale L. Goodhue, Michael D. Wybo, and Laurie J. Kirsch. The impact of data integration on the costs and benefits of information systems. *MISQ*, 16(3):293–311, 1992.
- [Hei89] Heidi Heilmann. Integration: Ein zentraler Begriff der Wirtschaftsinformatik im Wandel der Zeit. *HMD - Praxis der Wirtschaftsinformatik*, 26(150):45–58, 1989.
- [JP01] Paul Johannesson and Erik Perjons. Design principles for process modelling in enterprise application integration. *Information Systems*, 26(3):165–184, 2001.
- [Jun06] Reinhard Jung. *Architekturen zur Datenintegration: Gestaltungsempfehlungen auf der Basis fachkonzeptueller Anforderungen: zugl. Habilitationsschrift Universität St. Gallen*. Deutscher Universitäts-Verlag, Wiesbaden, 2006.
- [Ket09] Karl Heinz Kettner. The Integrative Effect of Electronic Data Processing: This paper was first published in elektronische datenverarbeitung 1(1)1959:14–19. *Business & Information Systems Engineering*, 1(1):84–88, 2009.
- [Kir08] Lutz Kirchner. *Eine Methode zur Unterstützung des IT-Managements im Rahmen der Unternehmensmodellierung: zugl.: Dissertation, Universität Duisburg-Essen, 2007*. Logos, 2008.
- [LZ11] Jerry N. Luftman and Hossein S. Zadeh. Key information technology and management issues 2010–11: an international study. *J. of Inf. Technology*, 26(3):193–204, 2011.
- [MH92] Peter Mertens and Jochen Holzner. WI - State Of The Art. Eine Gegenüberstellung von Integrationsansätzen der Wirtschaftsinformatik. *WIRTSCHAFTSINFORMATIK*, 34(1):5–25, 1992.

[Vog06] Petra Vogler. *Prozess- und Systemintegration: Evolutionäre Weiterentwicklung bestehender Informationssysteme mit Hilfe von Enterprise Application Integration: zugl. Habilitationsschrift Universität St. Gallen 2004*. DUV, Wiesbaden, 2006.



## **Process Modelling**



# Do We Need a Standard for EPC Modelling?

## The State of Syntactic, Semantic and Pragmatic Quality

Michael Fellmann, Sebastian Bittmann, Arne Karhof, Carl Stolze, Oliver Thomas

Information Management and Information Systems,

University Osnabrueck, Germany

{ michael.fellmann | sebastian.bittmann | akarhof | carl.stolze | oliver.thomas }@uni-osnabrueck.de

**Abstract:** The quality determination of business process models is a complex and demanding task. In literature, a plethora of different quality criteria can be identified and are respectively used by practitioners. The selection of quality criteria depends on the one hand on the respective preferences of the individual modeller. On the other hand, it is prescribed by the modelling language in use, as it has embodied specific criteria e.g. based on its syntax. For the widespread EPC, no comprehensive overview of existing aspects for the evaluation of an EPC model's syntactic, semantic and pragmatic quality exists. With this investigation, we present such an overview and put the identified aspects into a holistic perspective based on the identified and relevant literature.

## 1 Introduction

Process models are a widespread tool to design, represent, discuss and eventually manage chains of activities in all kinds of businesses – from small companies to global enterprises [HFL11]. Although many methods have been developed to model business processes [BNT10], there is no agreed standard to judge the quality of a process model. Having in mind, there is – besides the variety of methods – no shortage of modelling languages to choose from [TF09], quality criteria either need to be generic, respectively abstract or language-specific. Whilst generic frameworks for process model quality, such as the Generalities of Orderly Modelling [BRS95] or the SIQ-Framework [RMR10], provide general guidelines, they cannot reflect the particularities of specific languages with their individual syntax and semantics. Consequently, the need exists to evaluate particular modelling languages thus going beyond the general guidelines in order to propose implications to the model construction. Thus, in this paper, we focus on the widespread Event-driven Process Chain (EPC) to provide a comprehensive overview of aspects to evaluate the syntactic, semantic and pragmatic quality of models. The EPC is established in academia and practice alike for business process modelling [Fett09; MeRA10], thereby combining practical and theoretical relevance. In addition to this, the EPC language has been investigated for more than 20 years. However, despite its popularity and the large number of modelling tools available, the EPC has never been fully standardised. As no standard for the EPC is available, the quality judgement of the EPC and the resulting EPC models is a demanding task. In this sense, this paper might

contribute to a standard of EPC modelling. Moreover, the language is a candidate to identify important correctness aspects going beyond general frameworks and recommendations. These aspects and the body of research accompanied to them may serve as a reference both for designing new languages and to standardise existing ones.

The remainder of the paper is structured as follows. In Section 2 the literature analysis is described. In Section 3, EPC-specific aspects of syntactic, semantic and pragmatic quality are presented. In Section 4, the results are summarized by providing a comprehensive overview of quality aspects. The paper ends with a conclusion.

## **2 Preparing and Conducting the Literature Analysis**

The scope of our literature analysis is only limited by content. We did not exclude any work based on its age. The content scope was defined as the intersection of the modelling language EPC and evaluation criteria for process model quality. Therefore, only articles mentioning specifically the EPC were considered in the following. We have not limited our search to publications only addressing the basic EPC, instead we also considered publications that focus on quality aspects of eEPC models.

At first, leading and well established databases that are known for publications around the EPC were considered. In applying this approach, we have examined the working paper series of the IWi (Institute for Information Systems at DFKI) – the institute which published the original EPC back in 1992 [KNS92]. The selection of relevant articles from the IWi reports has been made by their titles and abstracts. Initially, all reports dealing with processes were considered (37 papers). From this set, only articles with business process and EPC subjects have been chosen (8 papers). In addition, we used the EPC Community conference proceedings and examined each conference proceeding available. With a similar strategy as previously explained the amount of hits has been too high (65 papers). Therefore, we adjusted our relevance conditions, so that articles that deal not explicitly with either syntactic, semantic or pragmatic quality aspects were excluded. Additionally, redundant contributions were excluded. With this strategy, also applied at the result set of the IWi reports, three relevant EPC Community papers and two IWi reports could be found. To complement our literature analysis, we used the academic search engine Google Scholar with queries such as “EPC syntax” (70 papers), “semantic syntactic pragmatic consistency EPC” (196 papers) or “modelling conventions business process” (213 papers). After applying our conditions for suitable papers, eleven articles have been chosen. Besides this process, we found relevant literature with the “Method of concentric circles” (Snowball system) [K99]. This method consists in evaluating the references of our already selected literature in order to identify often quoted, thus meaningful papers like [Ga10].

### 3 Quality Aspects of EPC Models

#### 3.1 The SIQ-Framework applied to EPC Models

Quality aspects for EPC models (other types of models as well) can be divided into different areas. As already mentioned, we only considered articles providing aspects that could be mapped in the three main categories of the SIQ-Framework (Figure 2): *Syntactic*, *semantic* and *pragmatic* quality.

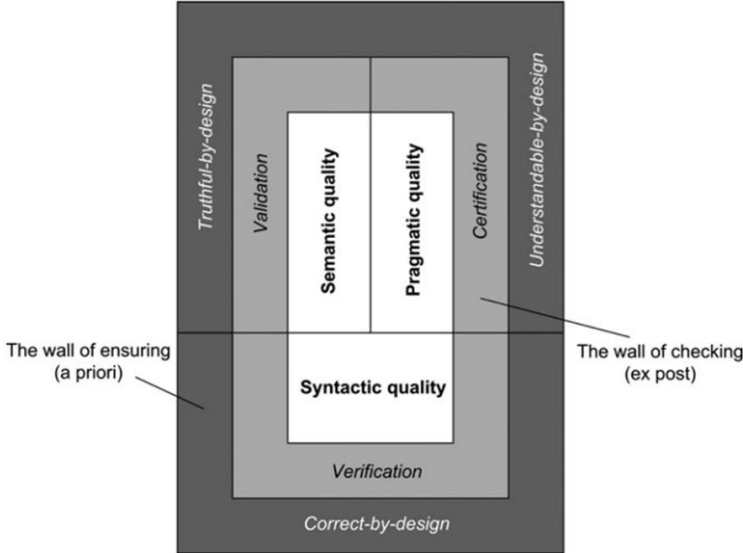


Figure 1: SIQ-Framework [RMR10]

We follow the SIQ-Framework conceptualization of these quality categories [RMR10]:

1. *Syntactic quality* is evaluated by the degree of conformance to a previously defined syntax. Any language defines rules for elements and connections between elements that need to be followed to create a model in this language.
2. *Semantic quality* refers to how well a model is representing the object under consideration. Two sub-goals exist: *completeness* and *validity*. A model is valid if all of its statements are correct and have a relevant reference to the underlying problem. The model is complete if it not only contains relevant references that are correct but also the references that *would be* correct.
3. *Pragmatic quality* is basically a matter of how good a model can be understood by its users. Whilst users might understand a model quite well, it could be still of low semantic quality by leaving out important characteristics of the underlying problem and vice-versa.

The SIQ-framework implicitly shows the relations between the different criteria. For example, syntactic quality is fundamental to semantic and pragmatic quality which is indicated by the fact that the latter two are based on the former. In Figure 2, this is depicted graphically by placing both *semantic quality* and *pragmatic quality* in rectangles on top of the rectangle *syntactic quality*.

## 3.2 Syntactic Quality

### 3.2.1 Representation

The notation of the concepts of the EPC is stated by means of Figure 2.

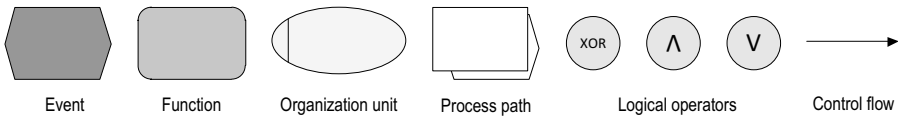


Figure 2: EPC Symbols, c.f. [Sc98, p. 19]

To begin with, there are syntactic rules whose non-conformance does not necessarily lead to a “bad” EPC. So, slight variations in the representation of the EPC elements are possible provided that these are still identifiable (cf. Organization unit in Figure 2). Another possible variation exists in the presentation of the logical operators [St06]. Staud proposes a graphical presentation where the logical operators are modelled as a split circle:

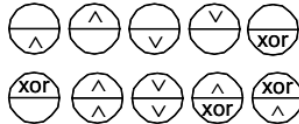


Figure 3: Alternative logical operators [St06]

The upper half represents the incoming control flows and the lower half represents the outgoing control flows. If there is just one out- or incoming control flow the respective half does not show an operator. Furthermore, operators splitting a single control flow to multiple ones are called a *split-operator* and operators joining multiple control flows are called *join-operator* [NR02].

### 3.2.2 Usage and Composition

In addition to the correct representation of model elements, additional rules are considered with the correct use and composition of these model elements. These rules are dealing with the correct connection between events and functions or the correct usage of logical operators. In the following section these rules will be introduced.

- (1) There is at least one start-event and one end-event [LM09; NR02; St06; KKS04; Ga10].
- (2) Events are able to trigger functions and functions have to be triggered by events. In other words: An event will be followed by a function and a function by an event. The only exception is the start- and end-event. The connection between events and functions are shown as a dashed arrow line [St06; KNS92; NR02; Ri00; KK05; Ga10].
- (3) Information objects and Organization units must be connected with functions. The connection is modelled with a dashed line (organization units) respectively with an arrow (information objects) [KNS92; St06]. An extension of this rule additionally requires a function involving human interactions to be associated with an organisational unit. Analogously, a function that does not require any human participation should be associated with an application system [KK05].
- (4) A logical operator connects several events and functions. The join operator needs to be of the same type as the logical operator that initially split the control flow [St06; KK05; NR02; Ga10].
- (5) Rule number 4 leads to the conclusion that logical operators have at least one incoming and one outgoing control flow arc [LM09]. Different sources expand on this: Split and join operators have either one incoming but multiple outgoing control flows or they have one outgoing but multiple incoming control flows [NR02; St06; KKS04; Ri00; Ga10].
- (6) Furthermore, because of the notation that just logical operators are able to connect one event/function with multiple ones (see rule 4), the guideline that events and functions possess just one incoming and one outgoing control flow arc can be deduced. The only exception is the start- and end-events just possess one arc [NR02; Ga10].
- (7) After an event an OR- or XOR-operator must not follow for the next functions. This results from the lack of decision-making power of an event [St06; NR02; KK05; KKS04; Ri00; Ga10]. But there are two papers mentioning that this opinion is not broadly shared [KKS04; KK05].
- (8) Process pointers are connected with events only [KKS04; NR02; St06].

### 3.3 Semantic Quality

#### 3.3.1 Linguistic Correctness

An EPC model is linguistically correct if every label of the model elements follow a specified naming convention. The resulting restrictions for the creator of the model aim

at preventing misinterpretations. At first the German naming convention will be presented. The events of an EPC are representing the current state, so the linguistic correctness requires the label to be created from a substantive and a verb in the past participle form. An EPC function represents the active and time consuming part of an EPC model. Thereby a function's name should be created from a substantive and an active. Objects like an organization unit emulate objects of the "real world" and have to be titled with one or compound substantives (e.g. Business Management) [Ba10]. The English naming convention suggestions for model elements differ from the German conventions. In English, events have to be modelled in a substantive form plus a verb in past participle, functions consists of a verb and a substantive [HF09]. The naming conventions are summarized in Table 1.

Table 7: Name convention for events and functions [Ba10]

|         | Function                                              | Event                                                                    |
|---------|-------------------------------------------------------|--------------------------------------------------------------------------|
| German  | Substantive(s) + Verb<br>(e.g. Bestellung bearbeiten) | Substantive(s) + Verb in Past Participle<br>(e.g. Bestellung bearbeitet) |
| English | Verb + Substantive(s)<br>(e.g. Processing order)      | Substantive(s) + Verb in Past Participle<br>(e.g. Order processed)       |

Additionally, to these generally accepted naming conventions, HUMM and FENGEL mention enterprise- and business-specific conventions. Accordingly, naming conventions of EPC model elements, which capture a rather enterprise-specific process, should conform to the according vocabulary of the enterprise (e.g. "Rent vehicle" instead of „Rent car“) [HF09].

### 3.3.2 Formal Semantics

Due to the possible automatic execution of EPC models, formal semantics require such models to ensure formal correctness, namely to avoid endless loops (livelock) or the unexpected halt of the execution (deadlock) [St06]. These two semantic mistakes will be shown in Figure 4. The model can be started with event E1 being activated resulting in a token which will be passed to function F1. After that function, the AND-split-operator delivers tokens to both control flows. So between the first event E1, the first function F1 and the AND-operator, a livelock occurs. In the subsequent control flow, the XOR-operator delivers just one token to one event, based on the decision in F1. But the following AND-operator expects both tokens, so the end-event will never be reached and a deadlock occurs [GLK].

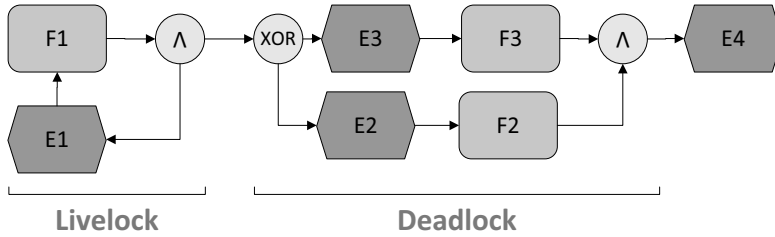


Figure 4: Deadlock and Livelock in an EPC model

A further criterion regarding the semantic correctness is *soundness*. It originally has been created for workflow nets and comprises three criteria: “option to complete”, “proper completion” and “no dead transitions”. Since these three requirements were designed for workflow nets, where just one source and one sink exist, the soundness requirements cannot be applied to EPC directly since they allow for multiple start- and end-events. Hence a slight variation is necessary which has led to the notion of *EPC soundness* [ME09]. EPC soundness also comprises three criteria: (i) a set of start-events exists so that the EPC is not a cycle and (ii) for every possible combination of activation of start-events proper completion is guaranteed, i.e. regardless of which start-event combinations are activated, the model terminates. Furthermore, (iii) the third point is derived from the *relaxed soundness*. For a particular set of final nodes the model has to be sound, i.e. every end-event in the model that is accessible from the start-events must be able to receive a final token. If this requirement is violated, the proper completion would not be guaranteed in the case that all start-events are activated. A taxonomy of soundness terms can be developed by the consideration of the different sets of anomalies. These sets of anomalies can be detected with criteria such as soundness, relaxed soundness and EPC soundness. While the soundness criteria are able to identify all errors, e.g. deadlocks, in an EPC model with just one start and one end node, the relaxed soundness criteria just demands the proper use of the soundness criteria at a particular part of the model (The part that is the desired behavior of the modeler). Therefore it is not possible to identify all errors in the model. The second soundness variation, the EPC soundness, deals with the case of multiple start- and end nodes. So by applying the EPC soundness criteria, it is possible to identify the errors in a model that occur for all possible start combinations. It could therefore be concluded that the original soundness term is the strictest criterion, followed by the EPC soundness and the relaxed soundness as the weakest criterion. Thus, soundness subsumes EPC soundness which in turn subsumes relaxed soundness. Beyond soundness, GRUHN and LAUE present different additional types of semantic anomalies and suggest solutions that will be presented in the following [GL09].

- (1) If two events occur before or after a join-operator with the identical content, the model often can be simplified. However, if the join-operator is a XOR-operator it shall be presumed that a modelling mistake occurred because a XOR-operator merely should be used for mutually exclusive events.

- (2) If an AND- or OR-operator is followed by two events whose content is mutually exclusive, it is a logical mistake because events after an AND- or OR-operator can occur together. Mostly a XOR-operator is the right choice.
- (3) A XOR-operator is followed by three events and two of them are their negation to each other. Therefore, the third event would eventually be needless because a XOR-operator normally models the decision between mutually exclusive events.
- (4) If an operator is used to compare a value (e.g. with a XOR-operator), the issue is often forgotten that the value remains constant.
- (5) If a function models a polar question that just can be answered with yes or no, the function has to be followed by an XOR-operator.

In an older publication GRUHN and LAUE elaborated on further aspects (styling rules) [GL05].

- (6) Functions and events of an EPC model must not be instantiated more than once, i.e. for example a function which is already active must not be activated by another token. Therefore it should be ensured that a control flow arc cannot be reached by several tokens.
- (7) XOR-joins do not block if multiple tokens are possible to arrive at the join, rather the incoming token will be passed to the outgoing arc. This also corresponds to the intention of EPC modellers in the practice, because they assume that anyway just one token will arrive.
- (8) Each split-operator must have a corresponding join-operator. Such models are also called *structured models*.

Other criteria are specified by GRUHN, LAUE, KERN and KÜHNE [GLK08] with reference to notion of soundness that has been introduced originally by VAN DER AALST.

- (9) Of each EPC model element that can be reached from the start-event it must be possible to reach an end-event.
- (10) If a model element does not have a following element, it must be an end-event.
- (11) There must be no element in an EPC model, for which there is no control flow from a start-event to an end-event.

In addition to these EPC-specific semantic rules, the literature often just discuss about concepts to transfer an EPC to a Petri net in order to detect anomalies relating to the formal semantics. Since these approaches do not focus on checking the semantic quality of EPC models directly, they are not covered in this paper due to space limitations.

### 3.3.2 Compliance

Beyond formal semantics, in the area of compliance checking, the correctness of the *models content* (i.e. what is represented in the model) is of central importance. Compliance can be understood as the conformity of something such as a process model to the entirety of relevant legal liabilities, directives and rules as well as to the internal guidelines and best practices of an enterprise [WL08]. Compliance rules can result from scenario-specific, project-specific or general requirements [DF09]. Often, compliance rules target the behavioural aspect of a model that is all possible execution traces. An example of such a compliance rule is: “After the decision that a service is free of charge, no fee calculation can take place”.

- (1) The behaviour of a model should comply with rules specified w.r.t the semantics of individual model elements [SPH04; SM06].

Further distinctions can be made (a) between compliance rules focusing on the *occurrence* of specific nodes in a process graph or the structure of the model and (b) regarding the type of nodes in the process graph between basic flow elements such as functions and events and elements representing resources such as organizational units [FHT11]. These distinctions lead to element flow rules, resource usage rules, element occurrence rules and resource occurrence rules [FHT11]. An example of a simple element flow rule is: “Within 2-3 process steps after order rejection, the customer has to be notified about the decision”.

- (2) The occurrence of model elements and the connections between model elements in an EPC model should comply with rules specified w.r.t the semantics of individual model elements [FHT11].

The attributes of model elements may also be the target of compliance rules [DF09]. An example of such a compliance rule is: “For every object which requires persistence, modifications have to be saved each time the document is read”.

- (3) The model should comply with rules covering both the model elements and their attributes [DF09].

### 3.4 Pragmatic Quality

An important criterion leading to a high pragmatic quality of an EPC model is the ease of understanding of the model for the intended audience. So an EPC model will suffice this demand if it e.g. follows the guideline of clearness comprises by the GoM [BRS95]. Among others, these are included in the following points:

- (1) There should not be overlapping arcs in the generated EPC [Br11].
- (2) If possible, a change of the flow direction should be avoided. In the best case the flow direction goes from the top to the bottom across the whole model [Br11].

- (3) The naming conventions should be applied [Br11].
- (4) Not every function in an EPC model receives an assignment to an organization unit. A new correlation should be done if the organization unit changes [BRS95].
- (5) Very large process chains should be divided with the aid of process pointers in order to promote clarity [St06].

Furthermore, there are a few rules given by [MRA10] that are aiming to a higher pragmatic quality. The relevant guidelines are listed as follows:

- (6) Use as few elements as possible. The bigger the model, the more confusing (and error-prone) it is.
- (7) Minimize the amount of control flows of an element. The more incoming and outgoing control flow arcs there are, the harder it is to interpret the model.
- (8) If possible, use just one start- and one end-event. In addition to the reduction of possible mistakes, it increases the intelligibility.
- (9) Design the EPC model as structured as possible, e.g. by reuniting every split-operator with its corresponding join-operator.
- (10) Avoid OR-operators because of their ambiguity in the semantics.
- (11) Apply the verb-object naming convention in order to prevent ambiguities.
- (12) Split your model if it contains more than 50 elements.

#### **4. EPC Quality Aspects reflected in the Literature**

In the following the quality aspects will be reflected based on the identified literature. The explicit consideration of syntactic aspects by the different authors is summarized in Table 8.

Table 8: Syntactic Quality Aspects in Literature

| <i>Citation</i> | <i>Authors Names</i>             | <i>Syntactic Aspects</i> |          |          |          |          |          |          |          |
|-----------------|----------------------------------|--------------------------|----------|----------|----------|----------|----------|----------|----------|
|                 |                                  | <i>1</i>                 | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> | <i>6</i> | <i>7</i> | <i>8</i> |
| [Ga2010]        | GADATSCH                         | •                        | •        |          | •        | •        | •        | •        |          |
| [KK05]          | KAHL, KUPSCH                     |                          |          | •        | •        | •        |          |          | •        |
| [KNS92]         | KELLER, NÜTTGENS, SCHEER         |                          | •        | •        |          |          |          |          |          |
| [KKS04]         | KLEIN, KUPSCH, SCHEER            | •                        |          |          |          | •        |          | •        | •        |
| [LM09]          | LAUE, MENDLING                   | •                        |          |          |          | •        |          |          |          |
| [MRA10]         | MENDLING, REIJERS, VAN DER AALST |                          |          |          |          |          |          |          |          |
| [NR02]          | NÜTTGENS, RUMP                   | •                        | •        |          | •        | •        | •        | •        | •        |
| [Ri00]          | RITTGEN                          |                          | •        |          |          | •        |          | •        |          |
| [St06]          | STAUD                            | •                        | •        | •        | •        | •        |          | •        | •        |

As it can be inferred the literature is to a certain degree clear about the syntactical aspects of an EPC model. A unique position is hold by rule number six, as this rule is only mentioned by two authors. However, rule number six is often implicitly given, as it can be deduced from other rules.

Table 9: Semantic Quality Aspects in Literature

| Citation | Authors Names                  | Semantic Aspects |   |                  |   |   |   |   |   |   |   |   |    |            |     |   |   |
|----------|--------------------------------|------------------|---|------------------|---|---|---|---|---|---|---|---|----|------------|-----|---|---|
|          |                                | Naming           |   | Formal Semantics |   |   |   |   |   |   |   |   |    | Compliance |     |   |   |
|          |                                | D                | E | 1                | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11         | 1   | 2 | 3 |
| [Ba10]   | BARTSCH                        | •                |   |                  |   |   | • | • |   | • | • | • | •  |            |     |   |   |
| [DF09]   | DRAWEHN, FEJA                  |                  |   |                  |   |   |   |   |   |   |   |   |    |            |     | • | • |
| [FHT11]  | FELLMANN, HOGREBE, THOMAS      |                  |   |                  |   |   |   |   |   |   |   |   |    |            | (•) | • |   |
| [GL05]   | GRUHN, LAUE                    |                  |   |                  |   |   |   |   | • | • | • |   |    |            |     |   |   |
| [GL09]   | GRUHN, LAUE                    |                  |   | •                | • | • | • | • |   |   |   |   |    |            |     |   |   |
| [GLK08]  | GRUHN, LAUE, KERN, KÜHNE       |                  |   |                  |   |   |   |   |   |   |   | • | •  | •          |     |   |   |
| [HF09]   | HUMM, FENGEL                   |                  | • |                  |   |   |   |   |   |   |   |   |    |            |     |   |   |
| [SM06]   | SIMON, MENDLING                |                  |   |                  |   |   |   |   |   |   |   |   |    |            | •   | • |   |
| [SPH04]  | SPECK, PULVERMÜLLER, HEUZEROTH |                  |   |                  |   |   |   |   |   |   |   |   |    |            | •   | • |   |

The coverage of semantic aspects is summarized by Table 3. In the column “Naming”, “D” and “E” represent German or English naming conventions. There is only a small overlap between the papers considered because there are no fixed EPC conventions in the area of semantics in particular. The coverage of pragmatic aspects is summarized by Table 4.

Table 10: Pragmatic Quality Aspects in Literature

| <i>Citation</i> | <i>Authors Names</i>             | <i>Pragmatic Aspects</i> |          |             |          |             |          |          |          |          |           |
|-----------------|----------------------------------|--------------------------|----------|-------------|----------|-------------|----------|----------|----------|----------|-----------|
|                 |                                  | <i>1</i>                 | <i>2</i> | <i>3/11</i> | <i>4</i> | <i>5/12</i> | <i>6</i> | <i>7</i> | <i>8</i> | <i>9</i> | <i>10</i> |
| [BRS95]         | BECKER, ROSEMAN, SCHÜTTE         |                          |          |             |          | •           | •        |          |          |          |           |
| [Br11]          | BRÜCK                            |                          | •        | •           | •        |             |          |          |          |          |           |
| [MRA10]         | MENDLING, REIJERS, VAN DER AALST |                          |          |             | •        |             | •        | •        | •        | •        | •         |

It is possible to summarize rule three and eleven as well as five and twelve according to their content. Through the presence of generally accepted notations for model design pragmatic aspects are barely discussed in literature. It is possible to deduce every rule mentioned in section 3 from the guideline of clearness stated in [BRS95] and [MRA10]. Overall, extensions like the eEPC, the SEQ-Operator [P95] or ET-Operator [R96] and the discussion about their implications for correctness and quality are disproportionately underrepresented in literature. Up to now, discussion about correctness has astonishingly just referred to the “simple” EPC with events, functions and operators.. In addition to the results apparent from the tables, we noticed that the events in the debate about correctness criteria are less important. They are playing a key role regarding the syntactic quality but while discussing the semantic and pragmatic quality of EPC models they are almost not even mentioned. Out of this observation we can point out two ideas for further modelling rules: Prospective modelling languages may not use elements similar to EPC events and insignificant events in EPC models might be abolished. This rule would just be applied on trivial events like “Check contract” – “Contract checked” and could both increase the pragmatic and semantic quality.

## 5. Some Considerations on Standardisation

As the use of EPCs in the practice is widespread, standardisation seems beneficial in order to support the users at modelling correct models and facilitate the implementation of conformant software tools. Especially, a standardisation is required for domain- or subject-specific extensions of the EPC. As far as it concerns the syntax, respectively the syntactic quality, there mostly is consensus among the authors regarding syntactical rules an EPC models has to follow. The problem is more dramatic regarding the semantics of an EPC model. Since the EPC was firstly introduced in 1992 without a complete semantic specification, nowadays there partly is a disagreement about the interpretation of different EPC elements (e.g. does an XOR-operator allow receiving several tokens, how the OR-operator does know when he has received all required

tokens etc.). Above all there is the discussion whether some rules restrict the freedom of modelling the EPC provides. Hence, the literature often looks for a way out by Petri nets whose semantic is sufficiently formalized and defined. The field of pragmatic quality is less strongly discussed in recent articles, presumably because some general guidelines have been published whose compliance already leads to a high pragmatic quality like the „Generalities of Orderly Modelling“ [BRS95] or the „Seven Process Modelling Guidelines“ [MRA10]. Although these were not only published specifically for EPC models, they can easily be applied to it. So ultimately, in order to increase the utility of the EPC as a language for business process modelling, a standard seems to play an important role. To name just two benefits from such a standardisation, first the cooperation between enterprises by means of the EPC would be increased, as common concept will have a common meaning and secondly, the derivation of software artefacts based from an EPC model in the sense of model-driven software engineering would be beneficial as well, as the more semantically enriched foundation tackles both: the humans that execute the business process and the technology that support and automate the execution. However, it is not likely that an “one size fits all”-approach to standardization will significantly increase the utility of the EPC and related tools. Instead, it can be assumed that different project requirements (e.g. documentation, re-engineering or automation) and modeller skills (e.g. novice modeller, intermediate or expert) will require different priorities and levels of conformance that can be consolidated to different EPC profiles in the future.

## 5. Conclusion

We have investigated the state of syntactic, semantic and pragmatic quality of EPC-modelling by conducting a literature analysis. The results show that there is some consensus regarding syntactic aspects, but less regarding semantic and pragmatic aspects. Finally, we advocate for the standardisation of the EPC which should reflect the needs and purposes in different modelling contexts and by different groups of modellers.

---

Acknowledgement: The research presented in this paper is part of the WISMO project and is funded by the German Federal Ministry of Education and Research (BMBF) within the framework concept “KMU-innovativ” (grant number 01IS012046B).

---

## References

- [Ba10] Bartsch, M.: Realisierung der semantischen Annotierung für pModeler, einem System zu semantischen Prozessanalyse. Linz: Johannes Kepler Universität Linz – Institut für Wirtschaftsinformatik, 2010
- [BNT10] Becker, Jörg; Niehaves, Björn; Thome, Irina: How Many Methods Do We Need ? – A Multiple Case Study Exploration into the Use of Business Process Modeling Methods in Industry. In: AMCIS 2010 Proceedings, 2010, p. 534.
- [Br11] Brück, S.: Modellierungskonventionen – Regeln bei der Erstellung eines Modells. Saarbrücken: Universität des Saarlandes: Institut für Wirtschaftsinformatik, 2011

- [BRS95] Becker, Jörg; Rosemann, Michael; Schütte, Reinhard: Grundsätze ordnungsmäßiger Modellierung. In: Wirtschaftsinformatik 37 (1995) 5, pp. 435–445.
- [DF09] Drawehn, J.; Feja, S.: Anwendung von grafischen Validierungsregeln bei der Entwicklung von IT-Integrationsprozessen. In: Münch, J.; Liggesmeyer, P. (Hrsg.): Tagungsband des Workshops "Modellgetriebene Softwarearchitektur – Evolution, Integration und Migration" (MSEIM 2009), 3. März, Kaiserslautern. Bonn : Köllen (GI LNI P150), 2009, S. 367–375
- [FHT11] Fellmann, M.; Högbe, F.; Thomas, O.: Semantic Verification of Business Process Models: Prospects and Limitations. In: Smolnik, S.; Teuteberg, F.; Thomas, O. (Hrsg.): Semantic Technologies for Business and Information Systems Engineering: Concepts and Applications. IGI Global, Hershey PA, USA, 2011, pp. 150–168
- [Ga10] Gadatsch, A.: Grundkurs Geschäftsprozessmanagement Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker. Vieweg+Teubner, Wiesbaden, 2010
- [GL05] Gruhn, V.; Laue, R.: Einfache EPK-Semantik durch praxistaugliche Stilregeln. In (Nüttgens, M.; Rump, F.J., Hrsg.): Geschäftsprozessmanagement mit Ereignisgesteuerten Prozesskette. Hamburg, 2005, S.176–189
- [GL09] Gruhn, V.; Laue, R.: Ein einfaches Verfahren zur Erkennung häufiger Fehler in EPKs. Universität Leipzig: Lehrstuhl für Angewandte Telematik und E-Business, 2009
- [GLK08] Gruhn, V.; Laue, R.; Kern, H.; Kühne, S.: EPK-Validierung zur Modellierungszeit in der bflow\* Toolbox.. Universität Leipzig
- [HF09] Humm, B. G.; Fengel, J.: Assessing Semantic Consistency of Business Process Models. Hochschule Darmstadt: Zentrum für Forschung und Entwicklung, 2009
- [HFL11] Houy, Constantin; Fettke, Peter; Loos, Peter; Aalst, Wil M. P.; Krogstie, John: Business Process Management in the Large. In: Business & Information Systems Engineering, 2011, vol. 3, no. 6, pp. 385–388
- [K99] Krämer, Walter: Wie schreibe ich eine Seminar- oder Examensarbeit? Frankfurt a.M.: Campus 1999
- [KK05] Kahl, T.; Kupsch, F.: Transformation und Mapping von Prozessmodellen in verteilten Umgebungen mit der ereignisgesteuerten Prozesskette. Saarbrücken: Institut für Wirtschaftsinformatik, 2005
- [KNS92] Keller, G.; Nuttgens, M.; Scheer, A.-W.: Semantische Prozessmodellierung auf der Grundlage "Ereignisgesteuerter Prozessketten (EPK)". Saarbrücken: Institut für Wirtschaftsinformatik, 1992
- [KKS04] Klein, R.; Kupsch, F.; Scheer, A.-W.: Modellierung inter-organisationaler Prozesse mit Ereignisgesteuerten Prozessketten. In (A.-W. Scheer, Hrsg.): Veröffentlichung des Instituts für Wirtschaftsinformatik. Heft 178. 2004
- [LM09] Laue, R.; Mendling, J.: Structuredness and its significance for correctness of process models. Springer-Verlag, Berlin, 2009
- [ME09] Mendling, J: Empirical Studies in Process Model Verification. Heidelberg u.a., Springer, 2009
- [MRA10] Mendling, Jan; Reijers, Hajo A.; van der Aalst, Wil M.P.: Seven process modeling guidelines (7PMG). In: Information and Software Technology 52 (Feb. 2010) 2, pp. 127–136.
- [NR02] Nuttgens, M.; Rump, F. J.: Syntax und Semantik Ereignisgesteuerter Prozessketten (EPK). In: Promise 2002 – Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen. Potsdam, 2002, S.14
- [P95] Priemer, J.: Entscheidungen über die Einsetzbarkeit von Software anhand formaler Modelle. Sinzheim: Universität Münster 1995

- [R96] Rosemann, M.: Komplexitätsmanagement in Prozessmodellen. Methodenspezifische Gestaltungsempfehlungen für die Informationsmodellierung. Wiesbaden, 1996
- [Ri00] Rittgen, P.: Quo vadis EPK in ARIS? Ansätze zur syntaktischen Erweiterungen und einer formalen Semantik. Koblenz: Universität Koblenz-Landau, 2000
- [RMR10] Reijers, H. A.; Mendling, J.; Recker, J.: Business process quality management. In: vom Brocke, J.; Rosemann, M. (Hrsg.): Handbook on Business Process Management 1 : Introduction, Methods, and Information Systems. Part II. Heidelberg u.a. Springer (International Handbooks on Information Systems), 2010, S. 167–185
- [Sc98] Scheer, A. W. (1998): ARIS – Vom Geschäftsprozeß zum Anwendungssystem. 3. Aufl. Berlin, Springer, 1998
- [SM06] Simon, C.; Mendling, J.: Verification of Forbidden Behavior in EPCs. In: Mayr, H. C.; Breu, R. (Hrsg.): Proceedings of the GI Conference Modellierung (MOD2006), March, 22 – 24, Innsbruck, Austria. GI (LNI 82), 2006, S. 233–242
- [SPH04] Speck, A.; Pulvermüller, E.; Heuzeroth, D.: Validation of business process models. In: Buschmann, F.; Buchmann, A. P.; Cilia, M. A. (Hrsg.): Proceedings of the 17th European Conference on Object-oriented Programming (ECOOP), July 21–25 2003, Darmstadt, Germany. Springer (LNCS 3013), Berlin u.a., 2004, 9 pages
- [St06] Staud, J.: Geschäftsprozessanalyse Ereignisgesteuerte Prozessketten und objekt-orientierte Geschäftsprozessmodellierung für Betriebswirtschaftliche Standardsoftware. Springer, Berlin u.a., 2006
- [TF09] Thomas, Oliver; Fellmann, Michael: Semantic Process Modeling – Design and Implementation of an Ontology-based Representation of Business Processes. In: Business & Information Systems Engineering 1 (Nov. 2009) 6, pp. 438–451.
- [WL08] Wecker, G.; van Laak, H. (Hrsg.): Compliance in der Unternehmenspraxis: Grundlagen, Organisation und Umsetzung. Gabler, Wiesbaden, 2008



# Ontology-Based Assistance for Semi-Formal Process Modeling

Michael Fellmann, Novica Zarvić, Annelene Sudau and Lennard Nobbe

Information Management and Information Systems

Osnabrück University

Katharinenstr. 3

49074 Osnabrück, Germany

{ michael.fellmann | novica.zarvic | asudau | lnobbe }@uni-osnabrueck.de

**Abstract:** The construction of semi-formal process models is a difficult task due to the complexity of today's business processes. In this contribution, we devise an ontology-based assistant for semi-formal process modeling. The assistant is intended to support the modeler in creating high quality process models with less time and effort by reusing knowledge captured in a reference ontology. At first, we derive functionalities for such an assistant. We then show how the approach can be implemented and integrated into an existing modeling tool. Finally, we describe an experimental evaluation of the approach.

## 1 Introduction

The construction of semi-formal process models is challenging due to the high complexity of today's business processes which do not impose an imminent solution for their representation. Therefore, process modeling can be considered as a form of problem solving rather than the simple mapping of known facts in a graphical model. Current tools have some deficiencies in supporting this problem solving, which can be characterized as *assessment problems* and *reuse problems*. *Assessment problems* can be summarized as questions regarding the fitness of the model for the intended purpose. They are hardly answered by existing modeling tools since they do not “understand” what is being modeled. Understanding in the sense of reasoning about the contents of a model e.g. in order to inform the user about modeling progress or correctness would require formalized domain knowledge incorporated into the modeling tool. Although such knowledge exists (e.g. in the form of ontologies), it is currently not embedded in the tools. *Reuse problems* arise during the creation of models if known solutions of business problems already achieved are not reused. Moreover, the comparison with processes of other companies is a useful and inspiring source of insight (especially spanning different sectors of industries) [APQC10] and it is not leveraged from existing process modeling tools.

In this contribution, we propose the extension of process modeling tools with an ontology-based assistance addressing the two problems introduced above. First we shortly outline some related work (section 2), before we discuss our research method (section 3) that belongs to the Design Science paradigm [HMP+04]. Then, we derive required functions that the assistant should implement in order to address the two problem areas (section 4). We then show how the approach can be implemented and integrated into the graphical user interface of an existing modeling tool and report on our ongoing development activities (section 5). Finally, we present an experiment-based evaluation approach (section 6). Finally we conclude the article (section 7). To the best of our knowledge, this is the first ontology-based approach exclusively geared towards advising the user in *business problems* of process modeling.

## 2 Related Work

The fundamental idea of our approach is to use an ontology in conjunction with assistance functionalities to support the user in modeling. Therefore, research regarding modeling support leveraging knowledge representations in a broad sense is related to our work. Such research can be found in several areas. The *reference-based model construction* aims at leveraging reference models to construct new models. They capture domain knowledge and hence serve as a knowledge representation. For model adaptation and reuse, several methods have been described [FL02; BB06]. Although methods such as configuration [De06; RA07; DK07a] are amenable to IT-support, they are not supported by most existing modeling tools [DK07b] and are still subject to research. Recently, *reuse-based model construction* is discussed in business process management as a more holistic approach in comparison to reference modeling in the sense of spanning different abstraction levels from process oriented systems. The main idea is to leverage existing process assets encompassing not only conceptual models, but also the accompanied software artifacts [FTS+11; TZD11]. Another research area focuses on *query-based support for model construction* where the goal is to find models or model fragments which can be used to complete the model under construction. [Ko07] describes an approach and a tool to determine the similarity of the model currently under construction compared to models which already exist in a repository. Analogous to reuse-based model construction and in contrast to reference-based model construction, no separate knowledge representation is required. In *pattern-based model construction*, the availability of patterns is essential. The idea of patterns originates from the discipline of architecture and has been introduced as analysis patterns in conceptual modeling by Fowler [Fo96]. They represent knowledge in the form of abstract templates used to apply well-known solutions to similar problems. In literature, different types of patterns have been suggested such as control flow [AHK+03], resource [RHE+05], service interaction, [BDH05], data [RHE+05] and activity patterns [TRI09]. Up to now, research has mainly focused on pattern detection [LIT+09; DHL+09], although a few approaches support modeling [TRC+08b] and a tool *ProWAP* has been developed [TRC+08a]. In comparison to other construction techniques such as reference-based model construction, patterns are more abstract in nature and usually require a thorough interpretation by the modeler.

In the area of *case-based reasoning support for BPM* the main idea is combining case-based reasoning and business process modeling in order to leverage past process executions when specifying a new process. Processes that have been executed (also referred to as cases or templates) are stored in a knowledge base and are used to create new processes or workflows [MZM04]. The approaches aim at improving flexibility and reuse [YCW06] increasing adaptiveness [PZW+08] and can act as a knowledge management technique for business process redesign [MMR03].

In order to position *our approach of ontology-based assistance* relative to the other areas of research, two criteria can be used: the amount of formalized knowledge and the complexity and sophistication of supporting tools. Regarding the amount of formalized knowledge, we make use of a comprehensive representation of business knowledge collected in the MIT Process Handbook [MCH03]. It comprises approx. 5000 business functions and processes which are described regarding their inputs, outputs, resources, specializations, generalizations, exceptions, related functions and other information (see <http://ccs.mit.edu/ph/> to explore the ontology and <http://ccs.mit.edu/ophi/> for the Open Process Handbook version). Regarding the amount of formalized knowledge, our approach is similar to automated model construction or case-based reasoning approaches which also rely on comprehensive knowledge bases. However, one decisive difference is that our approach does not require setting up a knowledge base with semantically described building blocks or recorded past process executions.

In contrast, our approach works “out of the box” since it leverages an existing reference ontology and is intended to assist and advice the modeler during modeling. With this paradigm of assistance, we do not want to suggest complete models or to improve models based on past executions. Consequently, the complexity of the tool which has to be developed is lower than the complexity of tools in the two aforementioned categories (cf. Table 1). In essence, we do not want to replace any of the other approaches belonging to other areas of research. We rather seek to offer new forms of additional advice and assistance which should make model creation more convenient by reusing the knowledge contained in a reference ontology.

Table 1: Comparison of research areas w.r.t. formalized knowledge and tool complexity

| Research area                                                          | FK | CT |
|------------------------------------------------------------------------|----|----|
| Reference-based model construction                                     | ○  | ○  |
| Reuse-based model construction                                         | ○  | ○  |
| Query-based support for model construction                             | ○  | ○  |
| Pattern-based model construction                                       | ○  | ○  |
| Automated, planning-based model construction                           | ●  | ●  |
| Case-based reasoning support for BPM                                   | ●  | ●  |
| <i>Ontology-based assistance for model construction (our approach)</i> | ●  | ○  |

Legend:

FK = usage of formalized business knowledge, CT = complexity of supporting tools;

● = high, ○ = medium, ○ = low

### 3 Research Method

The research approach taken in this work can be allocated to the design science paradigm [HMP+04, MS95]. Kuechler and Vaishnavi briefly describe design science research as the “construction of an information technology artifact and its evaluation” [KV11]. We address both mentioned issues – construction and evaluation – in this paper. With respect to the *construction* of the artifact, we first derived the functionalities of the ontology-based assistance by conceptually eliciting the requirements with the help of use cases. On the basis of the collected requirements the assistance functionalities have been implemented. This step implied the choice of a suitable reference ontology, the possibility to query the ontology in an adequate way, as well as the presentation of the query results in a modeling environment. With respect to the *evaluation* issue, we have first chosen a suitable design evaluation method. According to Hevner et al. [HMP+04] several methods can be used for evaluating designed artifacts. In this respect we have decided to perform an experiment in order to be able to study relevant characteristics of our artefact. Therefore we have first set up an experimental setting at our university, and analysed the results. Figure 1 conceptually depicts the described activities that have been performed in this work, where a clear distinction between the construction and evaluation activities can be seen.

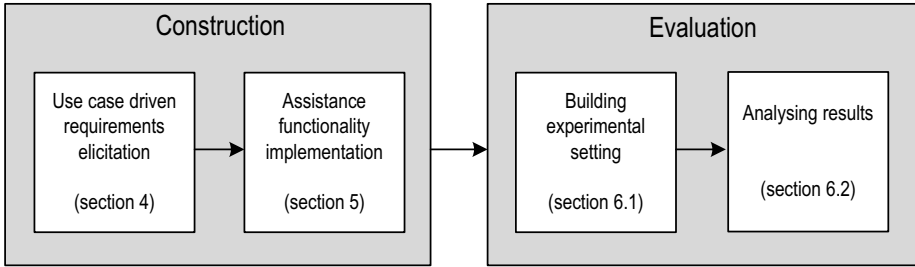


Figure 5: Design science research process taken in this paper

### 4 Assistance Functionality

The functionalities of the ontology-based assistance can be derived by considering the following broadly described use cases. (U1) The modeler has to select a process from the MIT Process Handbook ontology which is the basis of the process she wants to create. This is required to enable the other assistance functionalities. (U2) To support modeling, the assistance function has to be capable of suggesting model elements. (U3) To reuse business knowledge and expertise contained in the ontology, the assistance function has to be capable of providing information either upon user request or in an unobtrusive and pro-active way during modeling. (U4) To assess the model, the assistance function has to provide a verification feature for verifying the model during modeling and when it is finished. These four use cases are described in further detail below and in Figure 2.

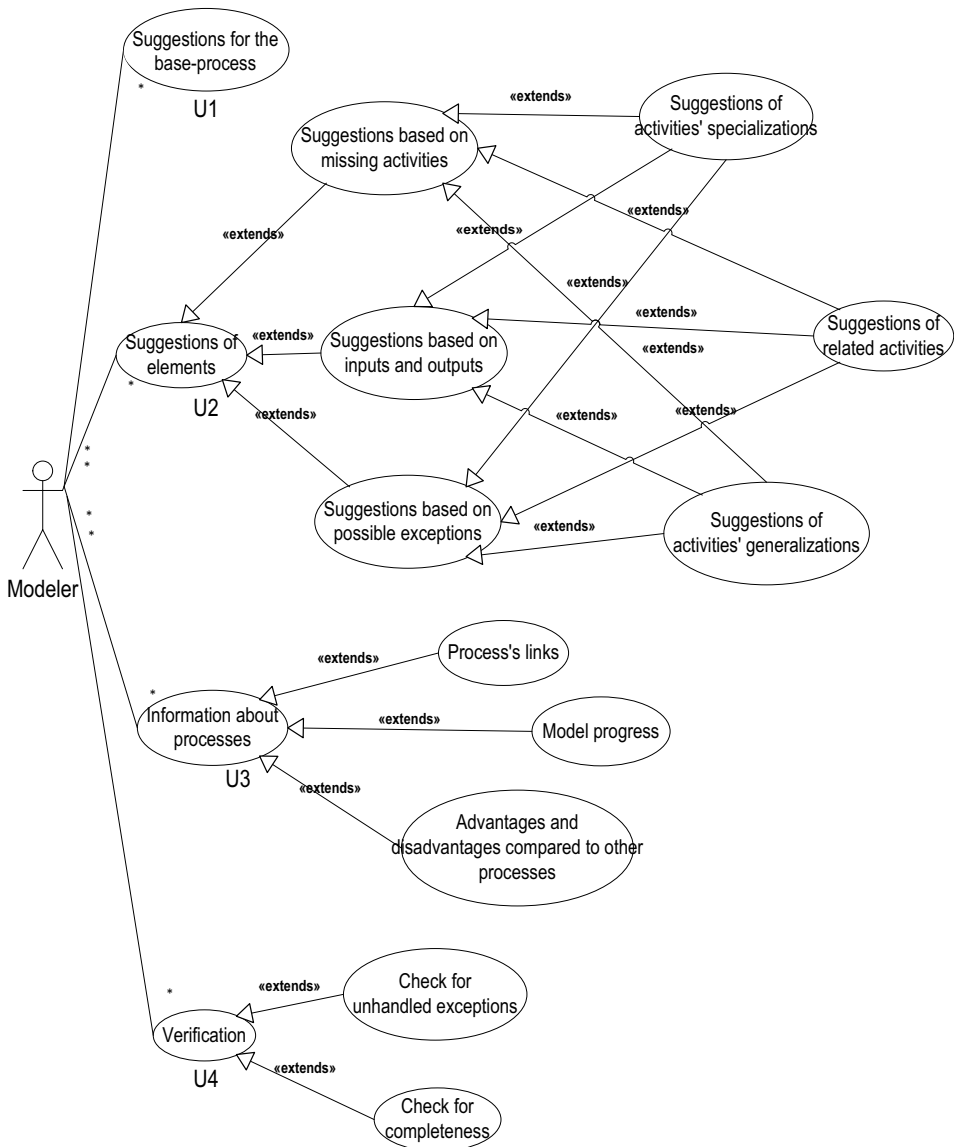


Figure 6: UML Use Case Diagram showing the functionality

*U1: Suggestions for the base-process.* Before modeling with the assistance functionality, the modeler has to select a base-process (e.g. “Hiring process”). Therefore, the modeler has to be able to search the ontology for an adequate base-process e.g. by using keywords or advanced browsing methods. Furthermore, linguistic procedures can be applied to enhance the quality of the search results.

*U2: Suggestions of elements.* Elements which can be inserted in the process model should be suggested based on the information available about the base-process or the elements already present in the model. Using the information about the base process, activities and resources can be suggested by detecting missing elements which are known to be part of the base process but which are missing. Also, suggestions based on the inputs, outputs and exceptions of elements already present in the model can be generated. In this way, the tool can suggest e.g. to insert additional activities for error handling if the current activity the modeler has inserted is known to produce exceptions of a specific type. For each of the suggested activities, the assistance functionality should also offer more specific and/or more generic alternatives and it should display related activities, as such information is captured in the MIT Process Handbook. If the modeler inserts a model element that has been suggested by the ontology-based modeling assistance, this element will be annotated with the ontological semantics which is a prerequisite of use case U4.

*U3: Providing information about processes.* The modeler should be provided with information about a process that is contained in or can be computed using the ontology. Such information encompasses e.g. the links the process has with goals and resources or the progress of model creation. The latter can be calculated from the number of process parts in the ontology and the number of those parts which are already included in the model. Moreover, a trade-off matrix (which is also part of the MIT Process Handbook) can be used to display advantages and disadvantages of alternative process designs.

*U4: Model verification.* Using the element suggestion functionality of use case U3 results in most (or even all) of the model elements being semantically annotated. Consequently, the model elements are backed with ontological (and hence machine processable) semantics enabling semantic verification functionalities such as checking for unhandled exceptions or incomplete process models.

## **5 Implementation of the Assistance Functionalities**

### **5.1 Management of the Ontology Data**

A key challenge in the development of software relying on ontologies is how to guarantee the availability as well as quality of knowledge contained in the ontology. In our research, we decided to use the MIT Process Handbook since it has obvious advantages.

First, its contents have been collected and formulated in scientific projects, which means that a certain quality level can be imputed. Second, it is comprehensive as it contains a substantial number of processes (currently approx. 8000), and thus standard processes are covered. Third, it has been translated into the Web Ontology Language (OWL) in a project at the University of Zurich so that no effort in “ontologizing” the knowledge was required. Fourth, it offers a very regular and clean structure (cf. Figure 3) making it amenable to implement the assistance functionalities on top of it.

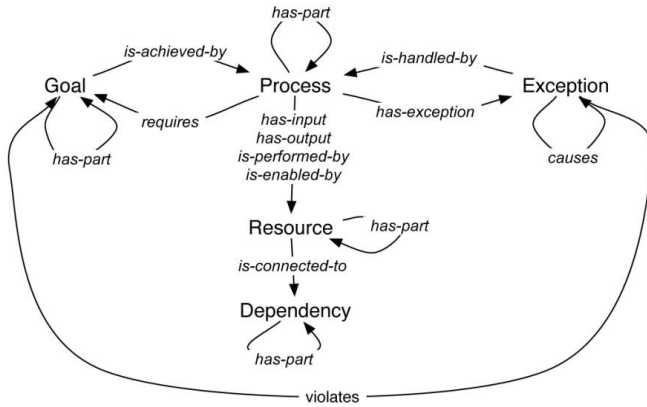


Figure 7: Structural scheme of the MIT Process Handbook (source: [www.ifi.uzh.ch/ddis/](http://www.ifi.uzh.ch/ddis/))

In the OWL version of the Process Handbook, every concept matches one OWL file. Regarding the expressivity, the subset of the MIT Process Handbook we used for implementing our demonstration system had the expressivity of the description logic SHIN. For our demonstration system, we used the Jena Framework combined with an in-memory approach (in the future, we plan to migrate to the OWL-API and OWL-DB).

## 5.2 Querying of the Ontology

To query the ontology, we use the query language SPARQL ([www.w3.org/TR/rdf-sparql-query/](http://www.w3.org/TR/rdf-sparql-query/)) which is the de-facto query language in the Semantic Web, standardized by the W3C and widely supported in triple stores and from various APIs associated with OWL.

In the following, we give an example of a SPARQL query which retrieves the parts of a process whereby the *has-part* relation of the MIT Process Handbook scheme (cf. Figure 3) is used. The query is generated inside the modeling tool and transmitted via HTTP to the server. The results of the query (ID, name and description of a respective process part) are used inside the modeling tool to suggest model elements (cf. use case U2 and section 5.3). The term `<BaseProcessID>` is a variable which is replaced by the ID of the current base-process which is known to the modeling tool since the user first has to select the base-process (cf. use case U1).

```

SELECT DISTINCT ?id ?name ?desc WHERE {
  <BaseProcessID> has-part ?id . ?id rdfs:label ?name.
  ?id rdfs:comment ?desc.
}

```

### 5.3 Presentation of the query-results in a modeling environment

Since it is easy to enhance through Add-Ins, we use Microsoft Visio 2010 for our demonstration system. The Add-In currently uses the template for event-driven process chains EPC. It catches events of inserted elements and matches their type against EPC shapes for the decision whether the assistant should show up or not. In the use case described by the query in section 5.2, the modeler would have modeled a few steps of a process and insert the next activity, so a screen of the application would appear.

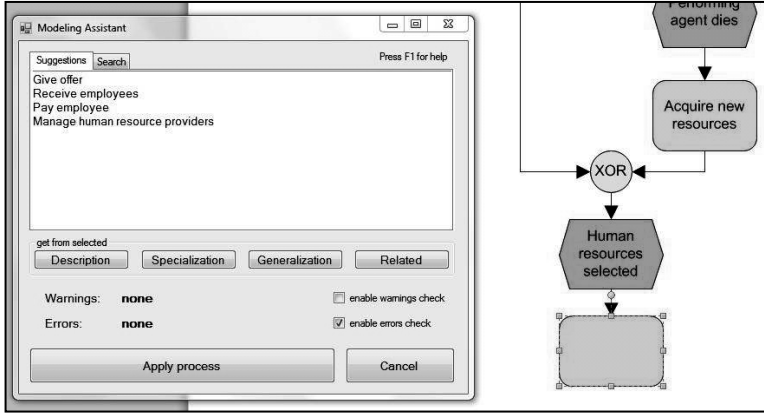


Figure 8: Appearance of the assistant tool while a process is created in MS Visio

The activities which are *suggested automatically* can be selected and applied to the model. Moreover, the modeler can navigate through the MIT Process Handbook by using the structure of specializations and generalizations or related processes. It is also possible to search the ontology for keywords if the modeler is not satisfied with the results.

## 6 Experimental Evaluation

### 6.1 Experimental setting

The goal of the experiment that has been performed was to get a first impression how assistance-based modeling is perceived by the participants while performing a predefined task. More specifically, it targeted at determining previously unknown problems as well as the technical functionality. For doing so, several scenario-based tasks have been formulated. This was accomplished by means of a browser interface, which provides a search function for processes, where the exact phrase can be typed in, the maximum number of results to be displayed as well as which exact elements and fields should be considered. If a process is selected, it is possible to obtain further information on the process. For instance, a full description is displayed or it can be seen whether the process has relation of type specialization or generalization respectively.

The experiment took place at our university and most participating modelers ( $n=6$ ) have been post-graduate students at Master or PhD level. After the experiment, the participants have been interviewed in an in-depth interview with no predefined structure in order to additionally capture participants' ideas and feelings about the assistance system.

For being able to assess the results of the experiment in an objective manner, the results had to be mapped against a predefined sample solution. By following this approach three sets of elements can be distinguished. The first set contains the elements of the predefined sample solution, the second set contains the elements that have been provided by the modeler, and finally the third set of elements, which the modeler used from the MIT process handbook. The Following figure 5 shows some situations that may occur.

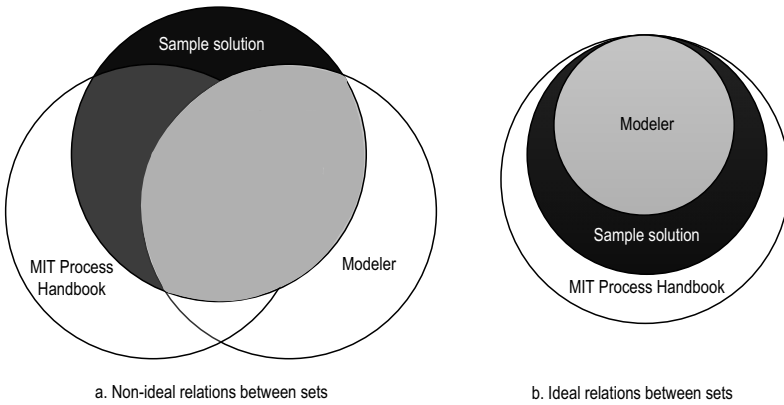


Figure 9: Relations between sets

On the left side we can see that the modeler has used some but not all elements from the process handbook and from the sample solution, because there are elements not covered by those sets. In Figure 5b on the right side we can see that all the elements used by the modeler are part of the sample solution, which again is part of the elements from the process handbook. This situation would – of course – mean that the elements used by the modeler are also all inside the lexical knowledge in the process handbook. This is not the case in Figure 4a and means that the elements from the handbook neither cover all elements in the sample solution nor from the modelers set.

## 6.2 Evaluation Results

The tasks on which the participants of the experiment worked on contained to model three different processes with and without support of the before presented assistant tool, which are: (i) hiring and new employment of a worker, (ii) the process of distributing an arbitrary product assortment, and (iii) the creation and placing advertisements for own products. The assessment of the results (cf. Figure 6) was centered on the difference of *elements* by modeling with or without support of the assistant tool, which explains the

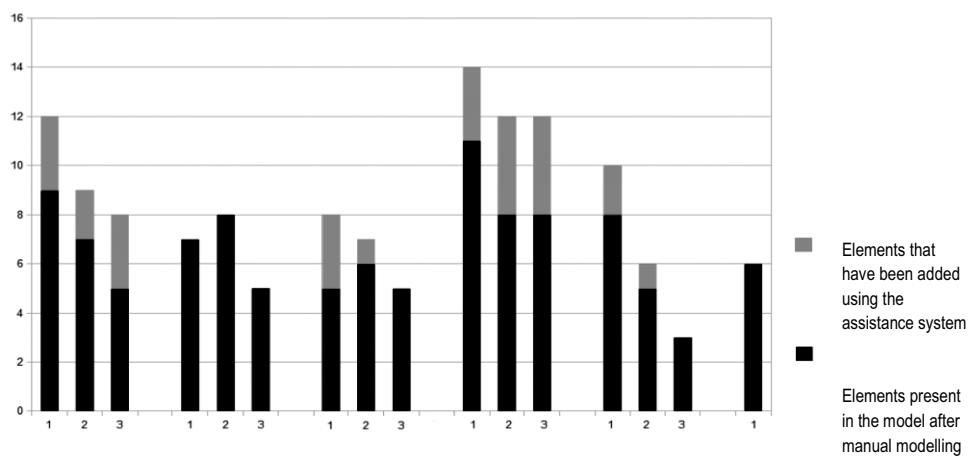


Figure 10: Assessing the relationship between inserted model elements by manual modeling and assisted modeling

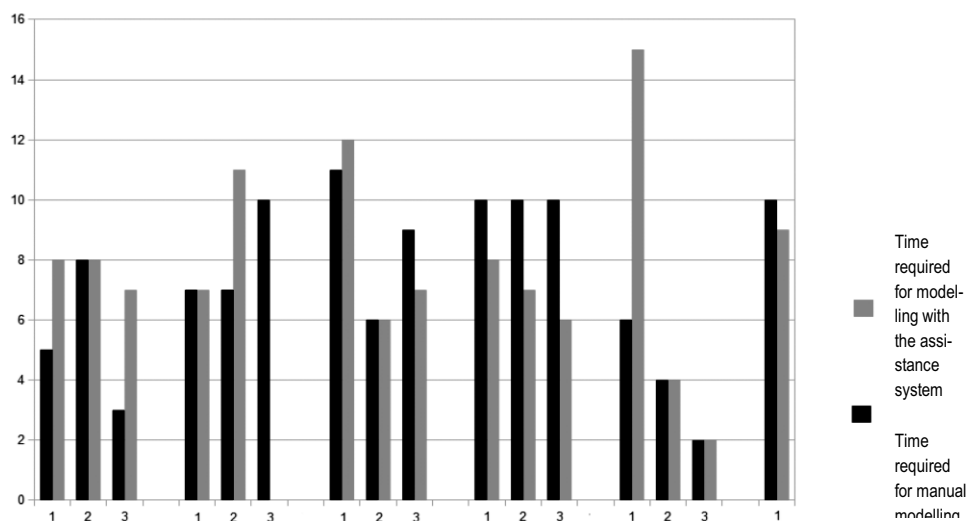


Figure 11: Assessing the time needed to create models with assistance and without assistance

relationship depicted in the diagram. Models that have been created by using the assistance function in most cases had benefit in the way that elements have been added to the model (red colour), which were beforehand absent while modeling manually (blue colour) without ontology-based assistance. However, the results of our evaluation show that one of the biggest problems is the degree of abstraction or level of detail respectively. The MIT process handbook contains process steps that are formulated in a rather generic way and not very detailed. This leads to the fact that the handbook was not very useful to the modeler, because it suggested only a small set of elements during process modeling.

The second criterion was to determine also differences in the *time* needed to create a process model (cf. Figure 7). Even though we have chosen parts of the handbook that seem to be more complete for our experiment, we had to notice that the information contained in the handbook for some areas seems to be incomplete by including also empty processes without any further information. In a future version of our system we will tackle this problem by integrating additional sources of process knowledge such as the Process Classification Framework (PCF) which is a collection of approx. 1000 business activities grouped on four levels. Despite the fact that our present prototype system has some shortcomings the participant's judgment has been positive. In the interviews that we conducted after the experiment, the participants emphasized that the system has helped them to grasp the essence of the process to be modeled – that is what to model and what to omit. One participant put it like this: *“If there had been suggestions, they helped me to focus on the right abstraction level and on what's really important. It somehow increased my confidence in what I'm modeling and therefore I liked it very much. It feels like the lane-assist in my car!”*.

## 7 Conclusion

We provided a first insight in our ongoing activity of developing an ontology-based assistance for semi-formal process modeling. The combination of a reference ontology and a modeling application provides a very promising and effective way to facilitate the creation of process models even for inexperienced modelers. It allows for assessing the model contents e.g. regarding the completeness of a model in contrast to a base-process selected from the ontology. It also leverages the business knowledge collected and represented in comprehensive ontologies such as the MIT Process Handbook.

In this contribution, we have derived functionalities for such an assistant based on use cases. We then showed how the approach can be implemented by elaborating on the storage and retrieval of ontological content. We thereby described the integration of query results into the graphical user interface of an existing modeling tool, which is the basis of our demonstration system. Finally, we provided a first evaluation of the approach describing an experiment.

In our future work, we plan to augment our demonstration system by using more comprehensive suggestions including also industry-specific suggestions provided e.g. by the industry-specific versions of the Process Classification Framework that have been released recently. Also, regarding the process of recommendation, we plan to incorporate concepts of recommender systems such as top-k rating or content-based recommendation strategies.

---

Acknowledgements: The authors are thankful to the European Regional Development Fund (ERDF) for supporting the project SEMPHis-(Semantische Modellierung prozessorientierter Informationssysteme) under number ZW 6- 80131109.

---

## References

- [AHK+03] van der Aalst, W. M. P.; Ter Hofstede, A. H. M.; Kiepuszewski, B.; Barros, A. P.: Workflow Patterns. In: Distributed and Parallel Databases 14, No. 1, 2003; pp. 5–51
- [APQC10] APQC: Process Classification Framework (PCF), Version 5.2.0, 2010. [www.apqc.org](http://www.apqc.org), accessed 12.06.2011
- [Ba+10] Balko, S.; Ter Hofstede, A. H. M.; Barros, A. P.; La Rosa, M.; Adams, M. J.: Business process extensibility. In: Enterprise Modelling and Information Systems Architectures Journal 5, No. 3, 2010; pp. 4–23
- [BDH05] Barros, A.; Dumas, M.; Ter Hofstede, A. H. M.: Service interaction patterns: Towards a reference framework for service-based business process interconnection. QUT Technical Report, FIT-TR-2005-02, Brisbane, Queensland University of Technology, 2005
- [BB06] vom Brocke, J.; Buddendick, C.: Reusable Conceptual Models – Requirements Based on the Design Science Research Paradigm. In: Hevner, A. R. (Eds.): Proceedings of the First International Conference on Design Science Research in Information Systems and Technology : February 24–25, 2006, Claremont, CA , 2006
- [De06] Delfmann, P.: Adaptive Referenzmodellierung. Methodische Konzepte zur Konstruktion und Anwendung wiederverwendungsorientierter Informationsmodelle. Logos, Berlin, 2006
- [DHL+09] Delfmann, P.; Herwig, S.; Lis, L.; Stein, A.: Pattern Matching in Conceptual Models – A Formal Multi-Modelling Language Approach. In: Mendling, J.; Rinderle-Ma, S.; Esswein, W. (Eds.): Proceedings of the 3rd International Workshop Enterprise Modelling and Information Systems Architectures (EMISA 2009) Concepts and Applications, September 10–11, Ulm, Germany, 2009; pp. 13–27
- [DK07a] Delfmann, P.; Knackstedt, R.: Konfiguration von Informationsmodellen – Untersuchungen zu Bedarf und Werkzeugunterstützung. In: Oberweis, A.; Weinhardt, C.; Henner, G.; Koschmider, A.; Pankratius, V.; Schnizler, B. (Eds.): Proceedings of eOrganisation: Service-, Prozess-, Market-Engineering: 8. Internationale Tagung Wirtschaftsinformatik (WI 2007), Karlsruhe, Germany, Feb. 28 – March 2, 2007; paper 62
- [DK07b] Delfmann, P.; Knackstedt, R.: Towards Tool Support for Information Model Variant Management – A Design Science Approach. In: Österle, H.; Schelp, J.; Winter, R. (Eds.): Proceedings of the Fifteenth European Conference on Information Systems. University of St. Gallen, 2007; pp. 2098–2109
- [DeBh11] Derguech, W.; Bhiri, S.: Reuse-Oriented Business Process Modelling Based on a Hierarchical Structure. In: zur Mühlen, M.; Su, J. (Eds.): Proceedings of the Business Process Management (BPM 2010) Workshops, September, Hoboken, NJ, United States. Springer (LNBIP 66), Berlin, 2011; pp. 301–313

- [FTS+11] Fantinato, M.; de Toledo, M. B. F.; de Souza Gimenes, I. M.; Thom, L. H.; Iochpe, C.: Introduction to the First International Workshop on Reuse in Business Process Management (rBPM 2010). In: zur Mühlen, M.; Su, J. (Eds.): Proceedings of the Business Process Management (BPM 2010) Workshops, September, Hoboken, NJ, United States. Springer (LNBIP 66), Berlin, 2011; pp. 223–225
- [FL02] Fettke, P.; Loos, P.: Methoden zur Wiederverwendung von Referenzmodellen – Übersicht und Taxonomie. In: Becker, J.; Knackstedt, R. (Eds.): Arbeitsberichte Nr. 90, Institut für Wirtschaftsinformatik, Westfälische Wilhelms-Universität Münster, 2002
- [Fo96] Fowler, M.: Analysis Patterns: Reusable Object Models. Addison-Wesley Longman, Amsterdam, 1996
- [HMP+04] Hevner, A.R.; March, S.T.; Park, J.; Ram, S.: Design Science in Information Systems Research. In: MIS Quarterly 28, No. 1, 2004; pp. 75–105
- [Ko07] Koschmider, A.: Ähnlichkeitsbasierte Modellierungssunterstützung für Geschäftsprozesse. Universität Karlsruhe, Fakultät für Wirtschaftswissenschaften, 2007
- [KV11] Kuechler, B.; Vaishnavi, V.: Promoting Relevance in IS Research: An Informing System for Design Science Research. In: Informing Science: the International Journal of an Emerging Transdiscipline 14, No. 1, 2011; pp. 125–138
- [LIT+09] Lau, J. M.; Iochpe, C.; Thom, L. H.; Reichert, M.: Discovery and analysis of activity pattern co-occurrences in business process models. In: Proceedings of the 11th International Conference on Enterprise Information Systems (ICEIS 2009), May 6–10, Milan, Italy, 2009; pp. 83–88
- [MS95] March, S.T.; Smith, G.F.: Design and natural science research on information technology. In: Decision Support Systems 15, No. 4, 1995; pp. 251–266
- [MCH03] Malone, T. W.; Crowston, K.; Herman, G. A. (Eds.): Organizing Business Knowledge: The MIT Process Handbook. The MIT Press, 2003
- [MMR03] Mansar, S. L.; Marir, F.; Reijers, H. A.: Case-based reasoning as a technique for knowledge management in business process redesign. In: Electronic Journal on Knowledge Management 1, No. 2, 2003; pp. 113–124
- [MZM04] Madhusudan, T.; Zhao, J. L.; Marshall, B.: A case-based reasoning framework for workflow model management. In: Data & Knowledge Engineering 50, No. 1, 2004; pp. 87–115
- [PZW+08] Pinggera, J.; Zugall, S.; Weber, B.; Wild, W.; Reichert, M. U.: Integrating case-based reasoning with adaptive process management. Tech. Rep. TR-CTIT-08-11, CTIT, University of Twente, The Netherlands, 2008
- [RA07] Rosemann, M.; van der Aalst, W. M. P.: A configurable reference modelling language. In: Information Systems 32, No. 1, 2007; pp. 1–23
- [RHE+05] Russell, N.; ter Hofstede, A.; Edmond, D.; van der Aalst, W.: Workflow data patterns: Identification, representation and tool support. In: Delcambre, L.; Kop, C.; Mayr, H. C.; Mylopoulos, J.; Pastor, O. (Eds.): Conceptual Modeling – ER 2005 : Proceedings of the 24th International Conference on Conceptual Modeling (ER 2005), October 24–28, Klagenfurt, Austria. Springer (LNCS 3716), Berlin, 2011; pp. 353–368
- [TRC+08a] Thom, L. H.; Iochpe, C.; Reichert, M.; Weber, B.; Matthias, D.; Nascimento, G. S.; Chiao, C. M.: On the Support of Activity Patterns in ProWAP: Case Studies, Formal Semantics, Tool Support. In: Revista Brasileira de Sistemas de Informação (iSys) 1, No. 1, 2008; pp. 27–53
- [TRC+08b] Thom, L. H.; Reichert, M.; Chiao, C.; Iochpe, C.: Applying Activity Patterns for Developing an Intelligent Process Modeling Tool. In: Proceedings of the 10th International Conference on Enterprise Information Systems (ICEIS'08), June 12–16, Barcelona, Spain, 2008
- [TRI09] Thom, L. H.; Reichert, M.; Iochpe, C.: Activity patterns in process-aware information systems: Basic concepts and empirical evidence. In: International Journal of Process Integration and Information Management (IJPIM), 2009

- [TZD11] Tran, H.; Zdun, U.; Dustdar, S.: Name-Based View Integration for Enhancing the Reusability in Process-Driven SOAs. In: zur Mühlen, M.; Su, J. (Eds.): Proceedings of the Business Process Management (BPM 2010) Workshops, September, Hoboken, NJ, United States. Springer (LNBIP 66), Berlin, 2011; pp. 223
- [VDB11] Vulcu, G.; Derguech, W.; Bhiri, S.: Business Process Model Discovery Using Semantics. In: zur Mühlen, M.; Su, J. (Eds.): Proceedings of the Business Process Management (BPM 2010) Workshops, September, Hoboken, NJ, United States. Springer (LNBIP 66), Berlin, 2011; pp. 326–337
- [YCW06] Yao, Q.; Chen, Z.; Wang, H.: Improving Flexibility and Reusage of Business Process Management: the Role of Cased-based Reasoning Technique. In: Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE 2006), October 24–26, Shanghai. IEEE, 2006; pp. 187–194

# Towards Auto-Suggested Process Modeling – Prototypical Development of an Auto-Suggest Component for Process Modeling Tools

Nico Clever, Justus Holler, Maria Shitkova, Jörg Becker

ERCIS – University of Muenster

Leonardo-Campus 3

48149 Muenster, Germany

{ nico.clever | justus.holler | maria.shitkova | joerg.becker }@ercis.uni-muenster.de

**Abstract:** Companies have accepted process modeling as a powerful instrument for business reorganization, requirements specification in software development, knowledge management and other activities. Huge amounts of processes are being modeled in organizations nowadays. However, the re-use of existing process knowledge in order to simplify the modeling process has not yet been thoroughly studied and applied. We propose an auto-suggest component for process modeling tools, which, based on existing process knowledge, “auto-suggests” the process model elements step-by-step, thus saving the modeler time and effort.

## 1 Introduction

Business process management (BPM) has established itself as a mature field of research during the last decades. There are several commonly used process modeling notations [Ag04] and process modeling tools as a part of complex business process management suites supporting modeling projects [Ga10]. Companies have acknowledged process modeling as a powerful instrument for business reorganization, requirements specification in software development, knowledge management and other activities [In09]. As a result, large process model collections are created in organizations nowadays, such as APROMORE (2000 models with 20-100 tasks in each) [LR11], SAP reference model (600 models) [Me08] or the repository of Dutch local governments council (500 models) [Di11].

Given the size of process model collections and single process models, manual creation of all processes within the repository becomes a tedious, time-consuming, and, thus, inefficient task. Therefore, we argue that the existing process knowledge from already created processes in the organization as well as from reference models should be re-used to facilitate modeling activities and enhance the efficiency in terms of model quality, modeling time and budget.

Driven by the research question, *how process modeling efficiency can be improved by using existing process knowledge*, we propose an auto-suggest component for process modeling tools. Similar to the T9 system for mobile phones or the auto-completion

function in a UNIX shell or web-search [In09], the active modeler is supported by automatic recommendations for the next process step. These recommendations are derived logically from a knowledge database filled with the process models created in the own or reference organizations.

The remaining paper is structured as follows: within the next section we describe the research design applied, followed by a literature review and the conceptual design of our auto-suggest prototype. The implementation and application section specifies the design decisions. In the last section we discuss the applicability of the conceptual design and prototypical implementation of the auto-suggest component and related future research and ongoing evaluation respectively.

## 2 Research design

In this article we follow the Design Science Research Methodology (DSRM) proposed by Peffers et al. [Pe07]. The DSRM is based on five consecutive phases depicted in Figure 1, namely problem identification and motivation, definition of the objective of the solution, design, demonstration and evaluation. Peffers et al. [Pe07] mention a sixth phase, communicating the research results, which is not depicted in the figure. This step is achieved by the ongoing research contributions.

In the first phase, we identify and motivate the problem by conducting a literature review [Vo09, WW02]. We used “auto-completion”, “auto-suggest” and “recommendation-based” as keywords to find existing approaches in the area of process modeling facilitation and selected those articles which had an explicit focus on BPM. On the basis of the literature review, we found that this topic is not widely discussed and present in the research literature at the moment. Based on the research gap, we define the objective of our solution.

In the design phase, based on auto-suggest approaches in different IS fields such as software engineering, web design and mobile application design, we develop a conceptual model of the auto-suggest component for a BPM tool prototype [Be13]. This prototype was then implemented as a web-based process modeling tool to demonstrate its applicability and practicability.

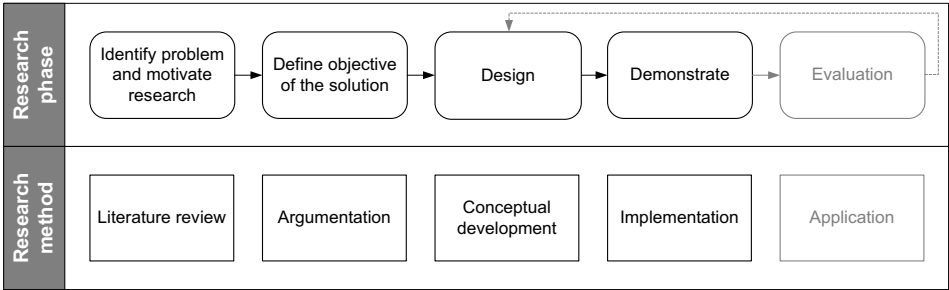


Figure 1: Research methodology

The last phase of the DSRM is greyed out in Figure 1: Research Figure , as the prototype evaluation is not addressed in this paper. The reason for it is the necessity of “learning” of the auto-suggest component, which will be discussed later in the paper. The evaluation is taking place at the time of this contribution and the results will be published after the evaluation is finished. Based on the evaluation results, we will address the improvement of the prototypical implementation.

### 3 Related work

In [HKO07], one approach for auto-completion based on business rules and structural constraints is presented. It uses Petri-nets described with the web ontology language OWL [MV04], which makes it possible to implement an efficient algorithm for (semi-) automatic similarity computation between process variants. Additionally, the semantic web rule language (SWRL) [Ho04] is used to describe structural and business-constraints. According to predefined rules, the system can automatically select process fragments that are then inserted at a particular step of the process model. The prototypical implementation of the conceptual idea from [HKO07] was presented in [HKO09]. The authors have extended the OWL description with two properties and created 18 SWRL rules which allow for automatic detection of suitable process fragments.

A recommendation-based modeling support system is proposed in [HKL08]. The system consists of a query interface, recommender and ranking functions and aims at re-using existing process fragments and, thus, at facilitating process modeling. Process fragments are described by tags or key words, which are derived from process descriptions according to the score function. [KO11] extends this approach by additionally integrating factors which influence process modeling (modeling purpose, view, role, model properties and complexity). The solution also includes a collaboration component – the recommender system tracks the users who have already utilized the recommended process fragment and shows this information in the query results screen which supports inexperienced modelers in their recommendation choice.

[Be06] presents a conceptual description of an approach for auto-completion of business process models based on the OWL which allows for an automatic analysis of process models. During the modeling process, a recommendation mechanism determines possible subsequent fragments of all templates by computing similarities. The system also ensures structural correctness of the models when suggesting fragments (check for deadlock freeness and soundness of the resulting model). The prototypical implementation of the approach is not yet finished or presented by the authors.

In [MP08], an alternative approach for reusing business process fragments is proposed. The authors construct a formal model for the description of process models using a  $\pi$ -calculus for dynamic aspects and an ontology stack for static aspects of process models. The authors have not yet finished the querying framework, so the approach is not completely implemented.

Table 11. Auto-suggest approaches in literature

| Source            | Modeling language                                    | Degree of automatisisation | Technology used                         | Impl. |
|-------------------|------------------------------------------------------|----------------------------|-----------------------------------------|-------|
| [HKO07]           | Petri nets                                           | Semi-automatic             | Ontology, structural correctness rules  | no    |
| [HKO09]           | Petri nets,<br>(possibly applicable to WS-BPEL, EPC) | Semi-automatic             | Ontology, structural correctness rules  | (yes) |
| [HKL08]<br>[KO11] | Petri nets                                           | Manual query               | Key words                               | yes   |
| [HKL08]<br>[KO11] | Petri nets                                           | Semi-automatic             | Key words, structural correctness rules | (yes) |
| [Be06]            | Petri nets                                           | Semi-automatic             | Ontology, structural correctness rules  | no    |
| [MP08]            | On the example of BPMN                               | Manual query               | Ontology, $\pi$ -calculus               | no    |

In Table 11, the approaches to auto-suggest process modeling described above are summarized. All of the approaches focus on the suggestion of process fragments, and not single elements. Most of the solutions, except [HKO09] and [MP08], are targeted to Petri nets, as this language has a formal representation which allows for checking the suitability of suggested fragments according to structural correctness properties such as deadlock-freeness or soundness. Most of the approaches are based on representing a process model with the web ontology language OWL, but two of them, [HKL08] and [KO11], use key words for describing the models. It has to be stated that, except for the query interface by [HKL08] and [KO11], the implementation of the approaches is either not performed or not clearly presented in the articles.

Thus, based on the literature review, we argue that there is still a gap in the area of auto-suggest functionality for process modeling. There is no universal language-independent approach to facilitate manual process modeling, which re-uses existing process knowledge and makes real-time suggestions during model creation. In the next section, we present a conceptual design for such an auto-suggest component for process modeling.

## 4 Conceptual design of the auto-suggest component for process modeling

### 4.1 Architectural design scenarios

Concerning the design of the auto-suggest functionality for process modeling, three different scenarios are applicable. The scenarios differ in their degree of specialization in favor of a concrete modeling environment and, thus, directly affect the flexibility of the

proposed solution. The first scenario seeks for a direct integration of the auto-suggest component with an existing process modeling environment. The second and third scenarios favor generic, environment-independent solutions, either based on a specific modeling language or not.

The first implementation scenario relies on a complete integration of the auto-suggest component into the modeling environment (Figure 2). This scenario does not require or make use of an external data source. It solely uses the existing models in the modeling environment or a predefined exemplary set of models. Due to the close integration with the modeling language and tool, semantic characteristics can be considered for providing the suggestions. For example, models in the notation of the Event-driven Process Chain (EPC) in its original form consist of alternating events and functions [Sc00]. When providing a suggestion for an element of type „event“, all the „events“ will be excluded from the suggestion set. Since no interface to an external database, e. g. other modeling environment has to be designed, no non-trivial generalizations of heterogenous process information is necessary on the one hand. On the other hand, this is a clear disadvantage of this solution, since only process elements can be transferred between two instances of the same modeling environment via export and import and, therefore, the reutilization of existing knowledge is hindered.

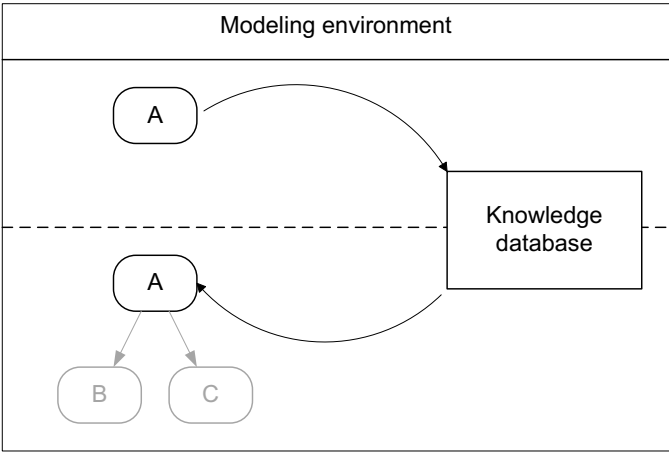


Figure 2: Completely integrated auto-suggest architecture

In the second scenario, the business logic and database of the auto-suggest component are outsourced from the actual modeling environment to a self-contained application. Both the modeling environment and the auto-suggest component communicate via a specifically designed interface. However, since the suggestions are – as they are in the integrated scenario – provided for one and the same modeling language, specific structural and semantic characteristics can still be utilized. As shown in Figure Figure 3, the architecture of this solution allows for the outsourcing of the auto-suggest component to an external server which, in turn, simplifies the exchange of knowledge between different process repositories.

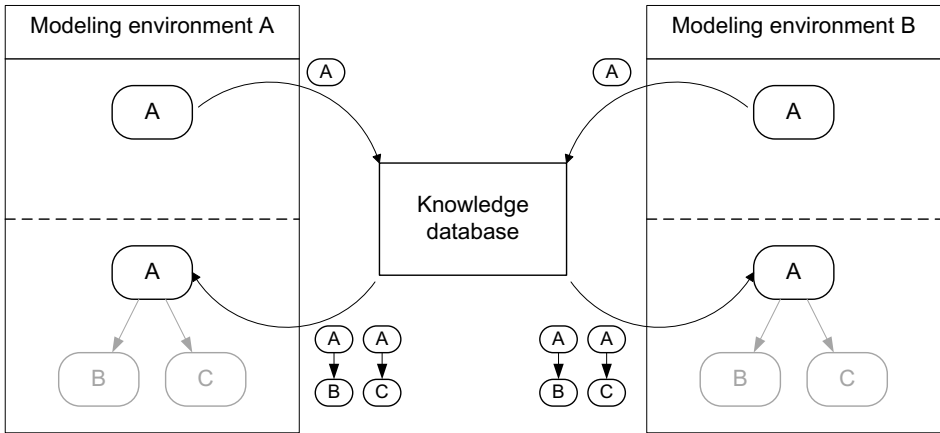


Figure 3: Self-contained, language-dependent auto-suggest architecture

In the third scenario, the auto-suggest component is not only self-contained but language-independent, as well. Here, only the denotations of the process elements – also called labels – are transferred to the knowledge base and used for suggestion. Unlike in the other scenarios, utilization of the language-specific semantic or syntactic characteristics for suggestion provision is not possible. A precondition for the conjoint utilization of such a central language-independent knowledge base for several different modeling environments is a consistent use of denomination conventions/patterns for process elements. In

Figure 4, the general architecture of such a central process modeling knowledge base scenario is shown. As outlined in the figure, the semantics incorporated by the previous solutions – in terms of concrete model elements represented by rounded rectangles – are left out. This represents storing process modeling knowledge in the external database purely based on the labels of the process elements.

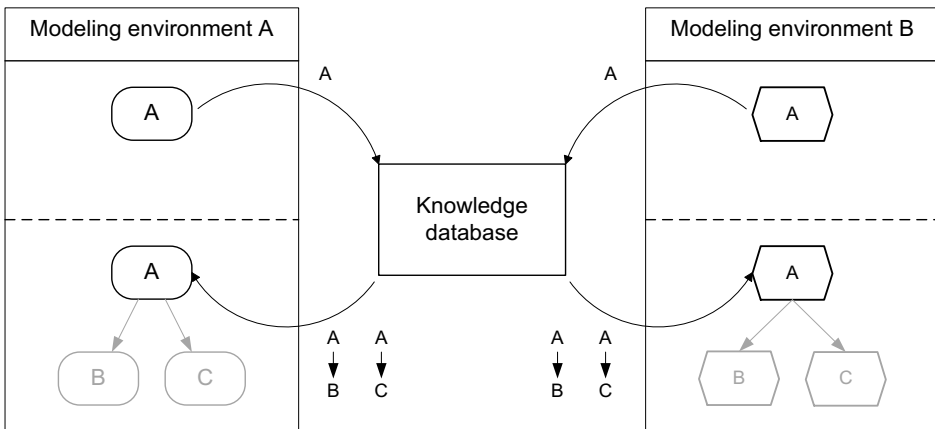


Figure 4: Self-contained, language-independent auto-suggest architecture

## 4.2 Additional requirements

Next to selecting the architectural design scenario for the prototype implementation, some additional requirements have to be considered. These requirements will be discussed in this section.

First of all, the selection process of a potential suggestions has to be determined. Several criteria are relevant in this case:

- Frequency of the sequence – an element would be selected if it is the successor of the predecessor in a large set of models in the knowledge base.
- Date of the insertion – an element would be selected depending on whether it was lately used as the successor of the previous element and would not be selected if it has not been used in a specific timespan.
- Previous users of the suggestion – following [KO11], social structures of process modeling is an important aspect to consider. Thus, entries from experienced users or own entries could be selected preferably.
- Frequency of suggestion-adoptions – a ratio between how often a successor was suggested and how often it was actually accepted as a valid process model element. The higher the ratio, the more likely an element will be suggested. After a certain threshold the element would not be suggested anymore.
- The similarity of element labels – calculated according to the Levenshtein distance<sup>3</sup> between two labels without taking semantic peculiarities of certain modeling languages into account [HM11].

On the basis of these criteria, learning of the knowledge database is realized. It can be divided into two phases. In the first phase, process models and their elements are loaded into the database by transferring predecessor-successor-pairs of elements. In the second phase, the frequency for suggestion (score) of these pairs is continuously adjusted as soon as new pairs of connected elements are created, or the existing ones are deleted in the modeling environment. For our prototype, a rather basic selection methodology is chosen. It is a combination of two aspects mentioned above, the frequency of usage and the date of the last usage counted in days. This is done to prove the applicability of such a selection method, but at the same time to be able to evaluate the learning functionality of the prototype in a comprehensible way. Thus, the selection formula is:

$$S(h, a) = h * e^{-\frac{\log 2}{365} * a}$$

$S$  represents the score assigned to a pair while  $h$  represents the frequency of usage and  $a$  represents the date of the last usage counted in days. In simple words, this exponential function can be interpreted in the way that the score of a certain connection is cut in half if it has not been used in a year.

Two additional aspects have to be considered prior to the development of the prototype. First, handling of sensitive data plays an enormous role in the development of an auto-suggest component for process modeling. Since business process modeling is often a

---

<sup>3</sup> The Levenshtein distance is a metric, measuring the distance between two string sequences.

matter of highly sensitive organizational information, the possibility of restricting the access to intra-organizational knowledge has to be accounted for in the concept.

Second, the extent of suggested elements is of high importance. Since the aim of auto-suggested process modeling is to save the modeler time, it would be suboptimal if too many suggestions would be provided by the tool. Hence, the challenge is to provide a number of choices which is as big as necessary and as small as possible. Based on the calculation of the weighted score  $S(h, a)$ , two approaches are feasible. First, a threshold could be defined which has to be exceeded by an element in order to be suggested. Second, a fixed number of suggestions provided by the tool could be defined ordered by the decreasing score  $S(h, a)$ . Advantage of the first method is a well-formed result set matched to the modeler's needs if the threshold is well-chosen and the quality of the knowledge in the database is good. Pitfall of the first method is a possibly too broad result set if the initialing element is an often used one. This, in turn, is not the case if the second method is applied. If an adequate number of elements to be suggested is defined<sup>4</sup>, the modeler can get an overview quickly and choose the fitting suggestion. Pitfall of this method could be that the modeler could be provided with suggestions which are of no interest in the explicit modeling context. For our prototype, we chose the latter method, because its pitfall of possibly providing unfitting suggestions is more acceptable than an excess supply of suggestions. However, extensive evaluation of the prototype will be carried out in order to assess and possibly correct the decisions made in the conceptual design phase of the auto-suggest component.

### 4.3 Architecture of the auto-suggest component

Given the multiplicity of process modeling tools and notations, as it was stated in the introduction of the article, the self-contained, language-independent approach is chosen for the auto-suggest component. Since it is detached from a specific modeling language, the application of the auto-suggest functionality is possible in a much broader context and is not limited by language-specific characteristics. Thus, a more meaningful application with respect to the potential of auto-suggest based process modeling itself is enabled.

Due to the choice of a self-contained solution, the auto-suggest component has two main tasks. On the one hand, it has to provide an interface to the process modeling tools which are delivering the learning input. As a matter of course, the tools which are to benefit from the functionality and to be included in the evaluation, have to match this interface when providing the information. On the other hand, it has to provide the business logic for the selection of the recommended successors of process elements. As an additional third, trivial task, the component serves as a process knowledge storage, containing all the information about existing internal and reference processes. The architecture of the resulting auto-suggest component and its interfaces is depicted in Figure 5.

---

<sup>4</sup> Adequate in this case always depends on the modeling purpose and the modeling context. Furthermore, a comprehensive evaluation of the prototype has to be carried out in order to provide a general statement in this context. Thus, no generally applicable rule can be provided at the current point.

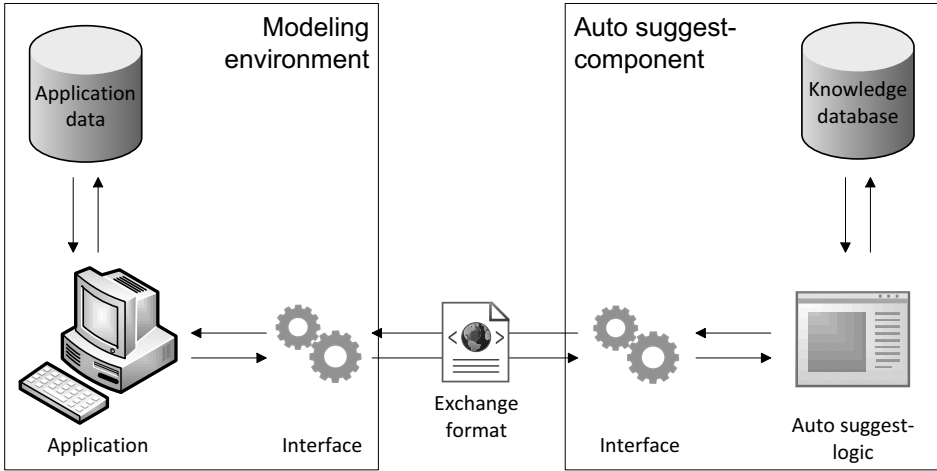


Figure 5: Architecture of the auto-suggest component

The data exchange format between the modeling environment(s) and the auto-suggest component, should account for three requirements. First, it has to be simple to be integrated into the interfaces of various modeling environments allowing for a wider application of the approach. Second, it has to be powerful enough to transfer all the necessary information. Third, in consideration of the execution time requirements of auto-suggested process modeling it has to allow for a short processing time of a suggestion-request and the respective response from the component. As the de-facto standard for data exchange on the web fulfilling all of these requirements, the JSON (JavaScript Object Notation) format is chosen for our prototype [Nu09].

## 5 Implementation and application

As pointed out in the conceptual design, our auto-suggest component for process modeling tools prototype is a self-contained, modeling language-independent web application. It can be viewed as a client-server application, where the auto-suggest component acts as a server and the associated modeling tools are clients. The application provides the following functionality:

- **Web-interface:** To respond to requests from the modeling tools, the application has a web-interface which provides the suggestions via the JSON-based exchange format.
- **Knowledge database:** To store the information about the process elements which is delivered by the modeling environments, the prototype possesses a high-performance database. The interface to this database allows for an efficient processing of the incoming requests, e. g. either storing information or accessing existing knowledge.

- Auto-suggest logic: To assess the quality of the suggestions and provide an adequate and fitting amount of recommendations, the prototype features the auto-suggest logic with the characteristics described in the previous section. Thus, it calculates the score of a certain suggestion based on the formula

$S(h, a) = h * e^{-\frac{\log 2}{365} * a}$ , which is dependent on the frequency of recommendation usage and the date when the recommendation was used for the last time. To put it simply, the score is halved if a connection has not been used in the last 365 days. This approach provides only a predefined number of suggestions to keep the time-saving effect for the modeler as high as possible. Furthermore, it accounts for sensitive (organizational) information as it allows the connections to be flagged for usage in a specific modeling environment only.

To enable the usage of the prototype, modeling environments have to be connected to the prototype in order to facilitate the learning of the knowledge base and to evaluate the effect of the auto-suggest functionality on the modeling time. Hence, a web-based modeling tool should be extended by an interface to communicate with the auto-suggest component. Likewise, it provides the following functionality:

- Obtain suggestions: For obtaining suggestions, the interface recognizes the label of the process element and sends a request for suggested elements to be placed after this element.
- Maintain connections: The interface is able to provide the auto-suggest component with the data set from the modeling environment to account for an initial knowledge database setup. Furthermore, it provides the component with changes in the modeling environment, either by inserting new connections between process elements or by changing the existing ones.
- Display suggestions: The modeling environment displays the suggestion provided by the auto-suggest component by outlining them in their potential position.
- Apply suggestions: The interface of the modeling environment allows for inserting the suggested element if the user accepts a certain suggestion.

Figure 6 shows the information flows between modeling environment and auto-suggest component. First, the user selects a process element in the modeling environment (1). The modeling tool then transfers this information to the web-interface (2) which, in turn, converts the information to the JSON-based exchange format and sends it to the auto-suggest component (3). The interface of the component converts the data to the native format and passes it to the auto-suggest logic (4). A query to the knowledge database is constructed to find possible successors of the element (5). The database responds with a set of possible recommendations (6). The auto-suggest logic calculates the scores for each returned element and passes the results via the web-interface (7) to the web-interface of the modeling tool (8). The data is converted back to the native format of the modeling tool and proceeds it to the application (9). The suggestions are then shown to the user (10). If the user decides to accept a suggestion (11), the further steps are carried out. As this is optional, i. e. that the user does not have to accept the suggestion, the further steps are denoted in brackets in the figure. The accepted suggestion is processed

by the modeling tool and stored in the process repository (12). The resulting process model is then shown to the user (13). In parallel, the information that the user has accepted the suggestion, is passed to the auto-suggest component via the web-interfaces (14, 15 and 16). The auto-suggest logic then increases the frequency of this particular connection in the underlying knowledge base (17).

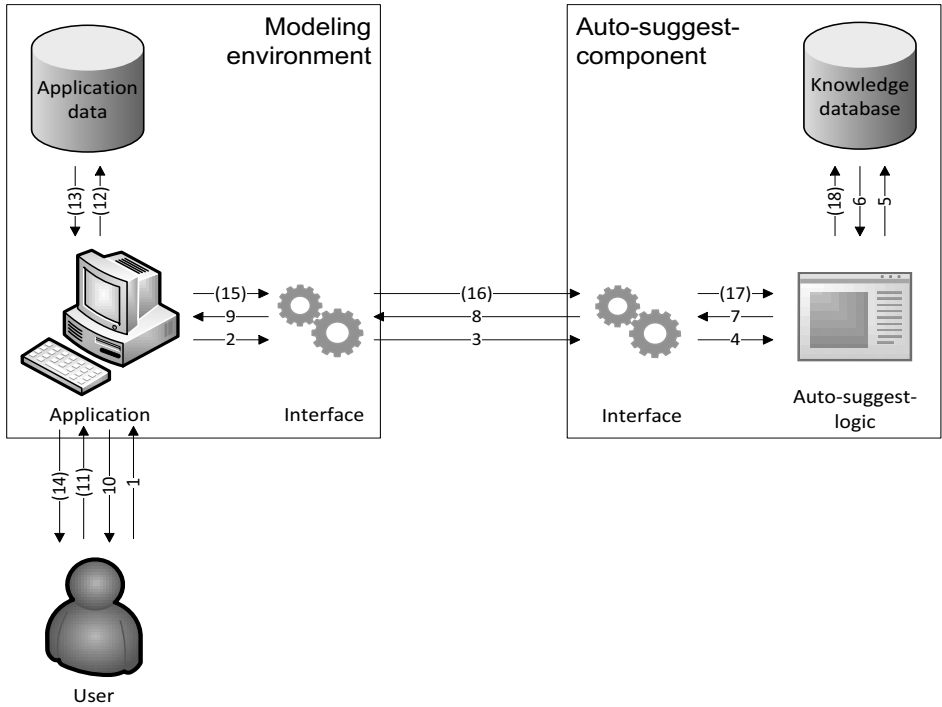


Figure 6: Prototypical information flow of the suggestion process

## 6 Discussion and further research

In this article, we have proposed an auto-suggest component for process modeling tools. An introduction of such a component should facilitate process modeling activities by supporting the user in manual creation of the process models.

We have designed our prototype as a self-contained, modeling language-independent web application with a rather simple recommendation mechanism to prove its applicability. We are going to extend and improve the recommendation mechanism after the evaluation of our solution has finished. The evaluation of the recommendation mechanism and the overall prototype is taking place at the time of this contribution and first results will be available in the near future. At the current state we can say that the prototype successfully demonstrates the concept's feasibility. As the perceived usefulness has to be evaluated by users in concrete modeling settings, we cannot yet tell if the research goal is fully reached. To prove it, we need to access the change in

modeling speed, quality and overall usability of the solution. Therefore, we are planning to conduct a number of experiments to measure time spent with and without the auto-suggest component. Furthermore, we want to elaborate on possibilities of increasing comparability of process labels and, therefore, process elements themselves. Possible approaches to solving this problem can be found in [HKO09, HKL08].

## Bibliography

- [Ag04] Aguilar-Savén, R. S.: Business process modelling: Review and framework. In *International Journal of Production Economics*. 90, 2004; pp. 129–149.
- [Be06] Betz, S.; Klink, S.; Koschmider, A.; Oberweis, A.: Automatic user support for business process modeling. In *Proceedings of the Workshop on Semantics for Business Process Management*. 2006; pp. 1–12.
- [Be13] Becker, J.; Clever, N.; Holler, J.; Püster, J.; Shitkova, M.: Semantically Standardized and Transparent Process Model Collections via Process Building Blocks. In *Proceedings of the The Fifth International Conference on Information, Process, and Knowledge Management - eKNOW 2013*. Nice, 2013; pp. 172–177.
- [BKR11] Becker, J.; Kugeler, M.; Rosemann, M.: *Process Management: A Guide for the Design of Business Processes*. Springer, Berlin, 2011.
- [Di11] Dijkman, R.; Dumas, M.; Van Dongen, B.; Käärik, R.; Mendling, J.: Similarity of business process models: Metrics and evaluation. In *Information Systems*. 36, 2011; pp. 498–516.
- [Ga10] Gartner: *Magic Quadrant for Business Process Management Suites*. 2010.
- [HM11] Haldar, R.; Mukhopadhyay, D.: Levenshtein Distance Technique in Dictionary Lookup Methods: An Improved Approach. In *arXiv preprint arXiv:1101.1232*. 2011.
- [Ho04] Horrocks, I.; Patel-schneider, P. F.; Boley, H.; Tabet, S.; Grosz, B.; Dean, M.: SWRL : A Semantic Web Rule Language Combining OWL and RuleML. In *W3C Member submission*, 21, 2004; 79.
- [HKO07] Hornung, T.; Koschmider, A.; Oberweis, A.: Rule-based autocompletion of business process models. In *Proceedings at the 19<sup>th</sup> Conference on Advanced Information Systems Engineering (CAiSE)*. 247, 2007.
- [HKO09] Hornung, T.; Koschmider, A.; Oberweis, A.: A recommender system for business process models. In *17<sup>th</sup> Annual Workshop on Information Technologies & Systems (WITS)*. 2009.
- [HKL08] Hornung, T.; Koschmider, A.; Lausen, G.: Recommendation based process modeling support: Method and user experience. In *Conceptual Modeling-ER 2008*. 2008; pp. 265–278.
- [In09] Indulska, M.; Green, P.; Recker, J.; Rosemann, M.: Business process modeling: Perceived benefits. In *Conceptual Modeling-ER 2009*. 2009; pp. 458–471.
- [KO11] Koschmider, A.; Oberweis, A.: Designing Business Process with a Recommendation-Based Editor. In *Data & Knowledge Engineering*. 70, 2011; pp. 483–503.
- [LR11] La Rosa, M.; Reijers, H.: APROMORE: An advanced process model repository. In *Expert Systems with Applications*. 38, 2011; pp. 7029–7040.
- [Me08] Mendling, J.; Verbeek, H. M. W.; Van Dongen, B. F.; Van der Aalst, W. M. P.; Neumann, G.: Detection and prediction of errors in EPCs of the SAP reference model. In *Data & Knowledge Engineering*. 64, 2008; pp. 312–329.
- [MV04] McGuinness, D. L.; Van Harmelen, F.: OWL web ontology language overview. In *W3C recommendation*. 10, 2004; 2004-03.

- [MP08] Markovic, I.; Pereira, A.: Towards a formal framework for reuse in business process modeling. In *Business Process Management Workshops*. 2008; pp. 484–495.
- [Nu09] Nurseitov, N.; Paulson, M.; Reynolds, R.; Izurieta, C.: Comparison of JSON and XML data interchange formats: A case study. In *Computer Applications in Industry and Engineering (CAINE)*. 2009.
- [Pe07] Peffers, K.; Tuunanen, T.; Rothenberger, M. A.; Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. In *Journal of Management Information Systems*. 24, 2007; pp. 45–77.
- [Sc00] Scheer, A.-W.: *Aris - Business Process Modeling*. Springer, Berlin, 2000.
- [Vo09] Vom Brocke, J.; Simons, A.; Niehaves, B.; Riemer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In *Proceedings of the ECIS 2009*. Verona, 2009; pp. 2206–2217.
- [WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In *MIS Quarterly*. 26, 2002; pp. xiii–xxiii.



# **Business Process Modelling**



# On the cognitive understanding of types in modeling languages

Dirk van der Linden<sup>1,2,3</sup> and Henderik A. Proper<sup>1,2,3</sup>

<sup>1</sup> Public Research Centre Henri Tudor, Luxembourg, Luxembourg

<sup>2</sup> Radboud University Nijmegen, the Netherlands

<sup>3</sup> EE-Team, Luxembourg, Luxembourg

dirk.vanderlinden@tudor.lu, e.proper@acm.org

**Abstract:** We investigate how enterprise modelers see common types (e.g., actor, event, process) used in most modeling languages in terms of their semantic feature structure (e.g., is human, is material). We hypothesize that modelers have specific interpretations for some of these common types that affect their range of conceptually valid instantiations (e.g., actors should not be instantiated as human things). Based on two exploratory psychometric studies performed with enterprise modeling practitioners and computing science students we discuss the way these typical interpretations affect their model(ing) semantics (e.g., results typically having to be modeled as well-described and non-natural entities, restrictions typically as logical necessities), and what consequences these findings have for modeling languages and the use and creation of models themselves, especially in an inherently collaborative effort like enterprise modeling. We conclude by arguing that insights into these conceptualizations are likely useful and should receive more attention and studies.

## 1 Introduction

While the quality of a conceptual model can be interpreted differently depending on what people expect from it [FHVL12, HFL12], its validity is contingent on its ability to communicate its intended meaning clearly and completely [Rob06]. As models are there to support the building and exchange of knowledge [Sta00], it is important that we can be sure that the semantics in a model reflect the semantics intended by the modeler. This goes further than merely ensuring that the model is capable of expressing those semantics that the original modeler wanted to (with, e.g., test scenarios). We need to additionally ensure that the model does not allow one to express situations that, in the real world, are considered conceptually invalid (or simply illogical) by the original modeler and stakeholders.

We focus here primarily on domain models that are created in order to communicate about a domain with other people. With such models it is important to avoid *mismatched understandings between different modelers and stakeholders* [KSKS12], as they can have a disastrous effect on both the produced models and the modeling process itself. As such, it is important to ensure that a user of the model does not have to guess about the semantics expressed in it. They might not have been involved in the original modeling

process, or come from a different background with their own set of typical understandings, and thus not share the same conceptualization as the original modelers.

What is typically already done to ensure the validity of such domain models is to reach consensus on what type to represent elements from the universe of discourse by, and how to call those representations. However, such discussions often take form of agreeing on the type to be used for such elements (e.g., "Let us model this pizza as a RESOURCE.", "Let us model the delivery of the pizza as a PROCESS."), while not necessarily going into detail on just what that type ('resource' or 'process') actually *is* in this context. Instead, the deeper semantics often stay implicit. For example, the business process for pizza deliveries might involve restrictions that are logical necessities and cannot be broken (e.g., a pizza can not be delivered before it is made, thus correctly observing temporal dependencies) but also restrictions that can be broken, but perhaps *ought* not to because of moral or financial constraints (e.g., a pizza should be delivered within 30 minutes of its order). It is thus important that whatever type we choose to represent these restrictions accommodate the conceptual distinction between alethic (i.e., logical *necessity*) and deontic (i.e., how the world *ought* to be) modality. But, if we only agree that we model these kinds of restrictions as "rules", we leave this distinction implicit and depend on the (selection of the) modeling language to dictate what exactly the semantic status of both of these rules can be (let alone that users at times deliberately or unconsciously ignore a language's semantics and invent their own, not always explicit, semantics [HS05]). As some languages accommodate less conceptual distinctions than others, that means relevant information might be lost.

Let us consider actors. If a modeler typically conceptualizes actors as human beings who take actions on their own accord, and wishes to model them so, it is necessary for whatever type is used from the modeling language to support the relevant parts of that conceptualization. In this case, making it explicit that any instantiation of an actor needs to be a single human being, and more specifically, a (in the given context) autonomous human being, for instance when modeling social decision making processes in an enterprise. If the produced model, however, is not explicit about these distinctions, there can be a host of conceptually related, but invalid, instantiations (e.g., an employee who needs permission for every action and can thus not be considered autonomous, a department with multiple persons). Such problems can be prevented by either using a modeling language that accommodates the necessary conceptual distinction, or by explicitly modeling such distinctions manually (e.g., by using unary constraints on the type for each necessary semantic feature). However, being aware of the relevant and necessary conceptual distinctions *before* the model creation process is what is often still lacking.

Thus, in order to ensure that a model can clearly and completely communicate its intended meaning, regardless of who is interpreting it at what time, a certain degree of conceptual alignment is needed. Specifically, an alignment between the conceptualization the original modeler has of his modeling concepts (i.e., the types used by a modeling language to represent the domain concepts) and the 'official', used, semantics of those modeling concepts (i.e., the semantics as found in the specification of the used language). This makes it easier to ensure that the modeling language (or dialect thereof) which will be used accommodates the needed conceptual distinctions. What is most lacking for this is either a simple and systematic procedure to uncover such

distinctions (which is difficult as poignant distinctions might simply never be explicitly talked about), or an initial set of commonly occurring distinctions that can be used as guidelines in such a process. When such distinctions have been discovered (either as generalized findings or specifically for a group), they can be of value to the modeling process by e.g., providing model facilitators with input for the facilitation process [RHLR11], or by providing explicit focus points for annotations in the model. Such factors should help enhance the quality of created models and improve consensus amongst stakeholders [SPS07].

Enterprise Modeling (EM) is an excellent environment to perform research into such commonly occurring distinctions. As it is a kind of collaborative modeling [SHP09,RGS+08], which is burdened with the difficulty of requiring “*people within the business to express their views in terms of a modeling language*” [BKV09], who often have a different view on things [RKV08] and then needing to integrate or link all those models written by different people in different languages different languages [Lan04,OB06], it is especially important to be aware of the different conceptualizations they might have of the types used in their models. Furthermore, it is possible that different aspects of the enterprise (e.g., goals, processes, value-exchanges) were modeled by separate groups and need to be integrated without access to (all) the involved people, showcasing again the need to be absolutely sure that the semantics expressed in the models can be relied on. Thus, one can expect a diverse amount of ways that modelers see different concepts.

We will first focus on investigating whether a small group of enterprise modelers have particularly specific conceptualizations of types used by conceptual modeling languages used in EM (e.g., i\* GHYA07], BPMN Obj10], e3Value [GYvdR06], RBAC [FCK95], ITML [FHK+09]), and then investigate whether there are common conceptualizations (e.g., actors are considered as singular autonomous human beings, resources as materially existing things) that could be used as pointers for a procedure to uncover conceptual distinctions about these types. With these results we hope to show it is useful to have insights into such conceptualizations, that our approach can be used as a systematic procedure to uncover them, and that larger studies with more participants are warranted.

The rest of this paper is structured as follows. In section II we clarify what the objects of our investigation are, and in section III we set out our research methodology for investigating them. We show the most important results in section IV, discuss them and their consequences in sections V and VI, and finally conclude in section VII.

## 2 What we investigate

We need to look at someone’s concept of a type in order to accurately investigate it. A concept in this context is the understanding that people have of the things they use to model, which is not necessarily the same as the understanding that is prescribed to them (by e.g., language specifications, official standards, or group ‘consensus’). This should be done by investigating people, because “*semantic memory for concepts is based on a subject’s memories of past experiences with instances of those concepts*” [Gee10] and because people generally do not think in the semantics of a given modeling language, but in the semantics of their own natural language [Sow10]. While we can investigate a

concept, we cannot measure it completely, but can instead characterize it (cf. [MAG+11,Pin07]) by looking which semantic features (i.e., distinctive attributes or properties of something that contribute to its meaning) resonate strongly with it for a person. This gives us information about the range and boundaries of the concept, which directly corresponds to the amount of conceptually valid instantiations of a type.

Different conceptualizations of a type can thus affect both the amount, and the content of the conceptually valid instantiations. When we look at the concept *ACTOR* with no given context, we could instantiate it with elements as human beings, computer hardware, abstract entities (e.g., agents), and organizational departments. However, when there is some context (i.e., from the characterization of the concept for that type), the amount of possible instantiatable elements becomes smaller. Assume we know that someone sees actors as human ‘things’. Computer hardware and abstract entities are now no longer conceptually valid instantiations. An organizational department might still be a conceptually valid instantiation of that concept as it is not uncommon for people to conceptualize as sets of human beings being essentially human. However, when we add to the context that actors are also considered single (non-composed) things, we do likely have to rule out a department as a conceptually valid actor. The way that an increasing amount of contextual information (in our case, relevant semantic features) affect the range of a concept can be understood in terms of quantum collapse in the semantic space of that concept (cf. [BC05,BW08]), which is backed up by recent psychological work on theory of concepts [GRA08]).

Consequently, we can effectively characterize the conceptual understanding people have of the types common to most modeling languages by investigating their concepts for them in the light of a set of relevant features. To do so, we will base ourselves on earlier work we performed [vdLHLP11] on the specifications of a number of languages and methods covering different aspects used in enterprise modeling (e.g., processes, value exchanges, goals, architecture, performance, security). This resulted in a set of high-level concepts that cover most types, and a set of features on which different languages were found to differ in their typical use of the concept.

The concepts we look at are actors, events, goals, processes, resources, restrictions and results. The different features which can be combined to characterize them are whether things are considered to naturally occur (natural), be human (human), are single things or composed of many (composed), are intentional or unintended (intentional), are logical necessities or not (necessary), are physically existing or not (material), and whether they are vague or well described (vague). A combination of any of these features can then be taken as a characterization of a given concept, for instance an actor being a natural, human, non-composed material thing.

It is important to note that not all features apply to each investigated concept (or are simply not informative), and that we investigate what people consider, regardless of whether it makes sense, such as e.g., people considering processes to be human things instead of abstract chain of events. Nonetheless, such findings of how people actually perceive (e.g., their personal ontology) the world, which might, for instance be choosing to look at a process in terms of the people performing the individual events in it instead of the events occurring can be useful for other purposes (see e.g., [Alm09] for an argument necessitating these endeavors to validate shared ontologies).

### 3 Method

We use a semantic differential [OST57] in order to investigate the conceptual understandings which participants have of the selected types. It is a widespread method used in different fields and areas of inquiry, and has well-researched specific quality guidelines for the Information Systems [VM07] field to ensure quality and validity of its results. The semantic differential can be used to investigate what connotative meanings apply to an investigated term, e.g., whether ‘red’ is typically considered good or bad. As input we will use the types and features described in the previous section. Each feature is described by a minimum of 5 adjectives (found through an earlier pilot study) that relates to the specific semantic of the feature to be investigated, ensuring statistical significance [VM07]. In order to ensure we investigate the typical understanding of each type, the differential task for each type is preceded by a semantic priming task in which participants are primed onto their typical understanding or use of the type.

We report on a study investigating practitioners ( $n=12$ ), which was performed amongst employees of two internationally operating companies that provide support to clients with (re)design of organizations. The participants all had several years of experience as enterprise modelers, creating and using conceptual models and using diverse modeling techniques. Furthermore, we include results from an ongoing longitudinal study ( $n=10$ ) into the (evolution of) understanding computing and information systems science students have of modeling types. This study started when the students began their university studies and had little to no experience, and will continue throughout their studies.

Results from the semantic differential are processed to give an average score for each type-feature combination stemming from the individual adjectives used for that feature. This results in a vector for each type, containing a score (a numerical value ranging from 2.0 to -2.0) describing for each feature how it relates to that type. Scores  $\geq 1.0$  were considered positive judgments, scores  $\leq -1.0$  were considered negative judgments. Other scores were considered neutral. The judgments are then used to calculate a percentage-wise breakdown of the amount of different judgments for each concept.

### 4 Results

Results for both the practitioners and students are shown in Table 1. The percentages are an aggregate reflecting the amount of negative, neutral and positive results. The amount of neutral responses can be taken as a measure of how open people’s typical conceptualization of a concept is, in that it allows for more flexible (and possible amount of) instantiation. On the other hand, the negative and positive responses indicate that an instantiation of the concept would either need, or not need to display a certain feature. For example, when it comes to ACTORS, practitioners have 9% negative responses, 53% neutral and 38% positive. This means that about 9% of the responses indicate a feature that *has to be false* for a typical actor instantiation (e.g., a typical actor should not be a composed thing), 38% of the response indicate a feature that *has to be true* for a typical actor instantiation (e.g., a typical actor should be a human thing), while the remaining 53% allow for features to be either false or true (e.g., an actor can be either material or immaterial thing).

Table 1: Comparison of neutral and polar responses. Practitioners average 43.86% neutral responses, students 38.14%, 39.29% and 36.86%, being on average 5.75% less neutral

|                    | actor | event | goal | process | resource | restriction | result |
|--------------------|-------|-------|------|---------|----------|-------------|--------|
| Practitioners      |       |       |      |         |          |             |        |
| negative           | 9%    | 27%   | 22%  | 31%     | 30%      | 45%         | 45%    |
| neutral            | 53%   | 53%   | 52%  | 40%     | 47%      | 32%         | 30%    |
| positive           | 38%   | 19%   | 26%  | 29%     | 23%      | 22%         | 25%    |
| Students (phase 1) |       |       |      |         |          |             |        |
| negative           | 26%   | 38%   | 44%  | 34%     | 30%      | 37%         | 29%    |
| neutral            | 41%   | 44%   | 25%  | 36%     | 38%      | 43%         | 40%    |
| positive           | 33%   | 17%   | 32%  | 30%     | 32%      | 20%         | 32%    |
| Students (phase 2) |       |       |      |         |          |             |        |
| negative           | 16%   | 41%   | 34%  | 30%     | 27%      | 36%         | 39%    |
| neutral            | 47%   | 41%   | 33%  | 47%     | 41%      | 37%         | 29%    |
| positive           | 37%   | 17%   | 33%  | 23%     | 31%      | 27%         | 33%    |
| Students (phase 3) |       |       |      |         |          |             |        |
| negative           | 7%    | 44%   | 39%  | 31%     | 33%      | 43%         | 31%    |
| neutral            | 47%   | 44%   | 27%  | 37%     | 34%      | 33%         | 36%    |
| positive           | 46%   | 11%   | 34%  | 31%     | 33%      | 24%         | 33%    |

Table 2 shows a number of strongly polarized type-feature combinations. The percentages here denote how many of the participants judged the type to need (or need not) to display the given (semantic) feature. For example, 91% of practitioners judged goals to need the ‘necessary’ feature, which means that the majority conceptualize goals as things that need to be achieved (or perhaps gained, if goals are also conceptualized as material objects). It can also be seen that 64% of practitioners judged processes to be non-vague things, meaning that they are (or should) be well-described things. The scores for students are shown as the scores per phase of the longitudinal study, and are only included for those type-feature combinations for which they also display strongly polarized judgments.

## 5 Discussion

As briefly discussed in the previous section, the results as seen in Tables 1 and 2, directly affect the total amount of conceptually valid instantiations for a given type. The higher the amount of neutral scores, and vice-versa the lower the amount of polarized (and more to the point, strongly polarized) scores, the lesser features need to be exhibited by an instantiation of a type. Most interesting for our purpose are the polarized scores, as they indicate a feature that has to be exhibited by the type, and thus restricts the amount of conceptually valid interpretations for that type.

Table 2: Some strongly (>55%) polarized type-feature combinations in the practitioner sample.

| type        | is...       | score (practitioner) | score (student) |
|-------------|-------------|----------------------|-----------------|
| goal        | necessary   | 91%                  | 90%,70%,80%     |
| process     | necessary   | 64%                  | 90%,70%,70%     |
|             | not vague   | 64%                  |                 |
| resource    | necessary   | 64%                  | 80%,70%,80%     |
|             | not vague   | 64%                  | 30%,60%,80%     |
| restriction | not natural | 73%                  |                 |
|             | not human   | 73%                  | 70%,80%,60%     |
|             | necessary   | 73%                  | 40%,70%,60%     |
|             | not vague   | 82%                  | 90%,60%,90%     |
| result      | not natural | 64%                  |                 |
|             | not human   | 73%                  | 80%,80%,90%     |
|             | intentional | 73%                  |                 |
|             | not vague   | 91%                  |                 |

The results also show that around half of the judgments for the types are of a polar (either positive or negative) nature, indicating that there are many features for all investigated types that are conceived of as respectively strongly typical and strongly atypical. On an individual level this may vary, as specific people might find all features (a)typical for a given type, or find no features particularly typical for a given type (i.e., remain neutral). Such effects are likely correlated with the expertise a modeler has in a specific area, which might exhibit by an increased amount of detailed scrutiny he will give to the types often used by that area. The variation found within the results (see in Table 3) is only significantly different from the average when it comes to actors and events (meaning people are more likely to have similar conceptualizations of those types).

Table 3: Variance of results for each investigated type in the practitioner study

|          | actor | event | goal | process | resource | restriction | result |
|----------|-------|-------|------|---------|----------|-------------|--------|
| variance | 0.38  | 0.57  | 0.68 | 0.93    | 0.73     | 0.94        | 0.92   |

Interestingly enough there is not a significant difference in the relative amount of neutral and polar scores between the practitioners and the students. One could expect that experienced modelers have come across a larger variety of viewpoints (through e.g., the use of different languages and methods and working with a multitude of people), and thus would be more accommodating in their conceptualizations of the investigated types. This is, however, not the case. It might again be correlated with the experience modelers have, in the sense that their often specialized focus leads them to have strong opinions on a specific set of types (e.g., value modelers expressing very specific demands on the features that a resource can and should exhibit). Some specific strongly polarized results are shown for a number of types in Table 2. Only features for which more than 55% of the participants judged them to be necessary (positively or negatively) are taken into

account. This does not mean that a participant is not capable of considering an instantiation of the type that does not conform to these features, but that their typical interpretation (and likely implicit usage) conforms to this set of features. We will discuss some of them in more detail below.

**Goals** are one of the few types where there is but one absolutely needed feature. Specifically, goals are considered by most (91% of practitioners) to be necessary. What this means is that goals are considered to be logical necessities in sense of their being achieved, i.e., they are the logical consequence of following a certain procedure to achieve them, and should be achieved then (and likely only) then. This is in contrast to the other option of goals being things that ‘merely’ *should* be achieved. As a consequence, this seems to imply an implicit bias towards hard-goals, as soft-goals are often used to model goals for which the achievement status is (or cannot) be known a priori. The absence of any other needed features is interesting as well. It could have been expected that, in line with the implicit bias towards hard-goals, the majority of practitioners would consider goals as being non-vague (i.e., well-described), yet only 45% does so.

**Resources** are mostly interesting because of the trend seen in the scores of the investigated students. A slight majority of practitioners deem resources to be necessary and not vague. The necessary feature is consistently shared by students, yet the not vague feature was not shared from the beginning. Over time, students also seemed to deem not being vague a needed feature of resources (going from respectively 30% to 60% to 80% of the investigated students). This seems to imply that resources, are commonly seen as things that need to be well-defined, before they can be reliably (and systematically) used for some other activity. Such typical interpretations might clash with the way resources are handled in some languages (e.g., the information object in ArchiMate [The12], which allows you to represent a functional piece of information needed by some business process, without requiring you to be explicit about what information it contains).

**Restrictions** already have a larger typical feature set. They are typically considered non-natural, non-human, necessary and non-vague things. The non-vague feature seems obvious in that it demands us to be clear about what a restriction actually does (perhaps making rules more typical restrictions than something like, say, architecture principles). Non-natural clearly implies that they are not naturally occurring restrictions (e.g., natural laws like gravity) but more like restrictions that are ‘created’ to control or restrict other things. Related to this restrictions are also not considered to be human things, which in this context might mean that the general way of conceptualizing a restriction is seeing it as the abstract entity itself, and not, for instance, the human being enforcing it. More interesting from a modeling language point of view is the, perhaps obvious, view of restrictions being logically necessary things. This means that a typical restriction is something that has to be adhered to, and not something that can be broken. To accommodate that, a modeling language should either have its restriction types be inherently alethic, or allow for a way to explicitly distinguish restrictions with different modality. While there are not many (especially domain-specific) languages that accommodate this, Object Role Modeling (ORM) [Hal05] does have explicit support for denoting elements as being either alethic or deontic, and thus could be a useful choice of

language when a domain involves those non-typical restrictions: ones that do not necessarily have to be adhered to.

**Results** are typically non-natural, non-human, intentional and non-vague. This seems to fit with the common conceptualization of results, because they are seen as ‘new’ things, and thus shouldn’t be already existing naturally occurring objects. This requires a certain level of ontological scrutiny though, because a tree is certainly a naturally occurring entity, as well as a tree branch that naturally fell off from a storm, but a log that results from cutting the tree into parts with a chainsaw is not. Furthermore, a branch that merely breaks off from a tree would not be considered a result in our sense because its breaking off was not intentional. Some results as modeled by languages might seem not to fit and evoke discussion on their ontological status, as, for instance, the human output concept in ARIS [SN00]. On a first glance, it would seem not to fit the characterization as it is likely a human kind of thing. However, ontologically speaking, the actual thing here is the activity performed by a human, not the human itself. Say, the successful delivery of a pizza. As results are not typically considered material things, this seems thus to be a conceptually valid instantiation of a result.

To show in some more detail how certain combinations of features change the amount of conceptually valid instantiations for a given type some examples are shown in Table 4. In line with the explanation of context (in this case, features) changing the amount of conceptually valid instantiations we show how some of the common types are affected by the results we found.

Table 4: Examples of difference in conceptually valid reflecting some of Table 2’s results

| type        | is a ... feature-wise                                                | example                                                                                                                                                                                                             | ounter-example                                                                               |
|-------------|----------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| goal        | Necessary thing                                                      | Any thing to achieve, whether physical (e.g., producing a re- presentation of a model) or non-physical (e.g., producing the information needed for it) entity to be acquired or some state to be reached, and so on | A non-necessary goal, or thing that is achieved as a side-effect of achieving something else |
| restriction | Not natural, not human, necessary to adhere to and well-define thing | Legal rules given in a state’s laws, well-known and documented natural laws                                                                                                                                         | etiquette, informal dinner rules, or natural laws like gravity                               |
| result      | Not natural, not human, intentionally achieved well-define thing     | A baked pizza, a sawed log, a specification of a modeling language, a representation of the specification of a modeling language                                                                                    | A newly hired person, the outcome of trial-and-error testing                                 |

When a type has only few absolutely needed features, the amount of conceptually valid instantiations is significant. Conversely, the amount of conceptually invalid instantiations is quite small. A goal is a good example of this, as the only needed feature

is it being a necessary thing, which means we can instantiate it as an actual physical object to be achieved (i.e., the goal is the pizza that will be baked), the activity of having done so (i.e., the goal is the final state of the pizza baking process) or even more abstract, an abstract object that follows from its creation (i.e., the goal is the information gained when the pizza is baked). The few counter-examples are those where the goal is not necessarily achieved, which could be either something that is not achieved, or something that was achieved by accident (e.g., as a side-effect from another process). The main problem here is that the counter-examples (and examples themselves too, perhaps) are so broad that it is difficult to figure out whether an instantiation was actually a valid instantiation or not.

On the other hand, when a type has many needed features, it is easier to find out whether something is a conceptually valid instantiation or not. When it comes to results, certain objects are clearly valid instantiations, like a freshly baked pizza or a specification of a modeling language. The latter one can actually be a result in two ways, as both the actual specification of the language is a valid result, as well as the representation of the language (i.e., the actual written documents describing it). Counter-examples are then also easy to come by, for instance, the actual person hired as the ‘result’ of a hiring process is not typically considered a result (by virtue of being human). Furthermore, any ‘results’ that were the outcome of trial-and-error approaches are also on shaky grounds, as they are not typical results by virtue of not having been necessary.

## 6 Consequences

The results we have presented entail a number of consequences for the creation and use of models, and modeling language design itself. As it is clear that there are some common type-feature combinations amongst the people we investigated (while the total amount of conceptual diversity is far greater than that), it is important that we revisit the way we model in order to deal with the effects of these different conceptual understandings.

While the data presented and discussed here is reporting on average and shared conceptualizations, the conceptualizations that individual people have can vary wildly. It is important that we take a moment in the modeling process before we start creating models to deal with these differences. A way to do so is by looking at the types that are needed to model the domain under investigation, and then discuss whether certain features (and combinations thereof) apply to it. This can be done by taking known lists of features that have proven to be a source for disagreements and conceptual misunderstandings, such as the features reported on in this work. For example, when we are modeling the process of baking pizzas, we might focus on the resources for a moment. We can then take some time before creating actual models to discuss what we find conceptually valid resources and what we do not (i.e., what we would rather not be possible to express in the model). From such discussions we gain a clearer understanding of how people see such a process and want to model it, which helps us to focus (e.g., by abstracting from such potential resources as time, only modeling material things as resources). By doing so, the semantics of the created model are shared and well understood by all the people involved in the process, not just those who are intimately familiar with the specification of the used modeling language.

Such discussions can go into more detail feature wise as the focus on what to model becomes narrower. For example, when there is already a strong focus on processes, the types most relevant to a process (e.g., processes and restrictions) can be explored more closely by looking at more relevant features in greater detail. For instance, a process model might benefit from having a level of traceability on the restrictions in it [PdKP12], which requires us to discuss whether restrictions are always made by specific people, or departments, and whether that implies that restrictions carry responsibility with them. These aspects require a more fine-grained characterization of the process type than if it is used in a general way. Equally so, when the focus is on modeling value exchanges, then the focus can be on, for instance, characterizing types like resource and actor more clearly by investigating more detailed features (e.g., value exchanges can be analyzed from an ethical point of view, and thus require actors to be moral agents).

As it is clear that there are a number of conceptual distinctions people have for the types we investigated, it is also important that the modeling languages we use allow us to express these distinctions. This can be ensured in a practical way by selecting a modeling language which allows us to explicitly express those distinctions that are important to us. For example, when it is important to make a distinction between alethic and deontic restrictions we could choose for ORM, when it is important to make a distinction between goals that are well-defined or not, we could choose for  $i^*$  or any of the goal modeling languages that explicitly support hard- and soft goals. In short, before selecting a modeling language and creating models, we need to find out the most important conceptual distinctions the people involved in the modeling process have, and then select a language based on them.

On a more theoretical side, we should also ensure that the modeling languages support these different conceptual distinctions in the specification of their semantics. While this would not necessarily be useful for general purpose languages (e.g., UML, ER), domain-specific languages could benefit from the added detailed semantics. This requires more fine-grained investigations into the conceptualization people have of types as they are used in specific domains. For example, a more detailed characterization of what a resource is could be useful to languages that center around value exchanges (e.g., e3Value), implementations (e.g., ITML) or deployments (e.g., ADeL [Pat11]).

Finally, while the results we have shown are interesting and could be used for a multitude of purposes, care must be taken not to immediately extrapolate the results of this relatively small study and infer general truths from them. Different groups with varying backgrounds and specializations might have different conceptualizations, and thus further studies might uncover additional and conflicting conceptualizations. As such, the contribution of this work has been in showing the demonstrating the possibility to systematically uncover such conceptual distinctions and laying the ground for further investigations.

## **7 Conclusion & Future Work**

We have shown how the conceptualization which someone has of types used by modeling languages can affect the range of conceptually valid instantiations for such types. This can have a negative effect on the validity and usefulness of created models if

we do not take care to discover and communicate these conceptualizations. Furthermore, some languages might be more suited to deal with certain conceptualizations by virtue of explicitly expressing conceptual distinctions that other languages leave implicit. Care should thus be taken to also select a modeling language that fits most with the conceptualizations of the involved modelers and stakeholders. Having shown the potential use of these kind of studies, we aim to extend their thoroughness for generalization purposes by repeating them with a larger amount of practitioners, and also by incorporating groups focused on specific sectors (e.g., modelers in healthcare, government, or telecom) and languages (e.g., the BPM community or the ArchiMate community).

## Acknowledgments

This work has been partially sponsored by the *Fonds National de la Recherche Luxembourg* ([www.fnrl.lu](http://www.fnrl.lu)), via the PEARL programme.

## References

- [Alm09] Mauricio Barcellos Almeida. A proposal to evaluate ontology content. *Applied Ontology*, 4:245–265, August 2009.
- [BC05] Peter D. Bruza and Richard Coles. *We Will Show Them: Essays in Honour of Dov Gabbay*, chapter Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning. College Publications, 2005.
- [BKV09] Joseph Barjis, GwendolynL. Kolfschoten, and Alexander Verbraeck. Collaborative Enterprise Modeling. In Proper et al., editor, *Advances in Enterprise Engineering II*, volume 28 of *LNBIP*, pages 50–62. Springer Berlin Heidelberg, 2009.
- [BW08] Peter D. Bruza and John Woods. Quantum collapse in semantic space : interpreting natural language argumentation. In Peter D. Bruza et al., editor, *Second Quantum Interaction Symposium*, 2008.
- [FCK95] David Ferrarolo, Janet Cugini, and Richard Kuhn. Role-based access control (RBAC): Features and motivations. In *Proc. of the 11th Annual Computer Security Applications Conference*, 1995.
- [FHK<sup>+</sup>09] Ulrich Frank, David Heise, Heiko Kattenstroth, Donald Ferguson, Ethan Hadar, and Marvin Waschke. ITML: A Domain-Specific Modeling Language for Supporting Business Driven IT Management. In *Proc. of the 9th OOPSLA workshop on DSM*, 2009.
- [FHV12] Peter Fettke, Constantin Houy, Armella-Lucia Vella, and Peter Loos. Towards the Reconstruction and Evaluation of Conceptual Model Quality Discourses – Methodical Framework and Application in the Context of Model Understandability. In Bider et al., editor, *Enterprise, Business-Process and Information Systems Modeling*, volume 113 of *LNBIP*, pages 406–421. Springer Berlin Heidelberg, 2012.
- [Gee10] Dirk Geeraerts. *Theories of Lexical Semantics*. Oxford University Press, 2010.
- [GHYA07] Gemma Grau, Jennifer Horkoff, Eric Yu, and Samer Abdulhadi. i\* Guide 3.0. Internet, August 2007.
- [GRA08] Liane Gabora, Eleanor Rosch, and Diederik Aerts. Toward an Ecological Theory of Concepts. *Ecological Psychology*, 20(1):84–116, 2008.
- [GYvdR06] Jaap Gordijn, Eric Yu, and Bas van der Raadt. e-Service Design Using i\* and e3value Modeling. *IEEE Software*, 23:26–33, 2006.
- [Hal05] Terry Halpin. ORM 2. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, pages 676–687. Springer, 2005.

- [HFL12] Constantin Houy, Peter Fettke, and Peter Loos. Understanding Understandability of Conceptual Models - What Are We Actually Talking about? In Paolo Atzeni, David Cheung, and Sudha Ram, editors, *Conceptual Modeling*, volume 7532 of *Lecture Notes in Computer Science*, pages 64–77. Springer Berlin Heidelberg, 2012.
- [HS05] Brian Henderson-Sellers. UML - the Good, the Bad or the Ugly? Perspectives from a panel of experts. *Software and System Modeling*, 4(1):4–13, 2005.
- [KSKS12] Julia Kaidalova, Ulf Seigerroth, Tomasz Kaczmarek, and Nikolay Shilov. Practical Challenges of Enterprise Modeling in the Light of Business and IT Alignment. In Sandkuhl et al., editor, *The Practice of Enterprise Modeling*, volume 134 of *LNBIP*, pages 31–45. Springer Berlin Heidelberg, 2012.
- [Lan04] Marc M. Lankhorst. Enterprise architecture modelling—the issue of integration. *Advanced Engineering Informatics*, 18(4):205 – 216, 2004.
- [MAG<sup>+</sup>11] Barbara C. Malt, Eef Ameel, Silvia Gennari, Mutsumi Imai, and Asifa Majid. Do words reveal concepts? In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 519–524, 2011.
- [OB06] Andreas L. Opdahl and Giuseppe Berio. Interoperable language and model management using the UEML approach. In *Proceedings of the 2006 International Workshop on Global Integrated Model Management*, pages 35–42, New York, NY, USA, 2006.
- [Obj10] Object Management Group. Business Process Model and Notation (BPMN) FTF Beta 1 for Version 2.0. Internet, 2010.
- [OST57] Charles E. Osgood, George J. Suci, and Percy Tannenbaum. *The Measurement of Meaning*. University of Illinois Press, Urbana, 1957.
- [Pat11] Susanne Patig. Modeling Deployment of Enterprise Applications. In Wil Aalst et al., editor, *Information Systems Evolution*, volume 72 of *LNBIP*, pages 253–266. Springer Berlin Heidelberg, 2011.
- [PdKP12] G. Plataniotis, S. de Kinderen, and H.A. Proper. EA Anamnesis: Towards an approach for Enterprise Architecture rationalization. In Sheridan Printing, editor, *Proceedings of the The 12th Workshop on Domain-Specific Modeling (DSM12)*. ACM DL, 2012.
- [Pin07] S. Pinker. *The stuff of thought: Language as a window into human nature*. Viking Press, 2007.
- [RGS<sup>+</sup>08] M. Rospocher, C. Ghidini, L. Serafini, B. Kump, V. Pammer, S. Lindstaedt, A. Faatz, and T. Ley. Collaborative enterprise integrated modelling. In *5th International Workshop on Semantic Web Applications and Perspectives (SWAP 2008)*, 2008.
- [RHRL11] Michael Rosemann, Anders Hjalmarsson, Mikael Lind, and Jan C. Recker. Four facets of a process modeling facilitator. In Dennis Galletta and Ting-Peng Liang, editors, *32nd International Conference on Information Systems*, pages 1–16, Shanghai International Convention Center, Shanghai, December 2011. AIS.
- [RKV08] Michiel Renger, Gwendolyn L. Kolfschoten, and Gert-Jan Vreede. Challenges in Collaborative Modeling: A Literature Review. In Jan Dietz et al., editor, *Advances in Enterprise Engineering I*, volume 10 of *Lecture Notes in Business Information Processing*, pages 61–77. Springer Berlin Heidelberg, 2008.
- [Rob06] Stewart Robinson. Conceptual modeling for simulation: issues and research requirements. In *Proceedings of the 38th conference on Winter simulation*, WSC '06, pages 792–800. Winter Simulation Conference, 2006.
- [SHP09] Denis Ssebuggwawo, Stijn Hoppenbrouwers, and Erik Proper. Interactions, Goals and Rules in a Collaborative Modelling Session. In Anne Persson and Janis Stirna, editors, *The Practice of Enterprise Modeling*, volume 39 of *Lecture Notes in Business Information Processing*, pages 54–68. Springer Berlin Heidelberg, 2009.
- [SN00] August-Wilhelm Scheer and Markus Nuettgens. ARIS Architecture and Reference Models for Business Process Management. In *Business Process Management*, volume 1806 of *LNCS*, pages 301–304. Springer Berlin / Heidelberg, 2000.

- [Sow10] John Sowa. The Role of Logic and Ontology in Language and Reasoning. In *Theory and Applications of Ontology: Philosophical Perspectives*, pages 231–263. Springer Netherlands, 2010.
- [SPS07] Janis Stirna, Anne Persson, and Kurt Sandkuhl. Participative Enterprise Modeling: Experiences and Recommendations. In Krogstie et al., editor, *Advanced Information Systems Engineering*, volume 4495 of *LNCS*, pages 546–560. Springer Berlin Heidelberg, 2007.
- [Sta00] Gerry Stahl. A model of collaborative knowledge-building. In *Proc. of 4th Int. Conference of the Learning Sciences (ICLS 2000)*, Ann Arbor, MI, pages 70–77, 2000.
- [The12] The Open Group. *ArchiMate 2.0 Specification*. Van Haren Publishing, 2012.
- [vdLHLP11] D. J. T. van der Linden, S. J. B. A. Hoppenbrouwers, A. Lartseva, and H. A. Proper. Towards an Investigation of the Conceptual Landscape of Enterprise Architecture. In T. Halpin et al., editor, *Enterprise, Business-Process and Information Systems Modeling*, volume 81 of *LNCS*, pages 526–535. Springer Berlin Heidelberg, 2011.
- [VM07] Tibert Verhagen and Selmar Meents. A Framework for Developing Semantic Differentials in IS research: Assessing the Meaning of Electronic Marketplace Quality (EMQ). Serie Research Memoranda 0016, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, 2007.

# Human-Oriented Challenges of Social BPM: An Overview

Nicolas Pflanzl and Gottfried Vossen

DBIS Group  
WWU Münster  
Leonardo-Campus 3  
48149 Münster, Germany  
vossen@uni-muenster.de, nicolas.pflanzl@wi.uni-muenster.de

**Abstract:** Social BPM is the practice of actively involving all relevant stakeholders into BPM through the use of social software and its underlying principles. This allows enhancing the correctness, completeness, and usefulness of process models and instances by leveraging the domain and method knowledge of the entire business community. However, the large number and variety of contributors and contributions also results in a number of challenges, which will be examined in this paper.

## 1 Introduction

Today, businesses must be willing to constantly revisit and improve their business practices to remain profitable in the light of challenges such as global competition and cost reduction. This can be achieved through business process management (BPM), which aims to ensure a high level of cost- and resource-efficiency, speed, accuracy, and flexibility of business operations [Ham10]. Traditional BPM approaches have been found to suffer from several shortcomings, such as a divergence between process models and execution reality, and the failure to leverage ideas for innovation held by process end-users [SN09]. To solve these problems, it is necessary to re-envision the management of business processes as a task that is carried out by the entire business community rather than a few individuals [SVOK11]. Social BPM aims to achieve this goal by establishing an “architecture of participation” for BPM through the use of social software and its underlying principles [EGH+10]. This paper examines human-oriented challenges that arise when such an approach is employed, i. e., challenges that arise either as a precondition for, or as a consequence of the inclusion of the entire business community.

Typically, most business process management activities are performed by method experts and IT developers according to requirements specified by domain experts and end-users who eventually enact processes with the support of information systems [SR12]. However, it can be argued that the correctness, completeness, and ultimately usefulness of a process can only be maximized if all relevant stakeholders participate in all phases of its life cycle, including end-users, and possibly even customers and external suppliers [SVOK11, Sch12]. This can be achieved through social BPM, which enables a large variety of internal and external actors with varying degrees of BPM training to contribute their domain knowledge and method expertise. By working together in large numbers, the business community can leverage its “wisdom of the crowd” [Sur05] to create better and higher-quality process solutions than a single expert can achieve alone. However, this can

only be successful if certain challenges are overcome, such as initiating and ensuring the ongoing participation of relevant actors. Solving this problem might not be straightforward and can affect other areas, thereby resulting in new challenges. For instance, the individuals that can make the most valuable contributions are not necessarily those with the most profound knowledge of BPM and therefore need to be supported through appropriate methods and technology. It can be seen that employing social BPM may have far-reaching consequences and requires overcoming problems in diverse areas, which may stem from it being a research topic orthogonal to other issues of BPM [vdA13].

This paper aims to answer the following question: What are the most important human-oriented challenges of social BPM currently recognized in literature that need to be overcome for its successful application? For this purpose, its remainder is organized as follows. In Section 2, previous work on which this article builds is named. Afterwards, Section 3, establishes the theoretical background of the paper. Section 4 outlines the drivers of social BPM, provides a definition, and describes its underlying principles. Section 5 examines human-oriented challenges that result from the application of social BPM. Lastly, a brief summary and an outlook are presented in Section 6.

## **2 Related Work**

Previous publications that this paper builds on can broadly be classified in two types, namely literature on social BPM in general, and literature addressing particular challenges, key concerns, and success factors of social BPM and other related fields.

The foundations of social BPM as a distinct research topic lie in the Workshop on Business Process Management and Social Software (BPMS2), which has been carried out in conjunction with the International Conference on Business Process Management since 2008 [NS09]. Notable publications resulting from these workshops subsequently published in journals include [EGH+10] and [BDJ+11]. Furthermore, social BPM has also been recognized as an important topic by practitioners and is discussed from various perspectives in the book “Social BPM - Working, Planning, and Collaboration Under the Impact of Social Technology” [Ric11].

A considerable amount of publications has, to some extent, addressed or raised awareness about challenges relevant for this paper, including social BPM literature (e. g., [BDJ+11], [EGH+10] and [JAW09]), and literature on BPM in general (e. g. [BICS07], [IRRG09] and [PCBV10]). Beyond these works, it is also necessary to consider publications that do not directly address social BPM, but concern other topics highly related to the former. This includes the following fields. Firstly, social BPM can be seen as part of the “Web 2.0” and “Enterprise 2.0” paradigm [VH07,KR09], and as such may also be affected by challenges resp. profit from solutions offered by literature in that area. Secondly, it is also related to inter-organizational, collaborative BPM initiatives which are often carried out in a top-down controlled fashion and need to overcome certain barriers [NH11, NP11]. Lastly, social BPM can also be interpreted as an application of crowdsourcing. It is thus subject to challenges specific to that field, such as those outlined in [DRH11], as well.

### 3 Background

**Social software** is defined as “software that gets better the more people use it” [VH07]. From a more technical perspective, it represents a class of Web-based applications for humans that support the exchange of information, the establishment of relationships and the communication in a social context [Hip06]. While such tools are often used with the intention of reaching a specific goal, this does not always have to be the case [JAW09]. SCHMIDT and NURCAN determine the purpose of social software as supporting the generation of digital goods that combine the contributions of multiple individuals who may not know each other and who are not initially organized in a hierarchical structure [SN09]. Social software is thus commonly seen as part of the “Web 2.0”, a term coined in 2004 describing then recent developments and applications of Web technologies that facilitate and encourage the voluntary and active participation of end-users in content creation [VH07]. These ideas have first found widespread use in personal contexts, but are nowadays increasingly employed by companies as well [KR09]. One very popular type of social software are Wikis. Typically, the users of such a collaborative authoring tool are not organized in any predetermined way and have equal editing and viewing rights. This allows documenting topics of interest exhaustively by making not only common knowledge available, but also knowledge held by only a few experts. Lastly, Wiki pages are dynamic by definition and are continuously expanded or revisited as new knowledge becomes available. Consequently, the quantity and quality of content managed in a Wiki increases with the number of active participants. Further types of social software include social tagging, blogs, microblogs, social networks and instant messaging [KR09].

**Business process management** is an approach for maintaining business performance through the management of business processes [Ham10], which are “[sets] of activities that are performed [...] to jointly realize a business goal” [Wes12]. It “includes concepts, methods and techniques to support [their] design, administration, configuration, enactment, and analysis [...]” [Wes12] and is supported through software called business process management systems [Wes12]. On a broader scale, BPM has also been described as a holistic approach to the management of an enterprise [SVOK11] that addresses the entire business layer and consists of six core elements [RvB10]: *strategic alignment, governance, methods, information technology, people, and culture*. When performed successfully, BPM allows companies to “create high-performance processes, which operate with much lower costs, faster speeds, greater accuracy, reduced assets, and enhanced flexibility” [Fehler! Verweisquelle konnte nicht gefunden werden.]. Its purpose is thus the continuous improvement of business operations through incremental or radical change of business processes [RvWML10]. Activities related to business process management are typically arranged as an iterative life cycle. Most BPM life cycles are very similar and implement a basic Plan-Do-Check-Act approach with steps such as *design, engineering, enactment, monitoring, and re-engineering* [SVOK11, Wes12].

### 4 Social Business Process Management

As mentioned in the Introduction, traditional BPM approaches suffer from several shortcomings that motivate the need for a social approach. First, current BPM

methodologies often follow a top-down decomposition approach from the strategy and business goals of an enterprise down to the level of individual business processes [SR12]. As a result, process improvement is a long-running process in itself that requires time-intensive negotiations and compromises, and restricts opportunities for implementing change to specific points in the BPM life cycle [WHR+10, BDJ+11]. This can become a problem when companies must be agile to react to new and unpredicted developments in a timely fashion [Ric11]. To ensure that business processes meet current needs, changes to a process model should be possible at any time and quickly reflect themselves in practice. Secondly, many aspects of process management have been conceptualized as activities carried out by a small number of experts. This is especially true for the process of process modeling and its three stages elicitation, modeling, and validation [RMH13]. In such an approach, process end-users are not empowered to directly integrate their knowledge and experiences from past enactments into process design [SR12]. However, as knowledge about current practices and improvement potentials is distributed across a potentially large number of different stakeholders, this is not suitable for corporate reality [RMH13]. If these issues are not addressed sufficiently, the following problems may arise [SN09, EGH+10]:

- *Model-reality divide*: This term describes a state of divergence between “planned” and “real” processes, i. e., between idealized models and the way in which processes are actually executed. This phenomenon occurs when end-users do not accept the created models because they do not reflect their day-to-day work.
- *Lost innovation*: In organizations that employ a top-down approach to BPM, employees may refrain from sharing ideas for process improvement and innovation, because the guidelines for change management are seen as too intransparent and restrictive and the chance of success considered too small. Therefore, this knowledge is either lost entirely, or applied on the local scale of individual process instances, thereby once again increasing the model-reality divide.

These problems can potentially be solved by *social BPM*. Today, there is no common understanding of what the term exactly entails [Puc11, Ric11]. From a purely technological point of view, social BPM can be seen as providing new tools for activities in the BPM life cycle by adding social software features to conventional BPM systems [SN10, Puc11, DC11]. For instance, a Wiki could be used for process design and documentation, a microblogging service to broadcast the execution status of a specific process instance, or a social network to find the best partner for collaboration on a specific task. However, while social software as enabling technology is of course an important aspect of social BPM, practitioners have recognized that reaping its potential benefits also requires embracing its underlying ideas, which will often result in organizational change [Puc11]. Therefore, for this article social BPM is defined as *the involvement of all relevant stakeholders in the BPM life cycle by applying social software and its underlying principles*. These principles are (cf. [SN09, VFL10, SN10]):

**Self-organization.** A system is self-organizing if it is capable of ensuring and refining its functionality through cooperation of its components without external influences [VFL10]. In this vein, social BPM is not regulated, and planning and control are performed by the business community in a democratized, bottom-up fashion rather than top-down [Hip06].

This is accomplished, e. g., by the signing and versioning of work activities, social feedback through discussions and ratings, and banning malicious actors [Hip06, JAW09].

**Egalitarianism.** All users are equal and possess equal rights [SN10]. This entails the transparency and open modification of contributions, i. e., all work results cannot only be viewed, but also edited by all other actors within the limits of reason [JAW09]. Consequently, no distinction is made between the roles of method and domain experts. Specifically, social BPM must avoid excluding any individuals from participation, be it explicitly due to not being part of a specific organizational unit, or implicitly due to not having the necessary method knowledge and training [BDJ+11]. Thus, the usability of BPM software for BPM novices also plays an important role. To ensure the desired level of quality, social BPM does not rely on access control but on trust and reputation [JF09, SN10].

**Collective intelligence.** Social BPM is based on the idea that the collective wisdom of a crowd can create better process solutions than individual experts alone [Sur05, SN09]. This requires users to establish and maintain relationships with one another and to perform many-to-many interactions rather than working in isolation [Hip06]. Sporadic connections of low emotional or temporal intensity, also called *weak ties*, are of special importance in this context as they form the “long tail” of knowledge and innovation accessible that would not be captured without them [SN09]. Leveraging the collective intelligence of a business community can only be accomplished if all relevant participants are actually included and their needs considered. Therefore, it is necessary to create an organizational environment that enables and fosters continuous contributions by all stakeholders [JAW09].

**Social production.** By using social software, individuals create *content* such as text and diagrams, and *context* information in the form of annotations, reputation and social links, which are both considered valuable [SN09, EGH+10]. These artifacts are developed by all participants interactively and all actors who consider themselves competent to contribute are enabled to do so [JAW09]. New information is continuously assessed and fused with the existing body of knowledge to perform a selection of the best available ideas. Additionally, changes become effective and visible immediately, thereby enabling an agile cycle of process improvement without any unnecessary delays between the detection of a desirable change and its transfer to the practice of process enactment [SN10, EGH+10, Rus11].

By respecting these principles, business process management can be re-envisioned as a task that is carried out by the entire business community rather than a few individuals [SVOK11]. This allows all stakeholders to contribute their specific method or domain knowledge about business practices and innovation potentials, which is one of the biggest opportunities of social BPM. Furthermore, the continuous integration and immediate effectiveness of new contributions provides the BPM life cycle with more agility, and thus the ability to react to internal and external events more quickly [SN09, BDJ+11]. As a result, the model-reality divide can be closed and lost innovation avoided.

## 5 Challenges

Compared to traditional BPM approaches, social BPM enables the participation of a much larger and more heterogeneous set of actors. However, such an approach can only be successful if a number of human-related challenges are overcome. While some of these issues should be addressed before social BPM is even initiated, other challenges become more important with an increasing number and variety of contributors and contributions. The following list of challenges has been compiled from an extensive review of the literature named in Section 2, enriched by keyword-based searches in relevant academic databases. The challenges have been extracted manually during the review, with a focus on those with a high perceived importance and specificity for social BPM. This list is not supposed to be finite or complete. However, the scope of the literature review has allowed capturing several important current challenges. For each challenge, further keyword-based as well as forward- and backward-searches based on the literature named in Section 2 were performed to obtain initial insight into how it is currently being addressed. Due to the subjective nature of the selection process, the list of challenges may be subject to extension and modification. Furthermore, it is not exclusive to social BPM, but may apply to top-down collaborative BPM approaches and other aspects of the Enterprise 2.0 as well.

### 5.1 Involving External Stakeholders

There is a broad consensus that integrating a wide variety of individuals from different backgrounds into BPM yields a higher rate of success and acceptance [NP11]. Through inter-organizational inclusion, boundaries between different enterprises can be blurred, and thus the definition of business processes expanded beyond the scope of a single company. Ultimately, this allows for an unhindered exchange of ideas and information and thus helps with closing the model-reality divide and avoiding lost innovation on a broader scale [NP11]. However, external stakeholders typically have their own interests and motivation for participation, and thus selecting the right actors at the right time for the right type of contribution is a challenging issue [VB10]. This may not always be straightforward, as there are certain factors that reduce the attractiveness of inter-organizational collaboration [NH11]. One of the most significant barriers is the fear of knowledge loss, i. e., of losing process knowledge as an important source of competitive advantage to competitors. Furthermore, if companies are not aware of potential benefits that may result from involving all stakeholders, they are not motivated to initiate or participate in such efforts. Lastly, boundary-spanning BPM is often perceived as being very costly due to the large number of transactions and time that it requires. These issues affect not only the focal enterprise, but external stakeholders as well.

To extend the reach of social BPM, research should find rules for good inter-organizational involvement, i. e., how, when, and which external actors should be included. Furthermore, it should be examined if the aforementioned barriers apply, and how to overcome them.

### 5.2 Motivating Participation

Clearly, the most important ingredient to social BPM are the stakeholders who are willing to contribute their own knowledge. Therefore, for such an approach to be successful, it is essential that all users who can make meaningful contributions are motivated to invest

their time and effort continuously to create and maintain process models and other content. While having only a small number of contributors who are highly active may be a valid scenario in some cases, this contradicts the principles of social software and thus presents a risk for closing the model-reality divide and avoiding lost innovation [EGH+10].

The first step towards continuous participation is accomplished by reaching a *critical mass* of users. This is necessary due to the fact that simply providing employees (and other stakeholders) with a platform for social production does not necessarily lead to its adoption without any meaningful initial content. Thus, companies have to carefully select the right individuals for the initiation of social BPM that can provide the required digital artifacts [KR09]. Until this “ramp-up” phase has been completed, the usefulness of social software may not reach its highest possible level [EGH+10]. Subsequently, it must be ensured that participants are motivated to continue with their involvement. This is a difficult task because the usage of social software is very time consuming and should be voluntary rather than being enforced [EGH+10]. Solving this problem requires either ensuring that employees are interested in the work itself and its results (intrinsic motivation), or providing them with monetary remuneration in dependence of some indicators of performance (extrinsic motivation). It has been noted that the latter may actually have a negative effect on participation and thus the former should be preferred [KR09]. While there are many different possibilities to invoke intrinsic motivation, the tools most often discussed in social BPM literature are based on the practice of *gamification* [Rit10]. For instance, users could be rewarded with “honor points” according to their contributions [JAW09]. The points themselves already provide incentive due to the fact that they are a measure of status and reputation, but can also be used as an input for extrinsic motivation [KR09].

As of yet, academics still have to address factors influencing the continuous participation in BPM on a broader scale [IRRG09]. Consequently, social BPM research should focus on finding and evaluating measures that reinforce the motivation of potential contributors. Furthermore, it should be examined how the “ramp-up” phase can be completed successfully and without unnecessary delays.

### 5.3 Training Novice Modelers

Modeling is an important part of social BPM, as it allows process end-users to make their process knowledge explicit without relying on method experts and long-running improvement cycles. However, experts often argue that novice modelers do not possess the necessary skills for this purpose, as modeling software, modeling languages and models themselves are too difficult to use for untrained individuals [NP12]. As a consequence, even though novices are allowed to contribute their own knowledge and experience in theory, they might not be able to do so in practice [SN09]. One way to overcome this problem is teaching contributors the skills required for successful participation. Measures for BPM education include teaching it at university, which is an opportunity for integrating it firmly within the mindset of future employees, and thus, organizational culture [BICS07]. In the short term, [RMH13] suggests that real-time collaborative process modeling can also serve as a measure for the training of employees as it animates the exchange of information about the modeled domain as well as modeling expertise. In this way, more experienced actors can help novices to increase their level of

knowledge, and thus enable them to make more contributions with a higher quality. Following this idea, a promising approach for familiarizing stakeholders with social BPM may be the organization of social BPM labs<sup>5</sup>. In such workshops, participants use social software and groupware to collaboratively model the processes of a business community over the span of one or two days. Bottom-up self-organization is combined with top-down guidance through, e. g., moderators, quality managers, and method experts to ensure that the learning goals of the lab are met. Ultimately, this not only teaches the required skills, but may also promote social BPM and increase the motivation for active participation.

The need for teaching BPM has been recognized and discussed in various publications, including [BICS07] and [IRRG09]. However, it is currently a topic addressed mostly by practitioners, with academics being more focused on challenges concerning methods and technology. Therefore, social BPM research should address this topic by examining which skills contributors require, and how they can be taught best.

#### 5.4 Providing Modeling Tools for Novices

Besides training, novice participation in process modeling can also be facilitated by tools that compensate for a lack in the former, such as highly usable modeling software and expressive, yet non-overwhelming modeling languages.

**Modeling software.** Usability has been shown to be one of the most highly valued qualities of BPM tools, which is true for novice users as well as modeling experts [PCBV10]. Inexperienced participants can further be supported by user interface agents that guide them through specific modeling tasks [Dry97], the automatic detection, prevention, and recovery of syntactic errors [KJ03], and the implementation of pragmatic guidelines that facilitate the creation of understandable models (see Section 5.5). The usability of modeling software can also be enhanced through social software features. For instance, social recommendations can provide actors with suggestions on how to proceed with modeling a specific business process based on other existing models, their past behavior and that of other users, as well as inter-personal relationships [SWMW09,KSR09]. Research on social BPM should address factors that make modeling software usable for novices, and the role of social software in this context.

**Modeling languages.** Examining the preferences and task performances of novice modelers when recording their process knowledge using different modeling archetypes, [RSR12] has found the “flowchart” to yield the most favorable results. As this type of diagram is very similar to conventional modeling techniques, it is reasonable to assume that the latter are also suitable for users with little modeling experience. Another consideration is the required *expressiveness* of the modeling language. Many languages offer a large variety of different notational elements to cover an equally large variety of modeling scenarios. However, not all of these elements are frequently used in practice [zMR08], and therefore it may be reasonable to provide a reduced set of elements with smaller expressiveness to novice users [Rit10]. Further research should examine how social BPM participants can be provided with the diagram types and subsets of modeling languages that match their task and level of capabilities.

---

<sup>5</sup> see <http://www.horus.biz/en/social-bpm/horus-social-bpm-lab.html> (last accessed 2013/06/05)

## 5.5 Ensuring Model Quality

The quality of a process model is a multi-faceted property that determines whether it can successfully be used for its intended goal. Maintaining this property has been recognized as a difficult task in modeling efforts that include non-expert users [MRvdA10] and is thus of special concern in the context of social BPM [EGH+10]. Based on three of the higher levels of the SEQUAL framework [KJ03], this section will examine the interrelationships between social BPM and model quality.

**Semantic quality** is given if a model is *valid* by containing only relevant and correct statements about a domain, and *complete* by containing all relevant and correct statements. In practice, this is difficult to achieve and thus has to be judged in relation to the modeling goal. Social BPM offers great opportunities for a high semantic quality as it includes a large variety of stakeholders and their in-depth knowledge about different parts of a domain. By integrating these contributions, reasonably complete and valid process models can be created, thereby diminishing the model-reality divide.

**Pragmatic quality** is given if a process model can be *understood* and is thus consistent with its interpretation by humans, which allows the reader to learn from it. [MRvdA10] have proposed seven pragmatic modeling guidelines intended to help non-expert users with the creation of understandable models, such as “use as few elements as possible” and “avoid OR routing elements”. To improve the adherence to these guidelines, modeling software can be extended with functionality that warns modelers of any violations.

**Social quality** requires participants to reach a shared understanding about a process model as the result of social learning. It is thus characterized by a convergence of the knowledge and model interpretations of all actors. This does not necessarily imply that a consensus is reached, but merely that inconsistent views are resolved where the benefits exceed the costs. Ensuring social quality not only reduces inconsistencies, but can also help with the detection of harmful contributions made by malicious actors. Clearly, social software can be an important enabler for social quality and many application scenarios have been studied in social BPM literature. For instance, [SMM+10] proposes a Wiki-like system that allows users to rate and comment on individual models.

Social BPM research should analyze how the use of social software and its underlying principles can positively and negatively affect the quality of business processes. Furthermore, the role of trust and reputation for ensuring social quality should be examined.

## 5.6 Handling Information Overload

In a successful social BPM setting with many participants and a high frequency and quantity of contributions, finding relevant people and content can become a significant challenge [JF11]. This problem is also known as *information overload* and, if not addressed, may result in users overlooking content to which they would otherwise contribute, providing inaccurate contributions, or abandoning the system altogether [HT85]. To cope with this situation, *filtering* important from unimportant information based on individual criteria and priorities must be possible. [HT85] points out that filtering should never be dependent on content, but only determine which content will be

shown. Therefore, the following paragraphs will briefly analyze how context information (cf. Section 4) can be used for this purpose.

**Annotations** are meta-data that provide additional information about contributions and can thus be used for search and evaluation [EGH+10]. A social approach to annotation-based filtering can be realized with the use of *social tagging*. By adding tags to process models, parts thereof, or other digital artifacts, users collaboratively provide machine-readable details about their meaning. This information can be used for filtering, e. g., by searching for models or model parts within larger repositories using tags [SMM+10], or finding similar models through tag-based similarity measures [LB13].

**Reputation** or ratings are a special type of annotation that support the decision whether to trust the validity of an object or the contributions of a specific actor. Providing a quantifiable measure of reputation is important for social BPM as users may not necessarily know each other [EGH+10]. To mitigate the effects of information overload, process models can be filtered and prioritized according to their rating or the reputation of their creators [DC11]. As a result, participants can choose to only consider contributions that have been judged to be of a high quality, or conversely, search for content of a low quality and help to improve it. Social BPM literature has examined different types and uses of reputation, such as rating the contributions of others to create a list of “top authors” [QJ08].

**Social links** are relational information that describes how humans connect themselves to form social networks [EGH+10]. Participants can use them to search for relevant or interesting contributions made by users that are, e. g., structurally close within the network, popular by having many relationships, or assume a specific role such as being distributors of information or bridges between different network parts. If a social networking site is used in the context of social BPM, participants have the opportunity to make these connections explicit by “befriending” other actors. Further types of relationships can be discovered from data in which they are implicitly contained, such as the event logs of software systems used for process enactment [dARS05].

To ensure that potential contributors can always find the content to which they would like to contribute or that they are interested in, further research should be carried out on the possible uses of context information.

## 5.7 Integrating Semantics

When individuals with diverse backgrounds collaborate, it is very likely that they will use different terminology to refer to similar concepts. This may result in misunderstandings and can ultimately have a negative effect on process performance [BDJ+11]. Most importantly, since social BPM integrates the contributions of people with varying levels of education, roles, and tasks, a so-called *language gap* may arise, as for each person, terms may hold different meanings [MS11]. For instance, this gap can manifest itself in a diverging use of synonyms and homonyms [BDJ+11]. Therefore, to achieve semantic integration, a common vocabulary of terms and their semantics must be established and maintained through social production. Different proposals on how to achieve this with social software can be found in social BPM literature, including the creation of a BPM folksonomy by means of social tagging [MS11], and the usage of Wikis with semantic extensions [BDJ+11]. However, it is not inherently clear how the work results should be

used. Therefore, additional measures must be defined and implemented to integrate them, e. g., into process modeling or textual discussions for conflict resolution.

## 6 Conclusion

By employing a social approach to BPM, companies can establish an “architecture of participation” that enables and encourages all process stakeholders to participate in process management. This allows them to contribute their own domain and method knowledge, thereby providing the necessary requirements for closing the model-reality divide and preventing lost innovation. However, social BPM presents its own challenges and problems that first need to be overcome. This paper has provided initial insight into the factors that influence the success of social BPM by analyzing its human-related challenges. Future work on the issues raised can be performed in different directions. First and foremost, measures that allow meeting them need to be found (or connected to social BPM), refined, and evaluated. Some such approaches have already been presented throughout Section 5, but still need to be critically reflected to determine where and how much additional work is required. Some suggestions on how to commence this research have already been given throughout the previous section. Secondly, the list of challenges, which has been derived from relevant literature, should be empirically validated by conducting expert interviews, or obtaining experiences through concrete applications of social BPM in real-world scenarios. Lastly, the scope of examination chosen for this paper can be broadened. Specifically, social BPM is a topic orthogonal to other aspects of BPM [vdA13], and may thus yield challenges specific to all of its core elements (cf. Section 3). Therefore, a future version of this paper will re-examine the challenges of social BPM from this higher point of view.

## References

- [BDJ<sup>+</sup>11] G. Bruno, F. Dengler, B. Jennings, R. Khalaf, S. Nurcan, M. Prilla, M. Sarini, R. Schmidt, and A. R. Silva. Key Challenges for Enabling Agile BPM with Social Software. *Journal of Software Maintenance and Evolution: Research and Practice*, 23(4):297–326, 2011.
- [BICS07] W. Bandara, M. Indulska, S. Chong, and S. Sadiq. Major Issues in Business Process Management: An Expert Perspective. In *Proceedings of the 15th European Conference on Information Systems (ECIS 2007)*, pages 1240–1251, St. Gallen, 2007.
- [DC11] M. Das and L. Chow. Next Generation BPM Suites: Social and Collaborative. In *Social BPM - Work, Planning and Collaboration Under the Impact of Social Technology*, pages 193–208. Future Strategies Inc., 1st edition, 2011.
- [DRH11] A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM*, 54(4):86, 2011.
- [Dry97] D. C. Dryer. Wizards, Guides, and Beyond: Rational and Empirical Methods for Selecting Optimal Intelligent User Interface Agents. In *Proceedings of the 2nd International Conference on Intelligent User Interfaces (IUI 1997)*, pages 265–268, Orlando, 1997.
- [EGH<sup>+</sup>10] S. Erol, M. Granitzer, S. Happ, S. Jantunen, B. Jennings, P. Johannesson, A. Koschmider, S. Nurcan, D. Rossi, and R. Schmidt. Combining BPM and Social Software: Contradiction or Chance? *Journal of Software Maintenance and Evolution: Research and Practice*, 22(6-7):449–476, 2010.

- [Ham10] M. Hammer. What is Business Process Management? In Jan vom Brocke and Michael Rosemann, editors, *Handbook on Business Process Management 1*, pages 3–16. Springer Berlin Heidelberg, 1st edition, 2010.
- [Hip06] H. Hippner. Bedeutung, Anwendungen und Einsatzpotenziale von Social Software. *HMD - Praxis der Wirtschaftsinformatik*, 43(252):6–16, 2006.
- [HT85] S. R. Hiltz and M. Turoff. Structuring Computer-Mediated Communication Systems to Avoid Information Overload. *Communications of the ACM*, 28(7):680–689, 1985.
- [IRRG09] M. Indulska, J. C. Recker, M. Rosemann, and P. Green. Business Process Modeling: Current Issues and Future Challenges. In *Proceedings of the 21st International Conference on Advanced Information Systems Engineering (CAiSE 2009)*, pages 501–514, Amsterdam, 2009.
- [JAW09] P. Johannesson, B. Andersson, and P. Wohed. Business Process Management with Social Software Systems â€” A New Paradigm for Work Organisation. In *Business Process Management Workshops*, pages 659–665. Springer Berlin Heidelberg, 2009.
- [JF09] B. Jennings and A. Finkelstein. Digital Identity and Reputation in the Context of a Bounded Social Ecosystem. In *Business Process Management Workshops*, pages 687–697. Springer Berlin Heidelberg, 2009.
- [JF11] B. Jennings and A. Finkelstein. Implicit Social Production: Utilising Socially Generated Data By-Products. In *Business Process Management Workshops*, pages 359–371. Springer Berlin Heidelberg, 2011.
- [KJ03] J. Krogstie and H. D. Jorgensen. Quality of Interactive Models. In *Advanced Conceptual Modeling Techniques*, pages 351–363. Springer Berlin Heidelberg, 2003.
- [KR09] M. Koch and A. Richter. *Enterprise 2.0 - Planung, Einföhrung und erfolgreicher Einsatz von Social Software in Unternehmen*. Oldenbourg Wissenschaftsverlag GmbH, 2nd edition, 2009.
- [KSR09] A. Koschmider, M. Song, and H. A. Reijers. Social Software for Modeling Business Processes. In *Business Process Management Workshops*, pages 666–677. Springer Berlin Heidelberg, 2009.
- [LB13] R. Laue and M. Becker. Evaluating Social Tagging for Business Process Models. In *Business Process Management Workshops*, pages 280–291. Springer Berlin Heidelberg, 2013.
- [MRvdA10] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst. Seven Process Modeling Guidelines (7PMG). *Information and Software Technology*, 52(2):127–136, 2010.
- [MS11] D. Martinho and A. R. Silva. ECHO - An Evolutive Vocabulary for Collaborative BPM Discussions. In *Business Process Management Workshops*, pages 408–419. Springer Berlin Heidelberg, 2011.
- [NH11] B. Niehaves and J. Henser. Boundary Spanning Practices in BPM: A Dynamic Capability Perspective. In *Proceedings of the 17th Americas Conference on Information Systems (AMCIS 2011)*, Detroit, 2011.
- [NP11] B. Niehaves and R. Plattfaut. Collaborative Business Process Management: Status Quo and Quo Vadis. *Business Process Management Journal*, 17(3):384–402, 2011.
- [NP12] A. Nolte and M. Prilla. Normal Users Cooperating on Process Models: Is It Possible at All? In *Proceedings of the 18th International Workshop on Groupware (CRIWG 2012)*, pages 57–72, Raesfeld, 2012.
- [NS09] S. Nurcan and R. Schmidt. Introduction to the First International Workshop on Business Process Management and Social Software (BPMS2 2008). In *Business Process Management Workshops*, pages 647–648. Springer Berlin Heidelberg, 2009.
- [PCBV10] S. Patig, V. Casanova-Brito, and B. Vögel. IT Requirements of Business Process Management in Practice - An Empirical Study. In *Proceedings of the 8th International Conference on Business Process Management (BPM 2010)*, pages 13–28, Hoboken, 2010.

- [Puc11] M. J. Pucher. How to Link BPM Governance and Social Collaboration Through an Adaptive Paradigm. In *Social BPM - Work, Planning and Collaboration Under the Impact of Social Technology*, pages 57–76. Future Strategies Inc., 1st edition, 2011.
- [QJSJ08] H. Qu, J. Sun, and H. T. Jamjoom. SCOOP: Automated Social Recommendation in Enterprise Process Management. In *Proceedings of the 2008 IEEE International Conference on Services Computing (SCC 2008)*, pages 101–108, Honolulu, July 2008.
- [Ric11] C. Richardson. Foreword. In *Social BPM - Work, Planning and Collaboration Under the Impact of Social Technology*, pages 5–8. Future Strategies Inc., 1st edition, 2011.
- [Rit10] P. Rittgen. Success Factors of e-Collaboration in Business Process Modeling. In *Proceedings of the 22nd International Conference on Advanced Information Systems Engineering (CAiSE 2010)*, pages 24–37, Hammamet, 2010.
- [RMH13] J. Recker, J. Mendling, and C. Hahn. How Collaborative Technology Supports Cognitive Processes in Collaborative Process Modeling: A Capabilities-Gains-Outcome Model. *Information Systems*, pages 1–15, 2013.
- [RSR12] J. C. Recker, N. Safrudin, and M. Rosemann. How Novices Design Business Processes. *Information Systems*, 37(6):557–573, September 2012.
- [Rus11] S. Russell. How Social Technologies Enhance the BPM Experience for All Participants. In *Social BPM - Work, Planning and Collaboration Under the Impact of Social Technology*, pages 113–122. Future Strategies Inc., 1st edition, 2011.
- [RvB10] M. Rosemann and J. vom Brocke. The Six Core Elements of Business Process Management. In Jan vom Brocke and Michael Rosemann, editors, *Handbook on Business Process Management 1*, pages 107–122. Springer Berlin Heidelberg, 1st edition, 2010.
- [RvWML10] H. A. Reijers, S. van Wijk, B. Mutschler, and M. Leurs. BPM in Practice: Who Is Doing What? In *Proceedings of the 8th International Conference on Business Process Management (BPM 2010)*, pages 45–60, Hoboken, 2010.
- [Sch12] R. Schmidt. A Framework for the Support of Value Co-creation by Social Software. In *Business Process Management Workshops*, pages 242–252. Springer Berlin Heidelberg, 2012.
- [SMM<sup>+</sup>10] A. R. Silva, R. Meziani, R. Magalhães, D. Martinho, A. Aguiar, and N. Flores. AGILIPO: Embedding Social Software Features into Business Process Tools. In *Business Process Management Workshops*, pages 219–230. Springer Berlin Heidelberg, 2010.
- [SN09] R. Schmidt and S. Nurcan. BPM and Social Software. In *Business Process Management Workshops*, pages 649–658. Springer Berlin Heidelberg, 2009.
- [SN10] R. Schmidt and S. Nurcan. Augmenting BPM with Social Software. In *Business Process Management Workshops*, pages 201–206. Springer Berlin Heidelberg, 2010.
- [SR12] A. R. Silva and M. Rosemann. Processpedia: an Ecological Environment for BPM Stakeholders’ Collaboration. *Business Process Management Journal*, 18(1):20–42, 2012.
- [Sur05] J. Surowiecki. *The Wisdom of Crowds*. Anchor, 1st edition, 2005.
- [SVOK11] F. Schönthaler, G. Vossen, A. Oberweis, and T. Karle. *Geschäftsprozesse für Business Communities - Modellierungssprachen, Methoden, Werkzeuge*. Oldenbourg Wissenschaftsverlag GmbH, München, 1st edition, 2011.
- [SWMW09] S. Smirnov, M. Weidlich, J. Mendling, and M. Weske. Action Patterns in Business Process Models. In *Service-Oriented Computing - Proceedings of the 7th International Joint Conference ICSOC-ServiceWave*, pages 115–129, Stockholm, 2009.
- [VB10] A. Voinov and F. Bousquet. Modelling with Stakeholders. *Environmental Modelling & Software*, 25(11):1268–1281, November 2010.
- [vdA13] W. M. P. van der Aalst. Business Process Management: A Comprehensive Survey. *ISRN Software Engineering*, 2013:1–37, 2013.
- [vdARS05] W. M. P. van der Aalst, Hajo A. Reijers, and Minseok Song. Discovering Social Networks from Event Logs. *Computer Supported Cooperative Work (CSCW)*, 14(6):549–593, October 2005.

- [VFL10] D. Vanderhaeghen, P. Fettke, and P. Loos. Organisations- und Technologieoptionen des Geschäftsprozessmanagements aus der Perspektive des Web 2.0. *Wirtschaftsinformatik*, 52(1):17–32, January 2010.
- [VH07] G. Vossen and S. Hagemann. *Unleashing Web 2.0: From Concepts to Creativity*. Academic Press, 1st edition, 2007.
- [Wes12] M. Weske. *Business Process Management - Concepts, Languages, Architectures*. Springer Berlin Heidelberg, 2nd edition, 2012.
- [WHR<sup>+</sup>10] H. F. Witschel, B. Hu, U. V. Riss, B. Thönssen, R. Brun, A. Martin, and K. Hinkelmann. A Collaborative Approach to Maturing Process-Related Knowledge. In *Business Process Management*, pages 343–358. Springer Berlin Heidelberg, 2010.
- [zMR08] M. zur Muehlen and J. C. Recker. How Much Language Is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In *Proceedings of the 20th International Conference on Advanced Information Systems Engineering (CAiSE 2008)*, pages 465–479, Montpellier, 2008.

# BPMN 2.0 Serialization - Standard Compliance Issues and Evaluation of Modeling Tools

Matthias Geiger, Guido Wirtz

Distributed Systems Group  
University of Bamberg  
An der Weberei 5

D-96047 Bamberg, Germany

{ matthias.geiger | guido.wirtz }@uni-bamberg.de

**Abstract:** *Business Process Model and Notation* (BPMN) 2.0 process models are used more and more, both in practice as in academia. Although academic research mainly focuses on sophisticated semantic checks and extensions there still exist problems in the basic usage of BPMN. This paper investigates issues in BPMN model serializations which arise as a result of the complexity and inconsistency of the standard document. We present a set of serialization constraints as a starting point for sophisticated compliance checks on serialized BPMN models. Furthermore, these constraints are used to perform an evaluation of current modeling tools. This evaluation reveals that the creation of standard compliant models is still a non-trivial endeavor.

## 1 Introduction and Motivation

The *Business Process Model and Notation* (BPMN) [OMG11a] is an international standard for process modeling developed and maintained by the *Object Management Group* (OMG). It is widely accepted in practice and in academia, as demonstrated by various BPMN process modeling tools, engines for executing BPMN processes and also in various academic papers, see for instance [CT09,DDO08,LVD09,WG08], BPMN is a topic of research. The first versions of BPMN [BPM04,OMG08,OMG09] focus on the standardization of a set of graphical shapes for process modeling - this was also denoted by the former meaning of the acronym: *Business Process Modeling Notation*. For that reason, the main scope of BPMN was conceptual modeling of business processes for documentation and visualization purposes.

With version 2.0 [OMG11a] the standard has been enhanced by an (informal) definition of execution semantics to provide executability on compliant BPMN engines. By adding executability for process models, the complexity of the standard (and also of the resulting models) naturally is increased: Especially, aspects such as data handling which have not been covered in the graphical models of former versions have to be defined now.

A problematic aspect of the former BPMN versions is that no serialization format has been standardized. This led to a plethora of different serialization formats which hampers model exchange between tools and engines: Modeling tools used their own proprietary file formats, but also mappings and transformations to other standards had been used. Especially, for processes which should be executable, a mapping to the *Web Services*

*Business Process Execution Language* (WS-BPEL) [OAS07] was intended as the serialization format for executable service-based process models and has been considered and evaluated in various academic papers [RM06,KMWL09,WDGW08,ODtHvdA08]. A mapping to WS-BPEL is still part of the BPMN standard document [OMG11a,Sec.14], but executable BPMN models based on the official WS-BPEL-mapping have some portability issues [LW13].

However, it is not necessary any longer to use WS-BPEL to define executable processes as the implementation of official serialization formats is the second important extension in [OMG11a]. There are two XML-based formats proposed in [OMG11a, Sec. 15]: *XML Metadata Interchange* (XMI) [OMG11b] and a serialization based on *XML Schema Definitions* (XSDs) [W3C04b,W3C04a]. For both formats the OMG provides normative schema definitions<sup>6</sup>. Especially the XSD-based serialization is referred to throughout the whole standard document and is already used in various modeling tools and engines.

Both standardized serialization formats have a serious issue: They do not ensure the correctness of serialized BPMN models. In [OMG11a] various requirements and constraints are defined which have to be respected in well-formed, standard compliant models. These constraints range from graphical rules (such as the appearance of shapes) to rules for execution semantics. But there do also exist hundreds of constraints which are relevant for correct model serialization. Using the standardized serialization formats most of the structural constraints can be verified by performing checks such as a schema validation. But a lot of other rules cannot be checked in this way. This implies that just using the standardized formats does not guarantee that the model is BPMN compliant. The problem is exacerbated as [OMG11a] provides no list of all constraints which well-formed models have to respect. Instead, rules and constraints are spread all over the standard document in tables, figures, XML schema excerpts and the running text. As we sketched in previous work [GW13], besides vendor policy this is a main issues which hinders model exchange between different tools. To be able to determine whether a BPMN model is *correct* by means of standard compliance, it is important to know which rules models have to adhere.

The paper at hand provides two main contributions: First, we present an extensive list of constraints relevant for correct serialization stated in [OMG11a]. These rules are independent from the concrete underlying serialization mechanism. Therefore, they can be used to validate the standard compliance of models in the standardized as well as in proprietary formats. Furthermore, we assess which constraints are already covered by the XSD based serialization format and highlight some issues and inconsistencies. Second, we evaluate the serialization mechanism of several modeling tools. The evaluation consists of checks whether the standardized serialization format is supported and of an evaluation of the generated models' standard compliance. The focus of our analysis is the assessment of whether the tools implement checking mechanisms for the constraints we revealed.

The remainder of the paper is organized as follows: In the next section, the approach of extracting the constraints from [OMG11a] and the resulting set of BPMN serialization constraints is presented. The subsequent section evaluates how model serialization is

---

<sup>6</sup> sec: <http://www.omg.org/spec/BPMN/2.0/>

implemented in state-of-the-art modeling tools. Related work is summarized and assessed in section 4 Finally, section 5 concludes the work and gives an outlook on future work.

## 2 BPMN Serialization Constraints

Before presenting our approach and results, we have to clarify the scope and limitations of our work: We strictly focus on the serialization of BPMN process models. We provide an overview of constraints stated in [OMG11a] regarding the serialization of BPMN process models. Although bound to serialization, we do not limit the standard screening to the normative XML-based serialization formats. Therefore, the set of rules is technology independent and can be used to analyze all types of BPMN serialization formats.

The following aspects are not covered in this work: Requirements regarding the visual appearance of BPMN shapes are out of scope. The serialization of BPMN diagrams using the *BPMN Diagram Interchange* format [OMG11a, Sec. 12] is left out as well. Moreover, we exclude all aspects regarding the execution of BPMN process models on compliant engines. First and foremost this involves all advanced execution semantics aspects but also all constraints affecting instance attributes and variables.

Due to the limitations of space, it is not possible to present all constraints here. A detailed description of our work and all extracted 611 rules can be found in a technical report [Gei13].

### 2.1 Constraint Categorization and Extraction Approach

The constraints stated in [Fehler! Verweisquelle konnte nicht gefunden werden.] can be divided and categorized in four major categories which are described hereafter.

**Basic Attribute/Sub Element Cardinality (CARD):** The most essential constraints define the general structure of BPMN models. The standard comprises a description of all possible elements, no matter whether they are depicted graphically or not, their attributes and their relations. In particular, for each attribute and model association, it is of interest whether the attribute is mandatory or whether some minimum or maximum occurrence constraints apply.

**Basic Value Restrictions and Default Values (VAL):** This addresses value restrictions, such as an enumeration of allowed values, and the definition of default values.

**Basic Reference Constraints (REF):** Another important aspect is the usage of references in [OMG11a]. BPMN allows to reuse and to refer to specific elements. The main source for references in concrete model instances is the definition of control and message flows. Each control flow link is realized as a *SequenceFlow* element which has to reference a source and a target element. The same applies to the definition of *MessageFlows*. In such cases an element can be referenced by other elements through a unique identifier and reference attributes. But reuse through referencing is also possible for a lot of other elements. An example is the definition of a *Message* which is referenced by a *SendTask* and a corresponding *ReceiveTask* using their *messageRef* attribute [OMG11a, p.159-162].

Two aspects are important for each reference: First, the reference must be resolvable that is, the referenced item must exist in the same model or must be imported from

another model or artifact. Second, in most cases only specific elements are allowed to be referenced. In the previous example, only *Messages* are valid elements. If the *messageRef* attribute references another element (e.g., an *Operation*), this would violate the model constraints in [OMG11a].

**Extended Constraints (EXT):** All other rules revealed in [OMG11a] are categorized as extended constraints. Typical rules for this category are constraints which apply only under certain preconditions. A frequently used precondition is that constraints only apply, if the process is defined as executable. Besides, restrictions which cannot be expressed by the basic cardinality, value and reference constraints, such as interdependence between model elements, are also listed as extended constraints.

The approach of determining rules basically consists of an in-depth analysis of the standard document. The identification and extraction of rules for the former three categories CARD, VAL and REF is rather straightforward: The standard document [OMG11a] provides tables listing all attributes and relations for (almost) each BPMN element. These tables provide (among other information) all aspects which are relevant to extract constraint definitions: attribute names, type definitions, value and cardinality restrictions.

More crucial is the definition of extended rules. Only some aspects may be derived from the previously mentioned tables. However, the majority of the extended constraints are spread over the standard's running text. In this cases not only the whole document is relevant, but also some interpretation is needed to identify requirements. Moreover, some other rules are not mentioned explicitly, nonetheless they are important. An example is that at least one *messageEventDefinition* must be present, if a *StartEvent* is target of a *MessageFlow* definition.

## 2.2 Results

In total we identified more than 600 different constraints. Figure 1 shows an overview of the distribution to the different categories. For a complete overview please refer to [Gei13]. However, the main findings and peculiarities are described in this section.

The standard [OMG11a] defines 311 different attributes and associations (CARD) for 108 BPMN elements. For each attribute the attribute name, the datatype and the required cardinality has been extracted. In combination with the defined inheritance structure these rules already allow the creation of structural correct BPMN models. It is problematic that the tables are not always well aligned with the UML class diagrams which are also included in [OMG11a]. Both, class diagrams and tables, contain information about the minimum and maximum occurrence of the relations between BPMN model elements and in some cases the information is inconsistent. For example, the *operationRef* attribute of a *SendTask* is marked as optional in the class diagram ([0..1]) but the corresponding table states that the attribute is mandatory [OMG11a] p.160-161].

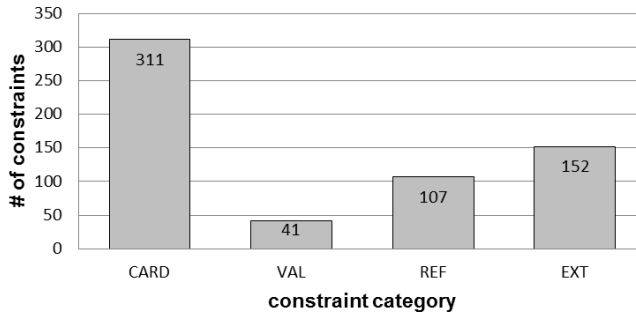


Figure 1: Number of extracted constraints by category

The 41 extracted value restrictions (VAL) can be divided in two groups: The first group covers default values for basic datatypes such as *boolean* and *string*. The second group comprises enumerations of allowed values. An example for this is the definition of allowed values for the attribute *gatewayDirection* for BPMN Gateways. This attribute indicates whether the gateway splits or merges the control flow and only the values *Unspecified*, *Converging*, *Diverging* and *Mixed* are valid [OMG11a, p.91].

As mentioned before, flow definitions (*Sequence* and *Message Flow*) are one important example for reference usage (REF). But the standard comprises 107 different associations between elements which are implemented as references. In order to determine the reference correctness for each reference definition the allowed datatypes have been determined.

Finally, 152 extended constraints (EXT) have been identified in the standard document. The complexity of the extended constraints varies. Examples for simple rules are constraints which only apply if the process is defined as “*executable*” that is, the attribute *isExecutable* of a process definition is set to `true`. But also more complex rules affecting various elements have been identified.

Generally, it is hard to prove that our list of constraints is complete and free of errors. Especially in cases where interpretation is needed or rules are stated only implicit, it is essential to ensure the quality and correctness of our extracted rules. However, we ensured the quality by internal discussions in our group and comparing our results with a less extensive set of BPMN constraints provided by Silver [Sil11]. And also the interpretations and according solutions implemented in different modeling tools have been considered. Moreover, we published all constraints and a technical report [Gei13] on our web page<sup>7</sup> and are looking forward for feedback from the community for an ongoing process of improvement for the constraint set.

### 2.3 XSD-based Standard Serialization

As mentioned in section 1 the standard is closely linked to the normative XSDs and also more and more modeling tools support the import or export in the XSD-based serialization format. Therefore, it is of particular interest which standard constraints are

<sup>7</sup> <http://www.uni-bamberg.de/pi/bpmn-constraints>

already covered when using this format and whether a schema validation is able to reveal violations.

The structural aspects defined in the cardinality constraints are directly mappable to XML schema definitions. So, it is no surprise that most of these constraints are correctly implemented in the normative XSDs. Exceptions are some aspects which are not directly viable using XSDs, such as multiple inheritance, or deviations which depend on modeling decisions. An example is the implementation of mandatory attributes with default values. Using XML schema definitions, it is sufficient to define a default value and leave out the definition that an attribute is mandatory, as the default value always applies when no other value is defined (and therefore a value is always available). Another modeling decision is the omission of bi-directional references on XML level. An example is the attribute *boundaryEventRefs* for *Activities*. This attribute is omitted as each *BoundaryEventDefinition* references the activity to which it is attached by the attribute *attachedToRef*. Thus, it is still clearly defined which *BoundaryEventDefinition* is attached to which *Activity*. There also exist several clear violations of CARD constraints: The most frequent violation is that attributes are defined as mandatory in the standard document but the XSD marks them as optional. This is the case for 24 constraints. But, four other cardinality requirements are enforced incorrectly in the normative XSD.

Default values can be easily defined in schema definitions as well as value enumerations based on basic datatypes such as strings. But, four out of the 41 VAL constraints are implemented faultily, as required default values are not implemented. Two more inconsistencies exist affecting the attributes of the elements *Transaction* and *BoundaryEvent*. There, default values are defined in the XSD which are not stated in [OMG11a].

To implement references, BPMN proposes two different mechanisms [OMG11a, p.476-477]: If an element must be defined in the same file, the reference is realized using an *xs:IDREF* [W3C04b]. If references potentially cross file borders the reference attribute's datatype is set to *xs:QName* [W3C04b]. A schema validation can be used to determine whether an *xs:IDREF* is resolvable [W3C04a]. Missing element definitions for *xs:QName* references cannot be detected. Moreover, in both cases the reference type violations are undetectable using a basic schema validation and therefore there is no support for any of the 107 reference constraints (REF).

As with the extraction, the XSD-based implementation of extended constraints (EXT) is hard. The observance of most rules can not be enforced by schema definitions. Out of the 152 extended constraints only seven are directly implemented in XSDs. Examples are constraints regarding the mutual exclusion of attributes (implemented as sub-elements) using the *xs:choice* [W3C04a] operator.

To conclude, the XSD-based serialization format introduced in [OMG11a] is a good starting point to define standard compliant BPMN process models. Especially the basic model structure and value restrictions are well covered: About 91% of the CARD and 90% of the VAL constraints are enforced by the XSDs. However, due to the missing reference checking support and only nominal (5%) support for the extended rules, the total constraint implementation is limited to about 54% of all 611 revealed constraints. Figure 2 shows the rule coverage in a graphical form.

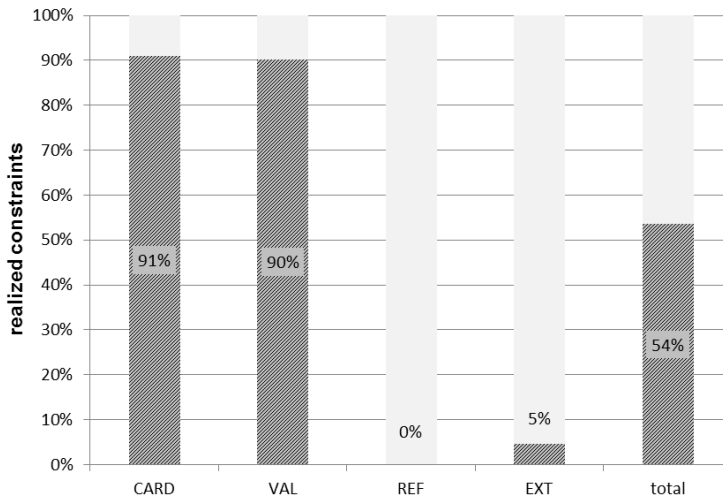


Figure 2: XSD-based serialization: constraint implementation coverage (percentage chart)

Regarding those numbers, it is evident that the usage of XML schema validation alone is not sufficient to enforce the correctness of models. Tool developers and end users who use such models need to perform more sophisticated checks to ensure that the model at hand is in fact standard compliant.

### 3 Evaluation of Existing Modeling Tools

Today, most BPMN modeling tools provide structural and semantic checks to some degree which enhance the abilities of XML schema validations. We compared and analyzed six BPMN editors regarding their ability to produce BPMN compliant models and their support to enforce the rules presented in section 2.

#### 3.1 Analyzed tools

The analysis of BPMN editors comprises six different process modeling tools namely: BizAgi Process modeler<sup>8</sup>, camunda Fox modeler<sup>9</sup>, eclipse BPMN2 modeler<sup>10</sup>, itp-commerce Process Modeler 6 for Visio<sup>11</sup>, Signavio Process Modeler<sup>12</sup> and Yaoqiang BPMN Editor<sup>13</sup>. All tools are dedicated for BPMN process modeling. Several tools are freely available as freeware or open source software, but we also evaluate two commercial editors (Signavio and itp-commerce).

Table 1 shows an overview of the tested modeling tools. Besides the vendor and the editor name, the actual version under test is stated. Furthermore, it is indicated whether

<sup>8</sup> <http://www.bizagi.com/download/>

<sup>9</sup> <http://www.camunda.org/design/modeler.html>

<sup>10</sup> <http://www.eclipse.org/bpmn2-modeler/>

<sup>11</sup> <http://www.itp-commerce.com/>

<sup>12</sup> <http://www.signavio.com>

<sup>13</sup> <http://sourceforge.net/projects/bpmn/>

the tool is commercial (*comm.*), freely available (*freeware*) or published as open source (*OS*). The base platform is depicted in the subsequent row. Some editors are implemented as a standalone tool, three editors are implemented as plugins for eclipse or Microsoft Visio and one editor is cloud-based and usable through a standard web browser. Moreover, for each modeling tool the default serialization format and the ability to import or export in the standardized XSD-based format is expressed.

Table 1: Overview of evaluated BPMN process modeling tools

| Vendor/<br>Developer                                      | BizAgi             | camunda        | eclipse          | itp-<br>commerce                  | Shi<br>Yaoqiang            | Signavio           |
|-----------------------------------------------------------|--------------------|----------------|------------------|-----------------------------------|----------------------------|--------------------|
| Tool name                                                 | Process<br>modeler | Fox<br>modeler | BPMN2<br>modeler | Process<br>Modeler 6<br>for Visio | Yaoqiang<br>BPMN<br>editor | Process<br>Modeler |
| Version                                                   | 2.4.0.8            | 2.0.12         | 0.2.6            | 6.3438                            | 2.1.36                     | 7.0.0              |
| License                                                   | freeware           | OS             | OS               | comm.                             | OS                         | comm.              |
| Platform                                                  | standalone         | eclipse        | eclipse          | MS Visio                          | standalone                 | cloud-<br>based    |
| Default<br>Serialization<br>Format                        | <i>prop.</i>       | BPMN<br>XSD    | BPMN<br>XSD      | MS Visio                          | BPMN<br>XSD                | <i>unknown</i>     |
| Import/Exp<br>ort support<br>(XSD-based<br>serialization) | —/—                | +/+            | +/+              | (+)/+                             | +/+                        | +/+                |

### 3.2 Evaluation Method

To assess the six different modeling tools we perform a three-step approach: First, we evaluate the supported BPMN conformance level and validation features of the tools under test. Depending on the determined supported conformance class, we check the modeling abilities and the serialization correctness in an second step. Finally, some advanced features and extended constraints are checked. Each step is described in more detail in the following paragraphs.

**Supported conformance level and validation features:** [OMG11a, Sec.2] introduces several *conformance levels* which define different levels of BPMN support. Especially the conformance level *Process Modeling Conformance* is relevant for the evaluation at hand. This level is refined to four sub-classes: *Descriptive*, *Analytic*, *Common Executable* and *Full Process Modeling Conformance* [OMG11a, p.2]. Depending on the level a modeling tool must support a specific set of BPMN elements and attributes [OMG11a, p.2-7]. In order to assess the modeling abilities of the tested modelers it is important to know which conformance class is supported by each tool. As the tools do not indicate which level they support, the level has to be determined by analyzing the palette of supported elements and their attributes. Furthermore, for each editor it is checked whether validation features are implemented. If validation techniques are used, we check whether the validation has to be performed manually, whether it is started automatically when the model is saved or exported, or whether the editor supports real-time validation during modeling.

**Modeling abilities and serialization correctness:** Depending on the supported conformance classes the editor specific modeling abilities are assessed in the next step. To achieve this, for each conformance level example processes are modeled with each editor. The resulting models are checked, whether a) it was possible to recreate the model, b) the model passes an XML schema validation and c) the model violates some other constraints.

The example models for the descriptive and analytic sub-classes are gathered from the *BPMN 2.0 by example* document provided by the OMG [OMG10]: We evaluate the ability of the editors to recreate the “Small Examples introducing Core Concepts” [OMG10, Sec.5]. In order to evaluate the modeling abilities for the common executable and full conformance sub-classes we use process models created in a student project containing advanced features regarding Web Services usage and data handling.

**Evaluation of advanced features and extended constraint validation:** We select several well known correctness aspects from our list of extended constraints<sup>14</sup> and check whether they are correctly implemented in the modeling tools.

- In [OMG11a, p.54] it is stated that “[i]mporting *Xml Schema 1.0*, *WSDL 2.0* and *BPMN 2.0* types *MUST* be supported”. We assess whether the tools support this requirement.
- BPMN allows the reuse of elements and processes defined in other files. So, it is evaluated whether tools are able to reference elements across file-borders.
- EXT.028/031: *SequenceFlows* must not cross the borders of pools; *MessageFlows* must cross pool borders.
- EXT.096/104: For *StartEvents* no incoming *SequenceFlow*, for *EndEvents* no outgoing *SequenceFlow* is allowed.
- EXT.100: For sub-processes only “none” *StartEvents* are allowed.
- EXT.016-019: The attribute *gatewayDirection* of *Gateways* defines whether the gateway is splitting (attribute value “diverging”) or joining (“converging”) the control flow. If a gateway has multiple incoming and outgoing sequence flows, the attribute must have the value “mixed” or “unspecified” (attribute mandatory for common executable and full conformance only)
- EXT.057: An *Event Sub-Process* must not have any incoming or outgoing *SequenceFlows*.

### 3.3 Evaluation Results

In the following, the main findings are described and Table 2 gives an overview of our evaluation results. In the table “+” denotes that the feature/constraint under test is fully supported, “-” is used if a feature is not supported and partial support is indicated by “(+)”.

---

<sup>14</sup> See [Gei13] for a detailed description of the constraints.

Table 2: Overview: Evaluation results

| Vendor/<br>Developer                                                      | BizAgi                   | camunda                   | eclipse            | itp-<br>commerce                  | Shi Yaoqiang            | Signavio                 |
|---------------------------------------------------------------------------|--------------------------|---------------------------|--------------------|-----------------------------------|-------------------------|--------------------------|
| Tool name                                                                 | Process<br>modeler       | Fox modeler               | BPMN2<br>modeler   | Process<br>Modeler 6<br>for Visio | Yaoqiang<br>BPMN editor | Process<br>Modeler       |
| <b>Supported conformance level and validation features</b>                |                          |                           |                    |                                   |                         |                          |
| Supported<br>conformance<br>class                                         | ( <i>analytic</i> )      | ( <i>analytic</i> )       | ( <i>full</i> )    | analytic                          | ( <i>full</i> )         | analytic                 |
| Model<br>validation                                                       | manual                   | while<br>opening<br>model | opening/<br>saving | real-time/<br>manual              | real-time               | manual/<br>saving        |
| <b>Modeling abilities and serialization correctness</b>                   |                          |                           |                    |                                   |                         |                          |
| <b>Descriptive</b><br>creation<br>schema valid                            | +<br><i>n/a</i>          | +<br>+                    | (+)<br>+           | +<br>+                            | (+)<br>+                | +<br>+                   |
| <b>Analytic</b><br>creation<br>schema valid                               | (+)<br><i>n/a</i>        | +<br>+                    | +<br>+             | +<br>+                            | +<br>+                  | +<br>+                   |
| <b>C. Executable</b><br>creation<br>schema valid                          | <i>n/a</i><br><i>n/a</i> | <i>n/a</i><br><i>n/a</i>  | +<br>+             | <i>n/a</i><br><i>n/a</i>          | +<br>+                  | <i>n/a</i><br><i>n/a</i> |
| <b>Full</b><br>creation<br>schema valid                                   | <i>n/a</i><br><i>n/a</i> | <i>n/a</i><br><i>n/a</i>  | (+)<br>+           | <i>n/a</i><br><i>n/a</i>          | +<br>+                  | <i>n/a</i><br><i>n/a</i> |
| <b>Evaluation of advanced features and extended constraint validation</b> |                          |                           |                    |                                   |                         |                          |
| WSDL/XSD/<br>BPMN import                                                  | –                        | –                         | +                  | (+), <i>only</i><br><i>BPMN</i>   | +                       | –                        |
| Cross-file<br>references                                                  | –                        | –                         | +                  | +                                 | +                       | –                        |
| EXT.028/031                                                               | +                        | +                         | +                  | +                                 | +                       | +                        |
| EXT.096/104                                                               | +                        | –                         | +                  | +                                 | +                       | +                        |
| EXT.100                                                                   | –                        | –                         | –                  | +                                 | +                       | –                        |
| EXT.016-019                                                               | <i>n/a</i>               | <i>n/a</i>                | +                  | +                                 | (+)                     | (+)                      |
| EXT.057                                                                   | +                        | –                         | <i>n/a</i>         | +                                 | +                       | +                        |

Regarding the conformance classes all tools support at least the analytic sub-class. However, some elements and minor aspects are missing for most tools. So technically the editors do not fully support the conformance classes. For example, the BizAgi editor is not able to model *CallActivities* which are mandatory for the descriptive conformance class. Generally the BizAgi Process modeler is a particular case. It is not able to serialize the model in the standardized XML format. Thus, it is hard to determine the supported conformance class and it is not possible to perform schema validation of the serialized models.

Some kind of model validation is implemented in each modeling tool, but only the itp-commerce and the Yaoqiang editor support real-time checks during modeling. In all other tools the validation has to be started manually or the validation is only performed when saving or opening a file.

It is mostly possible to recreate the example models from [OMG10] in the different tools, but some minor issues exist: Eclipse and Yaoqiang are not able to bind a *TextAssociation* to *SequenceFlows* which is allowed by [OMG11a], resulting in only partial support ((+))

for the descriptive sub-class. The BizAgi editor fails to provide support for modeling *Message* elements as required by the analytic conformance sub-class. Modeling processes for the common executable and full conformance class was only performed for eclipse and Yaoqiang and both editors provide good support for this conformance classes, but eclipse is not able to model event sub-processes.

Import and usage of WSDL, XSD and BPMN files in a BPMN model is supported by eclipse and the Yaoqiang editor. Itp-commerce provides support for BPMN imports only. Thus, only these three editors can handle cross-file references for sub-processes and other BPMN elements. It is to note, that Signavio supports referencing in the cloud-based application, but when the model is serialized the referenced elements are either left out or included in the model, instead of referencing the called element.

The constraints EXT.028 and EXT.031 stating sequence and message flow connection rules are well implemented in all of the six tools. Either, the editors do not allow invalid connections or invalid connections are marked during a validation. The *SequenceFlow* restrictions for start and end events are also covered in all tools, except for the camunda modeler which does not check the adherence to this rules. Rule EXT.100 regarding allowed *StartEvent* type for sub-process is checked by itp-commerce and Yaoqiang. The *gatewayDirection* constraints (EXT.016-019) cannot be assessed for BizAgi and camunda as the attribute is not existent in the models and their serialization. Yaoqiang and Signavio cannot handle “mixed” gateways, but eclipse and itp-commerce implement all rules correctly. EXT.057 is implemented correctly in most tools. Camunda does not implement the constraint check and eclipse is not able to model event sub-processes.

All in all, the itp-commerce editor has implemented most of the rules checked in our evaluation. A drawback is that the editor is limited regarding modeling executable processes and does not support the usage of WSDL and XSD imports.

## 4 Related Work

Previous research on the correctness of BPMN mainly focuses on establishing formal semantics for BPMN and providing support for semantic validation, verification and correctness checks for BPMN models (see for instance [DDO08, LVD09, WG08, GD12] without claiming to be complete). We also evaluate the “correctness” of BPMN models, but our work targets another level of abstraction: Correctness in this work refers to standard compliance, especially regarding model serialization.

Nowadays, various extensions for specific domains, such as cloud application management [KBBL12], and aspects such as security [BHLR12] and social BPM [BFV12] are developed. In contrast to this, our work is limited to the standard [OMG11a], domain-independent and (regarding the set of constraints) technology-independent.

An approach related closely to BPMN serialization is [CT09]. The authors develop a simplified meta model and an XML-based serialization format for BPMN 1.1 [OMG08]. For this serialization format they provide a mechanism to check references and also complex constraints can be checked using XPath expressions. Unfortunately, the approach is not applicable for BPMN 2.0 process models, as the proposed serialization format is not standard compatible any more.

Besides academic work, a set of BPMN constraints, which comprises 39 rules, is provided by Silver in [Sil11]. Our constraint set is far more extensive for two reasons: First, we do not exclude all cardinality, value and reference constraints which are already covered by the normative XSDs. Second, we do not limit our approach to the “*Analytic Process Modeling Conformance subclass*” as Silver does. And recently (since January 2013) also the OMG recognized model serialization and interchange between tools as an important issue. They initiated the *BPMN Model Interchange Working Group* (BPMN MIWG)<sup>15</sup> with the goal to provide support for modeling tool developers and to identify issues in the standard which inhibit interchange of models between tools. Planned outputs are test cases, feature tests, a set of BPMN 2.0 issues and interchange guidelines<sup>16</sup>. The group is still working in an initial project phase and no official outcomes have been released yet.

## 5 Conclusion and Outlook

In our work we have presented an extensive and technology-independent set of BPMN serialization constraints. This list is a basis for tool developers as well as for end users who want to check the standard compliance of serialized BPMN models. All in all, we have identified 611 different rules in [OMG11a] and have categorized them in four different categories. An analysis of the standardized XSD-based serialization format of BPMN shows that only 54% of all rules are enforceable by schema validation. Especially, extended constraints cannot be checked by such a validation. Therefore, the correctness of models and their serialization depends on the ability of the modeling tools to enforce the observance of all constraints.

Our evaluation shows that state-of-the-art modeling editors already implement the revealed constraints partly. The evaluated tools are able to generate and serialize standard compliant models with a basic feature set. However, the internal validation mechanisms of the tool versions tested (see Table 1) still do not cover all revealed constraints which have to be respected in correct BPMN models. Moreover, the evaluation also shows that the serialization of models is an important first step, but also the diagram serialization is still an issue. Thus, we plan to broaden the scope of our rule set to cover BPMN Diagram Interchange (BPMN DI) constraints as well.

Due to the extensive set of rules, a manual review of the standard compliance of concrete models is not feasible. In order to provide tool independent conformance checks of BPMN models we plan to implement a validation suite which checks all serialization rules automatically. We have implemented automatic checks of all reference rules and are now extending the validation tool to check extended constraints.

## References

[BFV12] Marco Brambilla, Piero Fraternali, and Carmen Vaca. BPMN and Design Patterns for Engineering Social BPM Solutions. In Business Process Management Workshops,

---

<sup>15</sup> <http://www.omgwiki.org/bpmn-miwg/doku.php>

<sup>16</sup> see: <http://www.omgwiki.org/bpmn-miwg/doku.php#outputs>

- volume 99 of Lecture Notes in Business Information Processing, pages 219–230. Springer Berlin Heidelberg, 2012.
- [BHLR12] Achim D. Brucker, Isabelle Hang, Gero Lückemeyer, and Raj Ruparel. SecureBPMN: modeling and enforcing access control requirements in business processes. In Proceedings of the 17th ACM symposium on Access Control Models and Technologies, SACMAT '12, pages 123–126. ACM, 2012.
  - [BPM04] BPMI (Business Process Management Initiative). Business Process Modeling Notation (BPMN) Version 1.0, May 2004.
  - [CT09] Michele Chinosi and Alberto Trombetta. Modeling and validating BPMN diagrams. In Commerce and Enterprise Computing, 2009. CEC'09. IEEE Conference on, pages 353–360. IEEE, 2009.
  - [DDO08] Remco M. Dijkman, Marlon Dumas, and Chun Ouyang. Semantics and analysis of business process models in BPMN. *Information and Software Technology*, 50(12):1281 – 1294, 2008.
  - [GD12] Pieter Van Gorp and Remco M. Dijkman. A visual token-based formalization of BPMN 2.0 based on in-place transformations. *Information and Software Technology*, 2012.
  - [Gei13] Matthias Geiger. BPMN 2.0 Process Model Serialization Constraints. *Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik*, no. 92, Otto-Friedrich-Universität Bamberg, May 2013.
  - [GW13] Matthias Geiger and Guido Wirtz. Detecting Interoperability and Correctness Issues in BPMN 2.0 Process Models. In Proceedings of the 5th Central-European Workshop on Services and their Composition (ZEUS), Rostock, Germany, CEUR Workshop Proceedings, pages 39–42. CEUR-WS.org, Feb 2013.
  - [KBBL12] Oliver Kopp, Tobias Binz, Uwe Breitenbücher, and Frank Leymann. BPMN4TOSCA: A Domain-Specific Language to Model Management Plans for Composite Applications. In *Business Process Model and Notation*, volume 125 of Lecture Notes in Business Information Processing, pages 38–52. Springer Berlin Heidelberg, 2012.
  - [KMWL09] Oliver Kopp, Daniel Martin, Daniel Wutke, and Frank Leymann. The Difference Between Graph-Based and Block-Structured Business Process Modelling Languages. *Enterprise Modelling and Information Systems Architectures*, 4(1):3–13, 2009.
  - [LVD09] Niels Lohmann, Eric Verbeek, and Remco Dijkman. PetriNet Transformations for BusinessProcesses ASurvey. In *Transactions on Petri Nets and Other Models of Concurrency II*, volume 5460 of Lecture Notes in Computer Science, pages 46–63. Springer Berlin Heidelberg, 2009.
  - [LW13] Jörg Lenhard and Guido Wirtz. Detecting Portability Issues in Model-Driven BPEL Mappings. In Proceedings of the 25th International Conference on Software Engineering and Knowledge Engineering (SEKE'2013), Boston, USA, June 2013.
  - [OAS07] OASIS. Web Services Business Process Execution Language, April 2007. v2.0.
  - [ODtHvdA08] Chun Ouyang, Marlon Dumas, Arthur H.M. ter Hofstede, and Wil M.P. van der Aalst. Pattern-based translation of BPMN process models to BPEL web services. *International Journal of Web Services Research (IJWSR)*, 5(1):42–62, 2008.
  - [OMG08] OMG (Object Management Group). Business Process Modeling Notation (BPMN) Version 1.1, January 2008.
  - [OMG09] OMG (Object Management Group). Business Process Modeling Notation (BPMN) Version 1.2, January 2009.
  - [OMG10] OMG (Object Management Group). BPMN 2.0 by Example, June 2010.
  - [OMG11a] OMG (Object Management Group). Business Process Model and Notation (BPMN) Version 2.0, January 2011.
  - [OMG11b] OMG (Object Management Group). OMG MOF 2 XMI Mapping Specification Version 2.4.1, August 2011.
  - [RM06] Jan C. Recker and Jan Mendling. On the translation between BPMN and BPEL: Conceptual mismatch between process modeling languages. In 18th International

- Conference on Advanced Information Systems Engineering. Proceedings of Workshops and Doctoral Consortium, pages 521–532. Namur University Press, 2006.
- [Sil11] Bruce Silver. BPMN method and style. Cody-Cassidy Press, Aptos, CA, USA, 2<sup>nd</sup> edition, 2011.
- [W3C04a] W3C. XML Schema Part 1: Structures Second Edition, October 2004.
- [W3C04b] W3C. XML Schema Part 2: Datatypes Second Edition, October 2004.
- [WDGW08] Matthias Weidlich, Gero Decker, Alexander Grosskopf, and Mathias Weske. BPEL to BPMN: TheMyth of a Straight-ForwardMapping. In On the Move to Meaningful Internet Systems: OTM 2008, volume 5331 of Lecture Notes in Computer Science, pages 265–282. Springer Berlin Heidelberg, 2008.
- [WG08] PeterY.H. Wong and Jeremy Gibbons. A Process Semantics for BPMN. In Formal Methods and Software Engineering, volume 5256 of Lecture Notes in Computer Science, pages 355–374. Springer Berlin Heidelberg, 2008.

## *GI-Edition Lecture Notes in Informatics*

- P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.
- P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.
- P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Language to Information Systems, NLDB'2001.
- P-4 H. Wörn, J. Mühling, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensor-gestützte Chirurgie; Workshop des SFB 414.
- P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.
- P-6 Hans-Jürgen Appelpath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.
- P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.
- P-8 Reinhard Keil-Slawik, Johannes Magenheimer (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.
- P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeits-tagung.
- P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.
- P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.
- P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.
- P-13 Jan von Knop, Peter Schirmbacher und Viljan Mahni\_ (Hrsg.): The Changing Universities – The Role of Technology.
- P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.
- P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.
- P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.
- P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletz-barkeit meistern, 16. DFN Arbeitstagung.
- P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informations-systeme – MobIS 2002.
- P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Infor-matik 2002 – 32. Jahrestagung der Gesell-schaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.
- P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Infor-matik 2002 – 32. Jahrestagung der Gesell-schaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungs-band).
- P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Metho-den und Werkzeuge für die Entwicklung von Informationssystemen.
- P-22 Sigrid Schubert, Johannes Magenheimer, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.
- P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit
- P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middle-ware für XML-Anwendungen
- P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003
- P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Daten-banksysteme für Business, Technologie und Web
- P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin
- P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Er-fahrungen und Visionen
- P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems
- P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications
- P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenber (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistence, Scalability, Transactions – Database Mechanisms for Mobile Applications
- P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning, E-Services
- P-45 Bernhard Rumpe, Wolfgang Hesse (Hrsg.): Modellierung 2004
- P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment
- P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society
- P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications
- P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven
- P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik
- P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004
- P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration
- P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science and Grid Ad-hoc-Netze Medienintegration
- P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government
- P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen
- P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality
- P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Rannenber, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments
- P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische Bildung
- P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen
- P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit
- P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and its Applications

- P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005
- P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolfrid Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web
- P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik
- P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)
- P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)
- P-69 Robert Hirschfeld, Ryszard Kowalczyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005
- P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)
- P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005
- P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment
- P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"
- P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology
- P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture
- P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz
- P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006
- P-78 K.-O. Wenkel, P. Wagner, M. Morgens-tern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel
- P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006
- P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce
- P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS'06
- P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006
- P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics
- P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications
- P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems
- P-86 Robert Krimmer (Ed.): Electronic Voting 2006
- P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik
- P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006
- P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur
- P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006
- P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006
- P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1
- P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2
- P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen
- P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies
- P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

- P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices
- P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik
- P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik
- P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten
- P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments
- P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)
- P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme
- P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007
- P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops
- P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications
- P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures
- P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1
- P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2
- P-111 Christian Eibl, Johannes Magenheimer, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik
- P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis
- P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1<sup>st</sup> Conference on Social Semantic Web (CSSW)
- P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3<sup>rd</sup> International Conference on IT-Incident Management & IT-Forensics
- P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007
- P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1<sup>st</sup> International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and management The 4<sup>th</sup> International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Hermann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

- P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)  
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit  
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)
- P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)  
9<sup>th</sup> Workshop on Parallel Systems and Algorithms (PASA)  
Workshop of the GI/ITG Special Interest Groups PARS and PARVA
- P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)  
Unternehmens-IT:  
Führungsinstrument oder Verwaltungsbürde  
Referate der 28. GIL Jahrestagung
- P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)  
10<sup>th</sup> Workshop Software Reengineering (WSR 2008)
- P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)  
Modellierung 2008
- P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)  
Sicherheit 2008  
Sicherheit, Schutz und Zuverlässigkeit  
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)  
2.-4. April 2008  
Saarbrücken, Germany
- P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)  
Sigsand-Europe 2008  
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems
- P-130 Paul Müller, Bernhard Neumair, Gabi Dreö Rodosek (Hrsg.)  
1. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-131 Robert Krimmer, Rüdiger Grimm (Eds.)  
3<sup>rd</sup> International Conference on Electronic Voting 2008  
Co-organized by Council of Europe, Gesellschaft für Informatik und E-Voting, CC
- P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)  
DeLFI 2008:  
Die 6. e-Learning Fachtagung Informatik
- P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)  
INFORMATIK 2008  
Beherrschbare Systeme – dank Informatik Band 1
- P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)  
INFORMATIK 2008  
Beherrschbare Systeme – dank Informatik Band 2
- P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)  
Didaktik der Informatik –  
Aktuelle Forschungsergebnisse
- P-136 Andreas Beyer, Michael Schroeder (Eds.)  
German Conference on Bioinformatics GCB 2008
- P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)  
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)  
Synergien durch Integration und Informationslogistik  
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)  
Industrialisierung des Software-Managements  
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)  
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)  
Modellierung betrieblicher Informationssysteme (MobIS 2008)  
Modellierung zwischen SOA und Compliance Management
- P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)  
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung
- P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)  
Software Engineering 2009  
Fachtagung des GI-Fachbereichs Softwaretechnik

- P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)  
Datenbanksysteme in Business, Technologie und Web (BTW)
- P-145 Knut Hinkelmann, Holger Wache (Eds.)  
WM2009: 5th Conference on Professional Knowledge Management
- P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)  
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung  
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)
- P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)  
Business Process, Services Computing and Intelligent Service Management  
BPSC 2009 · ISM 2009 · YRW-MBP 2009
- P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)  
9<sup>th</sup> International Conference on Innovative Internet Community Systems  
I<sup>2</sup>CS 2009
- P-149 Paul Müller, Bernhard Neumair, Gabi Dreö Rodosek (Hrsg.)  
2. DFN-Forum  
Kommunikationstechnologien  
Beiträge der Fachtagung
- P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)  
Software Engineering  
2009 - Workshopband
- P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)  
PRIMIUM  
Process Innovation for Enterprise Software
- P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)  
Enterprise Modelling and Information Systems Architectures  
Proceedings of the 3<sup>rd</sup> Int'l Workshop EMISA 2009
- P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)  
Lernen im Digitalen Zeitalter  
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik
- P-154 Stefan Fischer, Erik Maehle, Rüdiger Reischuk (Hrsg.)  
INFORMATIK 2009  
Im Focus das Leben
- P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)  
BIOSIG 2009:  
Biometrics and Electronic Signatures  
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures
- P-156 Bernhard Koerber (Hrsg.)  
Zukunft braucht Herkunft  
25 Jahre »INFOS – Informatik und Schule«
- P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)  
German Conference on Bioinformatics 2009
- P-158 W. Claudepein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)  
Precision Agriculture  
Reloaded – Informationsgestützte Landwirtschaft
- P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)  
Software Engineering 2010
- P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)  
Software Engineering 2010 – Workshopband  
(inkl. Doktorandensymposium)
- P-161 Gregor Engels, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)  
Modellierung 2010
- P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)  
Vernetzte IT für einen effektiven Staat  
Gemeinsame Fachtagung  
Verwaltungsinformatik (FTVI) und  
Fachtagung Rechtsinformatik (FTRI) 2010
- P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)  
Mobile und Ubiquitäre Informationssysteme  
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler Kollaboration
- P-164 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2010: Biometrics and Electronic Signatures  
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

- P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)  
10<sup>th</sup> International Conference on Innovative Internet Community Systems (I<sup>2</sup>CS) – Jubilee Edition 2010 –
- P-166 Paul Müller, Bernhard Neumair, Gabi Dreö Rodosek (Hrsg.)  
3. DFN-Forum Kommunikationstechnologien  
Beiträge der Fachtagung
- P-167 Robert Krimmer, Rüdiger Grimm (Eds.)  
4<sup>th</sup> International Conference on Electronic Voting 2010  
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)  
Didaktik der Informatik  
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik
- P-169 Michael Kerres, Nadine Ojstersek, Ulrik Schroeder, Ulrich Hoppe (Hrsg.)  
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.
- P-170 Felix C. Freiling (Hrsg.)  
Sicherheit 2010  
Sicherheit, Schutz und Zuverlässigkeit
- P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)  
Modellierung betrieblicher Informationssysteme (MobIS 2010)  
Modellgestütztes Management
- P-172 Stefan Klink, Agnes Koschmider, Marco Mevius, Andreas Oberweis (Hrsg.)  
EMISA 2010  
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme  
Beiträge des Workshops der GI-Fachgruppe EMISA  
(Entwicklungsmethoden für Informationssysteme und deren Anwendung)
- P-173 Dietmar Schomburg, Andreas Grote (Eds.)  
German Conference on Bioinformatics 2010
- P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)  
perspeGktive 2010  
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“
- P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)  
INFORMATIK 2010  
Service Science – Neue Perspektiven für die Informatik  
Band 1
- P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)  
INFORMATIK 2010  
Service Science – Neue Perspektiven für die Informatik  
Band 2
- P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)  
INFORMATIK 2010  
Business Process and Service Science – Proceedings of ISSS and BPSC
- P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)  
Vom Projekt zum Produkt  
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik (WI-MAW), Aachen, 2010
- P-179 Stefan Gruner, Bernhard Rumpe (Eds.)  
FM+AM'2010  
Second International Workshop on Formal Methods and Agile Methods
- P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)  
Datenbanksysteme für Business, Technologie und Web (BTW)  
14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)
- P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)  
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft
- P-182 Ronald Maier (Hrsg.)  
6<sup>th</sup> Conference on Professional Knowledge Management  
From Knowledge to Action
- P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)  
Software Engineering 2011  
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)  
Software Engineering 2011  
Workshopband  
(inkl. Doktorandensymposium)

- P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)  
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)
- P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)  
11<sup>th</sup> International Conference on Innovative Internet Community Systems (I<sup>2</sup>CS)
- P-187 Paul Müller, Bernhard Neumair, Gabi Dreö Rodosek (Hrsg.)  
4. DFN-Forum Kommunikationstechnologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn
- P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)  
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden
- P-189 Thomas, Marco (Hrsg.)  
Informatik in Bildung und Beruf INFOS 2011  
14. GI-Fachtagung Informatik und Schule
- P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)  
Enterprise Modelling and Information Systems Architectures (EMISA 2011)
- P-191 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2011  
International Conference of the Biometrics Special Interest Group
- P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)  
INFORMATIK 2011  
Informatik schafft Communities
- P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)  
IMDM 2011
- P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)  
Informationstechnologie für eine nachhaltige Landbewirtschaftung Fokus Forstwirtschaft
- P-195 Neeraj Suri, Michael Waidner (Hrsg.)  
Sicherheit 2012  
Sicherheit, Schutz und Zuverlässigkeit  
Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
- P-196 Arslan Brömme, Christoph Busch (Eds.)  
BIOSIG 2012  
Proceedings of the 11<sup>th</sup> International Conference of the Biometrics Special Interest Group
- P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)  
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur  
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012
- P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)  
Software Engineering 2012  
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)  
Software Engineering 2012  
Workshopband
- P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)  
ARCS 2012 Workshops
- P-201 Elmar J. Sinz Andy Schürr (Hrsg.)  
Modellierung 2012
- P-202 Andrea Back, Markus Bick, Martin Breunig, Key Poustchi, Frédéric Thiesse (Hrsg.)  
MMS 2012: Mobile und Ubiquitäre Informationssysteme
- P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreö Rodosek (Hrsg.)  
5. DFN-Forum Kommunikationstechnologien  
Beiträge der Fachtagung
- P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)  
12<sup>th</sup> International Conference on Innovative Internet Community Systems (I<sup>2</sup>CS 2012)
- P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)  
5<sup>th</sup> International Conference on Electronic Voting 2012 (EVOTE2012)  
Co-organized by the Council of Europe, Gesellschaft für Informatik und E-Voting.CC
- P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)  
EMISA 2012  
Der Mensch im Zentrum der Modellierung
- P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)  
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.  
24.–26. September 2012

- P-208 Ursula Goltz, Marcus Magnor,  
Hans-Jürgen Appelrath, Herbert Matthies,  
Wolf-Tilo Balke, Lars Wolf (Hrsg.)  
INFORMATIK 2012
- P-209 Hans Brandt-Pook, André Fleer, Thorsten  
Spitta, Malte Wattenberg (Hrsg.)  
Nachhaltiges Software Management
- P-210 Erhard Plödereder, Peter Dencker,  
Herbert Klenk, Hubert B. Keller,  
Silke Spitzer (Hrsg.)  
Automotive – Safety & Security 2012  
Sicherheit und Zuverlässigkeit für  
automobile Informationstechnik
- P-211 M. Clasen, K. C. Kersebaum, A.  
Meyer-Aurich, B. Theuvsen (Hrsg.)  
Massendatenmanagement in der  
Agrar- und Ernährungswirtschaft  
Erhebung - Verarbeitung - Nutzung  
Referate der 33. GIL-Jahrestagung  
20. – 21. Februar 2013, Potsdam
- P-213 Stefan Kowalewski,  
Bernhard Rumpe (Hrsg.)  
Software Engineering 2013  
Fachtagung des GI-Fachbereichs  
Softwaretechnik
- P-214 Volker Markl, Gunter Saake, Kai-Uwe  
Sattler, Gregor Hackenbroich, Bernhard Mit-  
schang, Theo Härder, Veit Köppen (Hrsg.)  
Datenbanksysteme für Business,  
Technologie und Web (BTW) 2013  
13. – 15. März 2013, Magdeburg
- P-215 Stefan Wagner, Horst Lichter (Hrsg.)  
Software Engineering 2013  
Workshopband  
(inkl. Doktorandensymposium)  
26. Februar – 1. März 2013, Aachen
- P-216 Gunter Saake, Andreas Henrich,  
Wolfgang Lehner, Thomas Neumann,  
Veit Köppen (Hrsg.)  
Datenbanksysteme für Business,  
Technologie und Web (BTW) 2013 –  
Workshopband  
11. – 12. März 2013, Magdeburg
- P-217 Paul Müller, Bernhard Neumair, Helmut  
Reiser, Gabi Dreö Rodosek (Hrsg.)  
6. DFN-Forum Kommunikations-  
technologien  
Beiträge der Fachtagung  
03.–04. Juni 2013, Erlangen
- P-218 Andreas Breiter, Christoph Rensing (Hrsg.)  
DeLFI 2013: Die 11 e-Learning  
Fachtagung Informatik der Gesellschaft  
für Informatik e.V. (GI)  
8.-11. September 2013, Bremen

- P-222 Reinhard Jung, Manfred Reichert (Eds.)  
Enterprise Modelling  
and Information Systems  
Architectures  
(EMISA 2013)  
St. Gallen, Switzerland  
September 5 - 6, 2013

The titles can be purchased at:

**Köllen Druck + Verlag GmbH**

Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn

Fax: +49 (0)228/9898222

E-Mail: [druckverlag@koellen.de](mailto:druckverlag@koellen.de)

