BUSINESS PROCESS ANALYSIS AUTOMATION FOR FINANCIAL AUDITS

A design science-oriented approach to support internal and external auditors in process audits by using process mining techniques

Kumulative Dissertation

von

Michael Werner

zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)
der Fakultät für Betriebswirtschaft
der Universität Hamburg

Hamburg

Oktober 2014

Tag der Einreichung:	7. Oktober 2014
Tag der Annahme:	15. April 2015
Tag der Disputation:	15. April 2015
Vorsitzender:	Prof. Dr. Mark Heitmann
Erstgutachter:	Prof. Dr. Markus Nüttgens
Zweitgutachter:	Prof. Dr. Tilo Böhmann

Foreword

This thesis presents the results of a challenging research endeavor that lasted over several years and stretched out to numerous different countries that were visited for discussion, reflection and knowledge exchange with international colleagues. The thesis presents innovative solutions to automate the analysis of business processes in the context of financial audits by using process mining techniques. It is set up as a cumulative dissertation that consists of ten published scientific papers. 28 high quality reviews from international fellow researchers have been received in the past years for the submitted papers. These reviews have contributed significantly to the progress of the presented research and several reliable relationships to experts in the domain of process mining and compliance checking have been established across the world. The preparation of the research was extremely work intensive and could not have been accomplished without the support of my family, friends and colleagues. I would like to thank my wonderful wife for her everlasting support. Without her motivation and encouragement I would never have been able to put the necessary dedication into the presented work as it has been the case. She guided me like a warming light through the ups and downs of this academic voyage. I would also like to thank my family and especially my parents who have always provided good advice and assistance. I would like to devote special thanks to my sister who reviewed my papers sometimes even in night work to ensure that my use of the English language is acceptable for international standards. I would also like to thank my PhD supervisors Prof. Dr. Markus Nüttgens and Prof. Dr. Horst Zündorf for providing advice during my PhD studies and especially Prof. Dr. Nüttgens who has always supported my participation in scientific conferences and exchange with international experts. This thesis would not have been written if the related research project had not been initiated by Prof. Dr. Nick Gehrke who also helped me to manage the sometimes difficult entry into the academic career and community. I would like to thank him for his assistance and fruitful discussions. Many persons have contributed to the success of the research that is presented in this thesis. My PhD colleagues Niels Müller-Wickop and Martin Schultz have been valuable companions in our common research project as well as my colleagues at our department at the Business School of the University of Hamburg and from the Department of Informatics at the Nordakademie. I would like to devote special thanks to Boris Böttcher for reviewing the final thesis and many thanks also to the many unknown reviewers who dedicated big amounts of their time to prepare constructive feedbacks that could be used to improve the results that are presented in this thesis.

For my parents Sylvia and Hans-Joachim

I. Abstract German

Enterprise Resource Planning (ERP) Systeme sind in modernen Organisationen heutzutage integraler Bestandteil zur Unterstützung und Automatisierung von Geschäftsprozessen. Unternehmen veröffentlichen Geschäftsberichte, um verschiedene Interessensgruppen über ihre wirtschaftliche und finanzielle Lage zu informieren. Eine wesentliche Datenquelle für die Erstellung dieser Berichte sind die Daten, die in ERP Systemen erzeugt werden. Aufgrund ihrer wesentlichen Rolle für das Wirtschaftssystem werden Jahresabschlussberichte von Wirtschaftsprüfern geprüft. Bilanzskandale der vergangenen Jahre haben gezeigt, dass Wirtschaftsprüfer nicht in der Lage waren, diese zu verhindern oder zumindest Verstöße frühzeitig aufzudecken. Ein wichtiger Bestandteil der Prüfung von Jahresabschlüssen ist die Prüfung von Geschäftsprozessen und relevanten internen Kontrollen. Die Prüfung von Geschäftsprozessen ist in der Annahme begründet, dass wohlkontrollierte Geschäftsprozesse zu vollständigen und richtigen Buchungseinträgen auf den Finanzkonten führen. Trotz der zunehmenden Integration von Informationstechnologie für die Automatisierung von Geschäftsprozessen verwenden Wirtschaftsprüfer weiterhin vorwiegend traditionelle und manuelle Prüfungsprozeduren, um die Prüfungen durchzuführen. Diese Prozeduren sind zeitintensiv und fehleranfällig. Das Ergebnis ist ein Ungleichgewicht zwischen automatisierter Transaktionsverarbeitung auf Seiten der Unternehmen und manuellen Prüfungsprozeduren auf Seiten der Wirtschaftsprüfer, welches zu ineffizienten oder ineffektiven Prüfungen führt. Der Einsatz von automatisierten Prüfungsprozeduren würde dieses Ungleichgewicht reduzieren. Diese Dissertation folgt einem gestaltungsorientierten Forschungsansatz und stellt unterschiedliche Artefakte vor, die es erlauben, Geschäftsprozessmodelle automatisiert mit Hilfe von Process Mining Techniken zu generieren. Solche Techniken analysieren die Ereignisdaten, die im Zuge der Transaktionsverarbeitung erzeugt werden. Diese Arbeit beschreibt als wesentliches Ergebnis einen Multilevel Process Mining (MLPM) Algorithmus, der speziell für den Einsatz in Jahresabschlussprüfungen entwickelt wurde. Verschiedene weitere Artefakte, die in dieser Arbeit präsentiert werden, sind jedoch auch für andere Einsatzszenarien nützlich. Der vorgestellte Algorithmus vereinigt die Kontrollfluss- und Datenflussperspektive. Er arbeitet auf verschiedenen Abstraktionsebenen, erzeugt präzise und passende Prozessmodelle, verarbeitet nicht-beschriftete und nicht-lineare Ereignisdaten aus ERP Systemen als Eingabedaten und verwendet Datenbeziehungen, um den Kontrollfluss abzuleiten. Sein Einsatz kann die Analyse von Geschäftsprozessen verbessern, die ein wichtiger Bestandteil von Prozessprüfungen ist. Die Automatisierung bestimmter Prozeduren für die Prüfung von Geschäftsprozessen wird zukünftige Bilanzskandale wahrscheinlich nicht verhindern können. Aber der Einsatz des vorgestellten Algorithmus kann Prozessprüfungen verbessern und Prüfungsressourcen freisetzen, die derzeit für die Prüfung von Standardgeschäftsvorfällen verwendet werden, um von Standardverfahren abweichende Transaktionen zu prüfen, die in der Regel ein wesentlich höheres Risiko aufweisen als Standardtransaktionen.

II. Abstract English

Enterprise resource planning (ERP) systems are key components in modern organizations to support and automate the operation of business processes. Companies publish financial reports to inform stakeholders about the economic and financial performance of the organization. A major data source for preparing these reports is the data that is produced by ERP systems. Due to their important role in the economic system financial reports are audited by public accountants. Accounting scandals in recent years have shown that auditors have not been able to prevent these scandals or at least to indicate any violations before the actual collapse. An important part of financial audits is the audit of business processes and related internal controls. The rationale for auditing business processes is the assumption that well-controlled business processes lead to complete and correct postings on the financial accounts. Despite the increasing integration of information technology for the automation of business processes public accountants primarily still use traditional and mostly manual audit procedures to carry out their process audits. These procedures are time-consuming and error-prone. The result is an imbalance between automated transaction processing on the companies' side and manual audit procedures on the auditors' side leading to inefficient or ineffective audits. The application of automated audit procedures would reduce this imbalance. This thesis follows a design science-oriented research approach and introduces several artifacts that can be used to create business process models by using process mining techniques. Such techniques analyze the event log data that is recorded during the processing of business transactions. This thesis presents a Multilevel Process Mining (MLPM) algorithm that has been especially tailored for the use in financial audits. Several other presented artifacts are also useful for other application areas. The algorithm integrates the control flow and data flow perspective. It operates on different abstraction levels, creates precise and fitting process models, accepts unlabeled and non-linear event logs from ERP systems as input, and considers data relationships to infer the control flow. Its application can improve the analysis of business processes which is an important part in process audits. The automation of certain process audit procedures will most likely not prevent accounting scandals in the future. But it can be used to improve process audits and to set free resources from auditing standard business transactions that can then be spent on the auditing of non-standard transactions that commonly exhibit a much higher risk than standard transactions.

III. Content

IV. List of Figures

V. List of Tables

VI. List of Abbreviations

ACE	Automated Controls Evaluator
ACL	Audit Command Language
ASC	Accounting Standards Codification
AS	Auditing Standard
BI	Business Intelligence
BPM	Business Process Management
BPMN	Business Process Model and Notation
CAATs	Computer-Assisted Audit Techniques
CORE	Computing Research & Education
CPN	Colored Petri Nets
DAAD	Deutscher Akademischer Austauschdienst
DFG	Deutsche Forschungsgemeinschaft
DRSC	Deutsches Rechnungslegungs Standards Committee
DSR	Design Science Research
EPCs	Event Driven Process Chains
ERA	Excellence in Research for Australia
ERM	Entity-Relationship-Model
ERP	Enterprise Resource Planning
FASB	Financial Accounting Standards Board
FPN	Financial Petri Net
GAAP	Generally Accepted Accounting Principles
GUI	Graphical User Interface
HTD	Human-Technical Dimension
IAASB	International Auditing and Assurance Standards Board
IASB	International Accounting Standards Board
IDE	Integrated Development Environment
IDEA	Interactive Data Extraction and Analysis
IDW	Institut der Wirtschaftsprüfer in Deutschland e.V.
IDW-PS	IDW-Prüfungsstandard
IFAC	International Federation of Accountants
IFRS	International Financial Reporting Standards

ISA	International Standards on Auditing
MLPM	Multilevel Process Mining
PN	Petri Nets
PCAOB	Public Company Accounting Oversight Board
RCD	Research Contribution Dimension
RQD	Research Question Dimension
SOX	Sarbanes-Oxley-Act
SQL	Structured Query Language
US-GAAP	United States Generally Accepted Accounting Principles
VHB	Verband der Hochschullehrer für Betriebswirtschaft e.V.

1 Introduction

Enterprise resource planning (ERP) systems are key components in modern organizations. They are a specific type of enterprise information systems that are primarily used to support and automate the operation of business processes. As they process business transactions they also create data that provides information about the economic and financial performance of an organization. Companies use this data source to prepare financial reports. Such reports get published as financial statements in periodic time intervals. These reports play a critical role for the smooth functioning of our economic system because they are an important prerequisite for stakeholders to direct their decisions. Due to their informative significance governments and regulatory institutions have issued laws and regulations that intend to safeguard the correctness of published financial statements (e.g. Deutscher Bundestag, 2013, para. 316–324; United States Congress, 2012a, 2012b). They have entrusted public accountants to carry out audits for ensuring that accounting standards are adhered to and that the published information is free of material misstatements.

Accounting scandals in recent years (e.g. Enron 2001, MCI WorldCom 2002, Parmalat 2003, Lehman Brothers 2008, Fannie Mae 2008, Satyam 2009, Olympus 2011 or HRE 2011) have shown that auditors have not been able to prevent these scandals or at least to indicate any violations before the actual collapse. This raises the question of how auditors can be supported to improve their audits. An important part of financial audits is the audit of business processes and related internal controls (IDW, 2009; IFAC, 2012; United States Congress, 2002). The rationale for auditing business processes is the assumption that well-controlled business processes lead to complete and correct postings on the financial accounts. Companies use information systems to support or automate the operation of their business processes. These produce increasing data volumes and become more and more complex with the increasing integration of information technology. Despite this technological progress public accountants primarily still use traditional and mostly manual audit procedures to carry out their process audits. These procedures become inefficient or ineffective in audit environments that are characterized by a high integration of information systems for the automation of transaction processing. The result is an imbalance between automated transaction processing on the companies' side and manual audit procedures on the auditors' side.

A solution to moderate this imbalance would be the application of automated audit procedures. Business processes are currently audited with the use of manual audit procedures like interviews and inspection of selected documents. These procedures are time-consuming and error-prone. Interviews become ineffective if the interviewee does not have a comprehensive knowledge of the relevant process activities. This can be the case if certain activities are automatically executed by an information system without any human interaction. It can also be doubted that the manual inspection of a small sample of transactions is adequate if millions or even billions of transactions are executed in a company every day as, for example, in the telecommunication industry. The application of automated audit procedures is always possible when the mere number of processed transactions makes manual audit procedures inappropriate. Whenever this is the case information systems must be involved that support or automate the processing. Otherwise the number of transactions could still be effectively audited with manual audit procedures. ERP systems produce data that is stored in the systems' databases. The stored data includes the journal entries and other logging information that can be used to reconstruct relationships between the stored data entries and their corresponding originally executed process activities. The content of the stored data in ERP systems that is related to financial transactions is generally suitable for automated audit procedures if such techniques are able to analyze and exploit the available data source.

Data analysis techniques that can be used for this purpose have to fit the specific requirements that are relevant for financial audits. This thesis deals with the question as to how data analysis techniques can be used in the context of financial audits to automate the analysis of business processes. The analysis of relevant business processes is the first step in a process audit to get an understanding about its structure, the data volumes that are processed and how the business processes relate to the financial accounts. The research that is presented in this thesis is embedded in a broader research project that lasted several years with the objective to develop tools and methods for the automation of process audits in financial audits.[1] The presented research focusses on the discovery and analysis of process models by using process mining techniques. The source data are the journal entries that are stored in ERP systems. The research is complementary to research efforts made by other researchers that cover the adequate representation of information in process models for the purpose of financial audits (Müller-Wickop, 2014) and the integration of internal controls (Schultz, 2015).

The main challenges for the development of the presented results are the specific requirements from the application domain that have to be accounted for as well as the complexity and volume of the source data. Process Mining is a research field that has emerged in the late 1990s. Research on process mining has matured in the past decade with the development of powerful general purpose mining algorithms that use simple deterministic (van der Aalst et al., 2004), heuristic (Günther and van der Aalst, 2007; Weijters et al., 2006) or genetic (de Medeiros, 2006) approaches. A fundamental challenge in process mining is to balance competing model criteria (van der Aalst et al., 2012, sec. C6; van der Aalst, 2011a, chap. 5.4.3) such as fitness, precision, simplicity and generalizability (Rozinat, 2007; Rozinat et al., 2008). Auditors require highly fitting and precise process models to avoid false negative and false positive audit results. The source data that is available in ERP systems and that can be used for financial audits is different compared to traditional event logs that are used for process mining because it is unlabeled and not linear (Werner and Nüttgens, 2014). A further challenge is the integration of the data perspective to model the relationship between process activities and financial accounts (Werner, 2013) and the ability to inspect created models at different levels of abstraction to trace a data entry from its point of origin to the final output on the financial accounts. The data perspective has widely been neglected in the academic community so far (de Leoni and van der Aalst, 2013; Stocker, 2012) and contemporary general purpose mining algorithms are not designed to provide process models at different abstraction levels.

The main research output that is described in this thesis is a special purpose process mining algorithm that can be used in the context of financial audits. It integrates the control flow

[1] Virtual Accounting Worlds sponsored by the German Federal Ministry of Education and Research (grant number 01IS10041) (University of Hamburg, 2014).

and data flow perspective, operates on different abstraction levels, creates precise and fitting process models, accepts unlabeled and non-linear event logs as input, and considers data relationships to infer the control flow. The application of the introduced artifacts will probably not prevent accounting scandals. But it would enable public accountants to audit standard transactions very efficiently and it would free resources that can be spent on auditing non-standard and high risk transactions and thus improve the overall audit process and results.

The research presented in this thesis follows a design science research approach (DSR). This approach was chosen because of the proximity of the investigated research questions to practical problems and the intention to develop artifacts that have a high contribution and relevance for the application domain. DSR commonly consists of four different research phases: analysis, design, evaluation and diffusion (Österle et al., 2010). The structure of this thesis follows these phases. It starts with a discussion of the research area and objectives in chapter 2 and the research structure and methodology in chapter 3. The different research phases were conducted in iterative cycles resulting in different research artifacts that were created in the overall research effort. The outputs of one iteration served as the input for the next iteration until a satisfying mining algorithm was designed. The analysis, design and evaluation results of each cycle are described in the corresponding chapters 4, 5 and 6. Chapter 4 deals with the analysis phase. It provides an overview of related scientific work in chapter 4.1 and the results of the requirement analysis in chapter 4.2. Chapter 5 describes the design phase and the different research artifacts that have been developed. Each main artifact is described in one of the chapters 5.1 to 5.6. Chapter 6 illustrates the different evaluation approaches and results that have been achieved with the description of the general experimental setup in chapter 6.1, the discussion of conducted simulations in chapter 6.2 and the presentation of the analysis results on mined process models in chapter 6.3. Chapter 6.4 summarizes how the requirements identified in the analysis phase have been satisfied. This thesis has been prepared as a cumulative dissertation. The research results have been published in ten scientific articles that form the core of this thesis. They are included in appendix A. Chapter 7 illustrates how these publications contributed to the diffusion of the research results as the last phase of design science-oriented research. The thesis closes with a summary in chapter 8.1, a discussion on identified limitations in chapter 8.2 and an outlook to future research in chapter 8.3.

2 Research Area and Objectives

2.1 Financial Audits

Financial audits are important for the smooth functioning of economic markets. They are a control mechanism to prevent the publication of false financial information. The reliability of published financial statements is crucial for stakeholders like shareholders, creditors, tax and regulatory authorities, employees, clients, financial analysts or competitors (Küting and Reuter, 2004) to direct their decisions. Organizations are legally obliged to prepare their financial statements fairly and truthfully. The requirements are specified in national laws such as the Handelsgesetzbuch (Deutscher Bundestag, 2013) in Germany or the Securities Act and the Securities Exchange Act (United States Congress, 2012a, 2012b) in the United States of America. Companies are required to apply generally accepted accounting principles (GAAP) and accounting standards when preparing their financial statements. This ensures that the published statements display correct and comparable information to the addressees. National governments have entrusted specific entities to issue accounting standards. Such standard setting bodies are the International Accounting Standards Board (IASB) on the international level or the Deutsches Rechnungslegungs Standards Committee (DRSC) in Germany and the Financial Accounting Standards Board (FASB) in the United States of America on the national level. They issue the International Financial Reporting Standards (IFRS), the Deutsche Rechnungslegungs Standards (DRS) and the FASB Accounting Standards Codification (ASC). In order to protect addressees from misinformation the financial statements are audited by public accountants. The obligation to engage registered public accountants for the auditing of financial statements is generally mandated by law. Public accountants act as referees ensuring the adherence to laws, regulations and accounting standards by auditing financial statements. They assess if the audited statements give a fair and true view of the financial situation of a company and if the statements are free of material misstatements. Standards on auditing provide guidance for conducting financial audits. They are issued by regulatory bodies such as the International Auditing and Assurance Standards Board (IAASB) for the International Standards on Auditing (ISA), the Institut der Wirtschaftsprüfer (IDW) for the IDW-Prüfungsstandards (IDW-PS) or the Public Company Accounting Oversight Board (PCAOB) for the Auditing Standards (AS). Laws, regulations and standards differ between countries. But a convergence has been taken place over recent years between internationally significant accounting frameworks (Schipper, 2005).

2.2 Contemporary Audit Approaches and Tool Support

Auditing standards require the application of a risk based audit approach that takes into account the internal control framework over relevant business processes and underlying information systems. The ISA 315 states that:

> "The objective of the auditor is to identify and assess the risks of material misstatement, whether due to fraud or error (…) through understanding the entity and its environment, including the entity's internal control (…)" (IFAC, 2012, sec. 3).

Business processes play a significant role in financial audits. The basis for the consideration of business processes is the assumption that well-controlled business processes lead to correct entries in the financial accounts. It is much more efficient to investigate the structure of business processes and related embedded controls than to inspect single transactions. ISA 315 explicitly refers to business processes and related information systems: "The auditor shall obtain an understanding of the information system, including the related business processes, relevant to financial reporting (…)" (IFAC, 2012, sec. 18). The focus on business processes and internal controls is a consequence of several accounting scandals at the beginning of the new millennium that resulted in legislation such as the Sarbanes-Oxley-Act (SOX) (United States Congress, 2002). Similar requirements can be found in national audit standards such as the IDW-PS 261 (IDW, 2009) or the Auditing Standard No. 12 (PCAOB, 2010).

Primary audit procedures to test internal controls and relevant business processes are inquiries of staff members, observation, inspection of documents or reports and the tracing of transactions through the information system relevant to financial reporting (IFAC, 2012, sec. A74). ISA 330 also states that computer-assisted audit techniques (CAATs) may be used to obtain additional evidence. CAATs are software tools that assist the auditor. Braun and Davis discuss that one possible interpretation of CAATs refers to "any use of technology to assist in the completion of an audit" (Braun and Davis, 2003, p. 726). This is a very wide definition and it would mean that simple project management and documentation tools could also be considered as CAATs. It is therefore appropriate to limit the use of the term to "tools and techniques employed to audit computer applications and to tools and technique that extract and analyze data from computer applications" (Braun and Davis, 2003, p. 726). Examples of commercial CAATs are ACL (Audit Command Language) or IDEA (Interactive Data Extraction and Analysis). Accounting firms like PwC also use proprietary software like the PwC SAP ACE (Automated Controls Evaluator). Such tools support certain audit procedures and provide functionality for data queries, sample extractions and statistical analysis or specific purposes like user access or segregation of duties analysis. The usage of CAATs is relatively low (Bierstaker et al., 2014) and they only support certain specific tasks in financial audits. None of currently available CAATs support process audits in a holistic manner.

2.3 Imbalance between Automated Transaction Processing and Manual Audit Procedures

Accounting scandals in recent years (e.g. Enron 2001, MCI WorldCom 2002, Parmalat 2003, Lehman Brothers 2008, Fannie Mae 2008, Satyam 2009, Olympus 2011 or HRE 2011) have shocked the economic markets. Public accountants have not been able to prevent one of these scandals or were able at least to reveal any violation before the actual collapse. Companies use information systems to automate the processing of business transactions resulting in increasingly complex processes and enormous data volumes that exhibit characteristics of Big Data (Chen et al., 2012) in terms of velocity of accumulations and volume.

The task of auditing business processes is getting more and more challenging with the increasing integration of information systems for the automation of transaction processing and the growing amount of produced data. Traditional audit procedures like interviews and inspections of selected documents become inefficient or even ineffective in such audit environments (Werner and Gehrke, 2011). Interview partners may no longer have overall information about a business process if parts of it are operated in an automated way and nontransparent in the information system without any human interaction. Furthermore it is questionable if the inspection of relatively few samples is a sufficient audit procedure when millions of transactions are processed. Software tools are rarely used and only for very specific tasks. This situation leads to an imbalance between automated transaction processing on the companies' side and manual audit procedures on the auditors' side which is illustrated in Figure 1. The outcomes are inefficient or even ineffective audits.

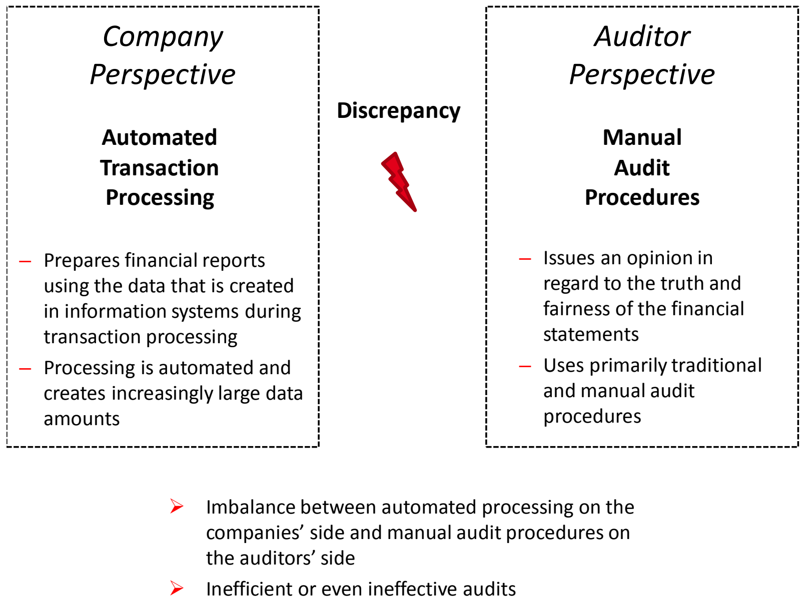


Figure 1 Imbalance Between Automated Transaction Processing and Manual Audit Procedures in Financial Audits

2.4 Research Questions and Objectives

A solution to reduce the imbalance between automated transaction processing on the companies' side and manual audit procedures on the auditors' side would be the application of automated audit procedures that exploit the data that is created and recorded in the course of transaction processing. The advantage of this data is that it is recorded automatically and independently from any involved person and therefore provides a much more reliable data source compared to the data sources that are used in traditional audit procedures (Jans et al., 2010). Automated audit procedures would have to be able to operate with the recorded data and comply with the specific requirements that generally have to be accounted for in financial audits. The overall research question that has to be answered for this aim can be formulated as followed:

Overall research question: How can data analysis techniques be used to automate the audit of business processes in the context of financial audits?

The overall research question is very broad and has been addressed by different researchers that focused on different research aspects to find solutions to this question in a common research effort. This thesis deals with the aspect of creating and analyzing business models by exploiting the source data which is stored in information systems that support or automate the processing of business transactions. The creation of process models from the source data is a core requirement for the development of automated audit procedures that support the audit of business processes in a holistic manner. The specific research question that is addressed in this thesis can be phrased as followed:

Specific research question: How can reliable process models be automatically reconstructed by analyzing data stored in information systems that process financially relevant transactions?

The objective of the described research is the development of research artifacts in form of construct, models, methods and instantiations that can be used to create and analyze process models on the basis of financially relevant data recorded in the source systems. Process mining is a business intelligence approach that provides powerful mining algorithms that are able to create process models from recorded event logs (van der Aalst, 2011a). It therefore serves as a starting point and important knowledge base for the research presented in this thesis.

3 Research Structure and Methodology

3.1 Epistemological and Ontological Orientation

The research approach and methodology for each research effort has to fit the intended purpose. Different paradigms and approaches prevail in the information systems research community. A fierce controversy has taken place in the international information systems research community in recent years between advocates of different research paradigms (Baskerville et al., 2010; Österle et al., 2010). Two differing research paradigms in information systems research are positivism and interpretivism. They differ in terms of the ontological and epistemological assumptions (Niehaves, 2007). Interpretivist research focusses on the perception of humans and the individuality of cognition. The basic ontological assumption is that a single objective world might indeed not exist and that hence objective cognition is impossible. Positivist research prevails in natural sciences and it is characterized by the ontological assumption that a real world exists that can be described unambiguously in combination with the epistemological assumption that objective cognition is hence possible. Social or behavioral science research tends to be descriptive in nature and to follow an interpretivist paradigm. Behavioral science research and interpretivist-oriented research methods are especially prevalent in the Anglo-Saxon research community (Palvia et al., 2004, 2003; Schauer, 2011; Wilde and Hess, 2007). Design science research is a research approach that has gained increased attention over the last decade in the international scientific community as an important research approach (Vaishnavi and Kuechler, 2004). It is normative in nature and DSR researchers tend to follow a positivist research paradigm. DSR is prevalent in German speaking countries (Wilde and Hess, 2007) where information systems research has traditionally been closely related to the natural and engineering sciences (Schauer, 2011) but also an important research approach in the Anglo-Saxon research community (Baskerville et al., 2010). DSR is characterized by the duality of the epistemological and design objective (Riege et al., 2009). It intends to create research artifacts in forms of constructs, models, methods (March and Smith, 1995) and theories (Gregor, 2006; Gregor and Hevner, 2013) that on the one hand contribute to the scientific knowledge base and that are on the other hand useful for the application domain (Hevner et al., 2004).

The research presented in this thesis intends to develop solutions to improve the audit of business processes in financial audits. The research questions are closely related to challenges in practice. DSR is therefore perceived as an adequate research approach to develop the intended research outputs. Several scholars have pointed out that focusing on a single research approach or research method might lead to paradigmatic bias (Mingers, 2001) and that an interaction between DSR and social science research is necessary to achieve progress in the scientific information systems discipline (Gregor and Baskerville, 2012). We therefore relied on a variety of research methods with differing paradigmatic orientation to prevent any methodological or paradigmatic bias and to benefit from observing the phenomena under investigation from different scientific angles. Detailed information on the research structure and methodology is described in chapters 3.3 and 3.4.

3.2 Research Scope and Subjects

3.2.1 Segmentation Framework

The presented research was conducted in the research project Virtual Accounting Worlds. This project was sponsored by the German Federal Ministry of Education and Research (grant number 01IS10041) and followed the objective to develop tools and methods to support auditors in process audits in the context of financial audits (University of Hamburg, 2014). The overall research task was divided into three different research sub-scopes that were assigned to three different researchers. A segmentation framework was developed to confine and assign the different work tasks to the involved researchers (Werner et al., 2014).

The framework itself was developed as part of the research project. It is used in this section to describe the research scope that is covered by this thesis. The framework is extensively described in (Werner et al., 2014). It consists of three dimensions that are suitable to divide the overall research task in well-defined and manageable research segments.

The first dimension refers to the human-technical interaction. The main subjects of interest in information systems research are information technologies and the man/machine interaction. The distinguishing characteristic of the information systems research discipline is the investigation of phenomena that emerge from the interaction of humans and technology in socio-technical systems (Gregor, 2006). Categorizations for different levels of information technology and human interaction can be found in various models. A common model from the field of information management is presented by Krcmar (Krcmar, 2010). The model was adapted from (Wollnik, 1988) and distinguishes between three different levels of management tasks. The lowest level contains tasks for the management of the technical infrastructure that are necessary for the use of information and communication technology at the higher levels. The second level deals with the management of information systems and includes the management of data, processes, applications and their lifecycles. At the highest level reside the tasks for the management of the information economy. The main objective of the tasks at this level is the management of the resource information, its supply, demand and usage. The two lower levels are mainly concerned with the technical aspects of information management. The human aspect is considered at the highest level where the requirements for the lower levels are defined on the basis of the needs of human information recipients and users of the applications, processes, data and technology that are located on the lower levels. The model of information management addresses both elements that are subject to research in information system science: information systems and human interaction in socio-technical systems. It is therefore useful for the categorization of research content because each researched artifact in design science-oriented research can be characterized if it addresses one or more of the different levels. The original model represents applications, data and processes at the same level. Prominent conferences in the information systems science area like the Business Process Management conference (BPM, 2014), comprehensive publications (Weske, 2012) and extensive reviews (van der Aalst, 2013, 2012) show that business processes are a key component in information systems research. We therefore considered it appropriate to divide the level of information systems into the two levels of software applications and processes. The process level is the connecting layer where process participants use components from the lower application level to satisfy information demands from the higher level.

The resulting four levels of the Human-Technical Dimension (HTD) are very broad. Although they can be used to distinguish research content on a technical vs. human-interaction dimension they are not sufficient to divide the research content into manageable segments (Werner et al., 2014). Hevner et al. present a research framework for information systems research that provides an illustration of how different concepts that are relevant for research projects relate to each other (Hevner et al. 2004). It describes the relationships between the main research activities for design science (build and evaluate) and behavioral science (develop and justify) research, the environment and the knowledge base. The environment or application domain defines the problem space. The phenomena of interest for design science-oriented research should be derived from the environment. The knowledge base represents the pool of already existing scientific expertise. Each research project should take into account already existing knowledge to execute the research activities, provide useful artifacts to the application domain and add additional generalized knowledge to the knowledge base. The objective to contribute to the application domain on the one hand and the scientific knowledge base on the other hand characterizes design science-oriented research projects (Riege et al., 2009). The distinction between the research contribution target domains can be used as a second Research Contribution Dimension (RCD) for the categorization of research content.

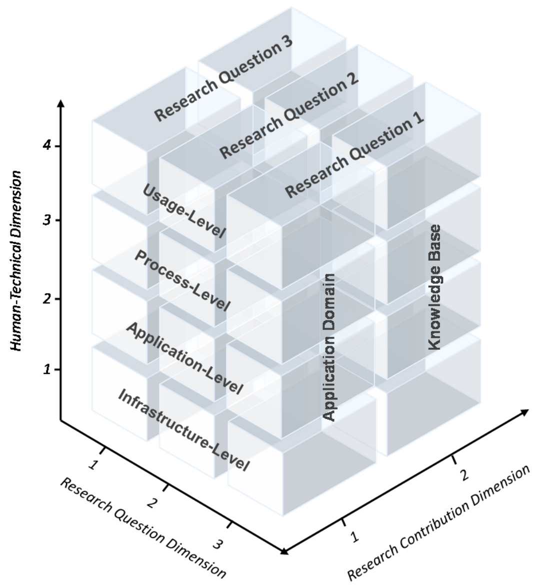


Figure 2 Segmentation Framework (adapted from Werner et al., 2014, p. 8)

Each research project addresses an overall research question. The research questions in large research projects are, as a rule, complex. Otherwise it would be debatable if such a project has the characteristics of a large project in the first place. Complex research questions can usually be divided into detailed lower-level research questions. These research questions can be used as a categorization criterion for a third Research Question Dimension (RQD). Figure 2 shows the segmentation framework with all three dimensions. It illustrates how separate segments emerge based on the different dimension categories. Each segment can be referenced by using its x- (RQD), y- (HTD) and z-coordinates (RCD) in the cube. The reference model shown in Figure 2 has to be instantiated to be useful for a specific research project which is described for the research at hand in in the next chapter.

3.2.2 Research Scope

Figure 3 shows the instantiated framework that was used to divide the overall research task into manageable research segments and to confine the scope for the research presented in this thesis. The primary application domain for the research work at hand is financial audits (compare chapter 2.1). It represents the first category in the Research Contribution Dimension.

The thesis focusses on presenting solutions for producing reliable process models automatically by analyzing data stored in information systems that process financially relevant transactions. Process models are abstractions from reality that are used to graphically represent business processes. A business process is of a set of related activities that are performed in an organizational and technical environment to realize a business goal (Reichert and Weber, 2012; van der Aalst and Stahl, 2011; Weske, 2012; Workflow Management Coalition, 1999). “A business process model consists of a set of activity models and execution constraints between them.” (Weske, 2012, p. 7). Modeling is the activity that leads to the definition of models (Hansen and Neumann, 2009). Process models are usually defined by using modeling languages like Business Process Model and Notation (BPMN), Event Driven Process Chains (EPCs) or Petri Nets. Business process modeling is a fundamental part of Business Process Management (BPM). It is therefore perceived as one important scientific knowledge base.

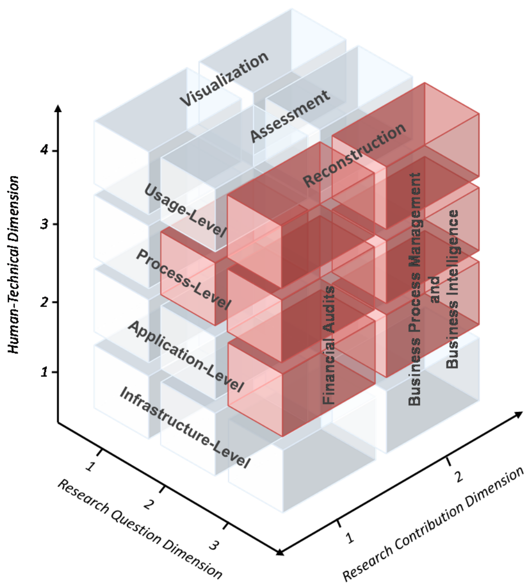


Figure 3 Research Scope (adapted from Werner et al., 2014, p. 12)

The fundamental idea to develop the solutions presented in this thesis is the application of process mining techniques. Process mining allows creating process models from recorded event log data (van der Aalst, 2011a). It is a Business Intelligence (BI) approach that bridges the gap between BPM and BI (van der Aalst, 2011b). BI can be seen as an amalgamation of business economics, operations research, data mining, statistics and data warehousing

(Müller and Lenz, 2013) and it is traditionally used to support decision making processes (Turban et al., 2007; Vercellis, 2009). The aim of process mining as an BI approach is to discover, analyze and enhance process models (van der Aalst et al., 2012). The segmentation framework was used to define and structure the research task for the overall research project. This thesis focusses on process mining techniques but it also relates to other data analysis techniques that were used to statistically analyze the source data and produced process models. In order not to be too restricted and also to account for the other research tasks from the overall research project that are not part of this thesis BI was chosen as the more general term as a relevant scientific knowledge base. BPM and BI in combination form the second category on the Research Contribution Dimension as illustrated in Figure 3.

The categories for the third Research Question Dimension were derived from breaking the overall research question (compare chapter 2.4) down into three detailed sub-questions. The detailed, lower-level research questions were labeled with the keywords reconstruction, assessment and visualization. They were formulated by logically reasoning what kind of questions on a more specific level had to be answered to find solutions for the overall research question. They were formulated as follows:

(1) Reconstruction:	How can reliable process models be automatically reconstructed by analyzing data stored in information systems that process financially relevant transactions?
(2) Assessment:	How can process models be automatically assessed from an audit perspective by integrating control data that is stored in the source systems?
(3) Visualization:	How can process models be graphically represented to display information that is relevant to auditors and that can be applied in real audit environments?

Figure 3 shows the precise scope of the research work that is presented in this thesis. The relevant research segments are highlighted in red. It is necessary to consider the application level (segments (3,2,1) and (3,2,2)) because software applications like ERP systems that process financially relevant transactions provide the source data that is necessary to reconstruct process models. The focus of the research lies on the reconstruction (segments (3,3,1) and (3,3,2)) and analysis (segment (2,3,1)) of process models for financial audits. The research scope on the horizontal process-level stretches into the vertical assessment category because the analysis of reconstructed process models already provides information that is useful for the assessment of business processes in the context of financial audits. The usage-level is concerned with the demand and supply of information and its consumption. This level is relevant to take primarily formal requirements into account that are important for the adequacy of the produced models (segments (3,4,1) and (3,4,2)). Figure 3 shows that contributions were planned for the application domain (segments (3,2,1), (2,3,1), (3,3,1), (3,4,1)) and the knowledge base (segments (3,2,2), (3,3,2), (3,4,2)).

The infrastructure-level (segments (x,1,z)) was out of scope because it is not necessary to consider this level in order to achieve the research objective as all necessary source data is provided on the application-level. Aspects of process visualization from a usability and end-user perspective (segments (1,y,z)) are not covered in this thesis and neither are aspects for the assessment of business processes that relate to internal controls (segments (2,y,z) except (2,3,1)).

3.2.3 Research Subjects

The research subjects are the data that is produced by information systems during the course of processing financially relevant transactions and the created process models.

The presented research focusses on data produced by ERP systems because they are the predominant type of information systems that are used to support and automate the processing of business transactions (Konradin Mediengruppe, 2011). The produced data from ERP systems is generally suitable for process mining purposes (van der Aalst et al., 2012).

The term model has many different meanings depending on the context and scientific discipline. A model according to the interpretation prevailing in information systems science is a simplified image of a selected part of reality (Heinrich et al., 2004) that serves a specific purpose (Becker et al., 2012). Model and reality are related to each other (model-relation) which means that it is possible to conclude from observed model characteristics on reality and vice versa (isomorphism-relation) (Heinrich et al., 2004).

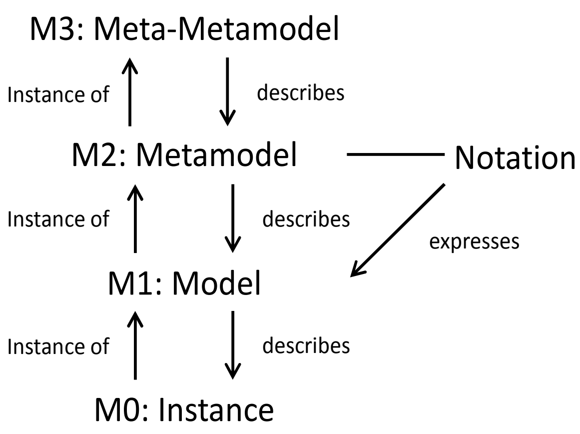


Figure 4 Horizontal Abstraction Levels (Weske 2012 p. 76)

Figure 4 illustrates that models in the context of BPM can be differentiated depending on the level of horizontal abstraction (Weske, 2012). Models (M1) are created by using constructs that are defined by metamodels (M2). These are associated with notations that often exhibit a graphical nature. The Petri Net metamodel (M2), for example, defines that Petri Nets (M1) consist of places and transitions that form a directed bipartite graph. "The complete set of concepts and associations between concepts is called metamodel." (Weske, 2012, pp. 76–77). Constructs of a metamodel (M2) can themselves be defined by a meta-metamodel (M3).

Process models reside on the model-level (M1) and are instances of metamodels (M2) that are expressed using an associated modeling language like BPMN, EPCs or Petri Nets. A process model is an abstraction of a business process and represents a set of similar executed entities of a business process. A process instance model (M0) is an abstraction of a single executed entity of a process model.

> "The instance level reflects the concrete entities that are involved in business processes. Executed activities, concrete data values, and resources and persons are represented at the instance level." (Weske, 2012, p. 75).

Process models and process instance models are especially relevant for the investigated research question. Process models provide an overview of the structure of a business process. Process instance models provide detailed information for each single execution of a business process that is relevant from an audit perspective to inspect individual transactions and associated data values. The two types of models are on the one hand the primary outputs that are produced by the designed research artifact (compare chapter 5) but on the other hand also the subjects of investigation in the evaluation phase (chapter 6).

3.3 Research and Thesis Structure

Different research frameworks exist that provide guidance on how to conduct design science-oriented research.[2] Österle et al. suggest a generally accepted framework that divides DSR into the phases analysis, design, evaluation and diffusion (Österle et al., 2010). The research and thesis structure follows these four different phases. Figure 5 provides an overview of the main artifacts that were developed in the design phase and the research methods that were used in the different research phases. It also shows the diffusion types[3] for the diffusion phase.

[2] Peffers et al. present a research methodology that consists of six phases: (1) identify problem and motivate, (2) define objectives of a solution, (3) design and development, (4) demonstration, (5) evaluation and (6) communication (Peffers et al., 2007, 2006). Their research framework is composed of process elements that have been identified by different scholars working in the information systems (Cole et al., 2005; Hevner et al., 2004; Nunamaker et al., 1991; Takeda et al., 1990; Walls et al., 1992) and engineering (Archer, 1984; Eekels and Roozenburg, 1991) discipline. Österle et al. suggest four phases for DSR: (1) analysis, (2) design, (3) evaluation and (4) diffusion (Österle et al., 2010). Gregor and Baskerville examine the research process from a philosophy of science perspective with the objective to provide a framework for the combination of design science and social science research. The presented research process consists of the phases (A) construct and test artefacts, (B) formulate prescriptive knowledge and theory, (C) study artefact(s) in use, (D) test knowledge of artefacts in use and (E) formulate descriptive knowledge (Gregor and Baskerville, 2012). All authors explicitly emphasize the iterative relationship between the different research steps in each model. Alturki et al. present the most detailed model that consists of 14 research steps (Alturki et al., 2011).

[3] The diffusion types are presented in Figure 5 to provide a complete overview of the research structure. They cannot be considered as research methods and are therefore not described in the following chapter 3.4 but instead in chapter 7.

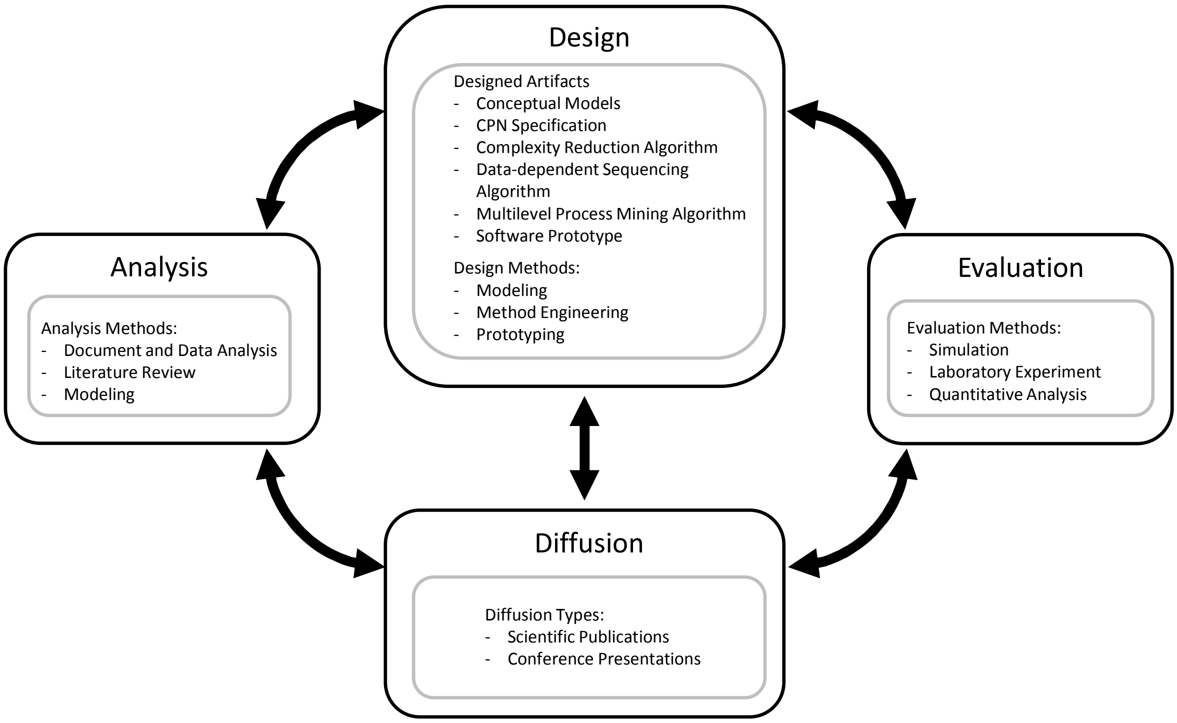


Figure 5 Research Structure

The different phases were undergone in iterative loops where the research outputs of one research cycle formed the input of the following one. Figure 6 shows how the different designed artifacts depend on each other and how the research structure refers to the structure of this thesis.

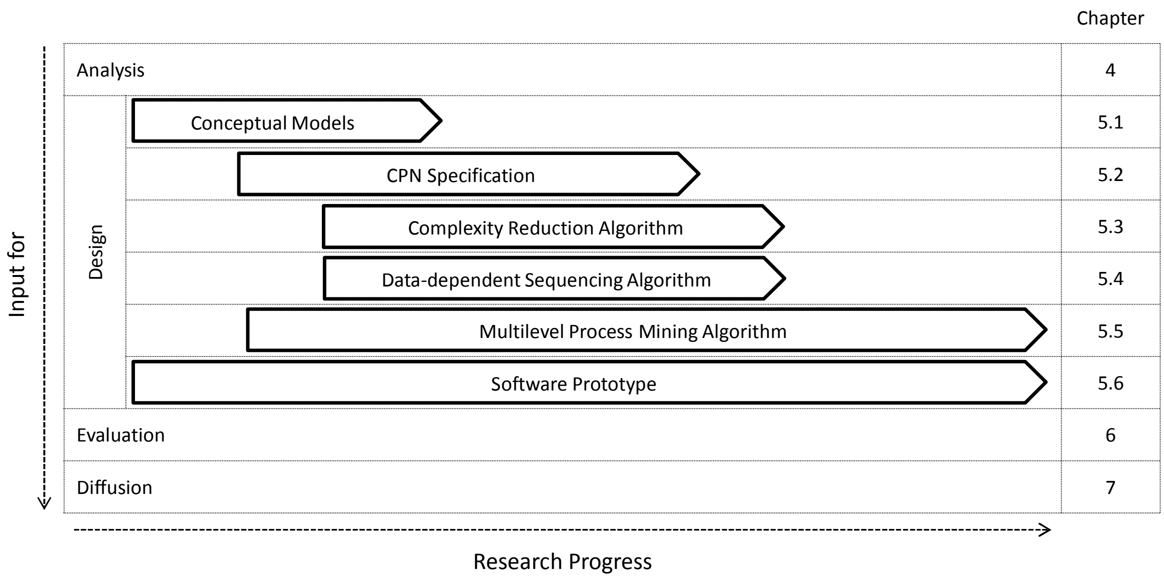


Figure 6 Research Phases, Artifacts and Thesis Structure

3.4 Research Methodology

The term methodology is used ambiguously in information systems science.[4] We use the term in the sense of a combination of research methods that are used for a specific research project. The presented research follows as design science-oriented research approach. An important aspect when following a specific research approach is the prevention of paradigmatic bias that might occur by focusing on a single research paradigm. We therefore used different research methods with diverse paradigmatic orientation to prevent such a bias and to benefit from triangulation (Jick, 1979) and an appropriate mix between qualitative and quantitative research methods (Venkatesh et al., 2013). Figure 5 in chapter 3.3 illustrates the variety of research methods that were used in the different research phases. The following chapters briefly describe these methods.

3.4.1 Analysis Methods

Document and Data Analysis

The primary analysis methods include a literature review and the analysis of source data and documents (Recker, 2012) by inspecting the data, data structure and system documentation from test and productive ERP systems.

Literature Review

The literature review referred to relevant scientific publications but also to laws, regulations and standards that specify the formal requirements that have to be taken into account in the context of financial audits. It was conducted following the guidelines published by various scholars (Fettke, 2006; Rowley and Slack, 2004; vom Brocke et al., 2009; Webster and Watson, 2002). The collection of empirical data regarding requirements directly from the application domain was not in scope of the research presented in this thesis but was conducted in the related research project (Müller-Wickop et al., 2013; Müller-Wickop and Schultz, 2013a; Schultz et al., 2012; University of Hamburg, 2014). The empirical investigation included the application of qualitative research methods such as structured interviews (Gubrium and Holstein, 2002) and the quantitative research methods in the form of surveys (Fowler, 1984). The achieved results were incorporated in this work as part of the literature analysis to achieve an adequate mix between requirements identified using a positivist perspective (objective laws, regulations and standards) and an interpretive perspective (results from interviews and surveys).

Modeling

Models are, besides constructs, methods, instantiations and theories (Gregor, 2006), one of the main research artifacts in DSR (March and Smith, 1995). Two types of models are used in this thesis: conceptual models to define the problem and solution space like entity-

[4] Mingers identifies three widespread and different interpretations for the term methodology (Mingers, 2001).The most general meaning refers to the study of methods. The most specific meaning is related to a particular research study and it refers to the actual research method(s) that are used in a specific piece of research. The third interpretation is a generalization of the second. A methodology is referred to as a specific combination of particular methods that is deliberately designed a priori in the sense of a blueprint that occurs many times in practice.

relationship-models (ERM) in the analysis phase and process models as output of the designed methods and software prototype in the design and evaluation phases. Modeling or conceptual modeling is the research method that was used to create different models. Modeling is an engineering process that creates simplified images of reality by using an inductive approach relying on observations or a deductive approach relying on theories (Wilde and Hess, 2006).

3.4.2 Design Methods

Method Engineering

The primary research methods for designing the research artifacts were method engineering and prototyping. Method Engineering is a commonly used research method in information systems research (Österle et al., 2010; Wilde and Hess, 2007, 2006) for the systematic design of methods (Brinkkemper, 1996). A method in this context consists of different parts (method fragments) that can be combined and reused (Harmsen et al., 1994). A new method can be engineered by combining existing method fragments in a new manner or by developing completely new method fragments.

Prototyping

Prototyping is a software programming approach originating from the area of software engineering (Naumann and Jenkins, 1982). Prototyping consists of the phases: requirement identification, development, implementation, revision and enhancement. The aim is to develop a software artifact that implements the intended core functionality in iterative cycles. It is a common research method in design science-oriented research (Österle et al., 2010; Schauer, 2011; Wilde and Hess, 2007, 2006). This research method is particularly suitable for the research at hand because it allows to design, implement and evaluate different research artifacts that relate to each other as shown in Figure 6.

3.4.3 Evaluation Methods

DSR is often criticized for a lack of research rigor (Österle et al., 2010) and several scholars highlight the importance of rigorous evaluation in design science-oriented research (Riege et al., 2009; Venable et al., 2012). The duality of the design and the epistemological objective (Riege et al., 2009) require the consideration of evaluation methods that can be used to verify the adequateness of the research outputs in regard to the identified research gap and to validate if the developed solutions are able to satisfy the demands from the application domain. Several scholars provide guidelines and frameworks for the selection of appropriate research methods.

The primary evaluation methods that were used in the presented research include simulation, lab experiment and quantitative analyses. Simulations and lab experiments are suitable for artificial ex-post evaluations (Venable et al., 2012). Quantitative analyses using descriptive statistics (Bhattacherjee, 2012) were added to describe and analyze the output data that was generated during the lab experiments. The design and evaluation methods are closely related. The implementation of the designed artifacts in a prototype already serves as a proof of concept from an evaluation perspective because it allows to verify if the theoretical constructs, models and methods can actually be executed. It also provides the foundation for simulations and lab experiments.

Simulation

Simulations are goal-oriented experiments to gain information on models that are difficult to represent as formal models due to their inherent complexity. The models that were created by the software prototype were made of up to several hundred thousand net elements (Werner et al., 2012b). Simulations were used to verify if the produced process models satisfied the identified requirements for example in respect to specific model characteristics like soundness (van der Aalst, 2011a; Weske, 2012).

Laboratory Experiment

The prototype was used in laboratory experiments. The conducted experiments differ from traditional experiments that are commonly used in behavioral science-oriented research. In such experiments one or more independent variables are manipulated by the researcher. Subjects are randomly assigned to different treatment levels and the results of the treatments on outcomes are observed (Bhattacherjee, 2012, chap. 10). The subject in the design science-oriented experiments is the prototype itself and the outcomes that are observed are the produced process models (Riege et al., 2009). The variables are different configuration parameters.

Quantitative Analysis

The output that was produced by the designed artifacts were finally analyzed with the use of descriptive statistics (Bhattacherjee, 2012; Schira, 2009). These analysis methods were used to inspect the voluminous output data on an aggregate level and to gain insights into the distributions and relationship between different model characteristics.

3.4.4 Method Overview

Table 1 provides an overview of the used research methods, paradigmatic orientation and literature references that were primarily used to apply the listed methods. The table shows that a mix between positivist and interpretive oriented research methods has been achieved by using diverse quantitative and qualitative analysis, design and evaluation methods.

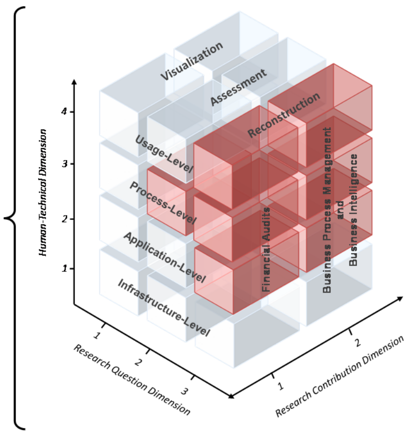
Table 1 Research Methods

Research Method	Paradigmatic Orientation	Type	Primary References
Document and Data Analysis	Positivist	Quantitative / Qualitative	(Bhattacherjee, 2012; Österle et al., 2010; Recker, 2012)
Literature Review	Positivist / interpretivist	Quantitative / Qualitative	(Fettke, 2006; Rowley and Slack, 2004; vom Brocke et al., 2009; Webster and Watson, 2002)
Modeling	Interpretivist	Qualitative	(Österle et al., 2010; van der Aalst and Stahl, 2011; Wilde and Hess, 2007, 2006)

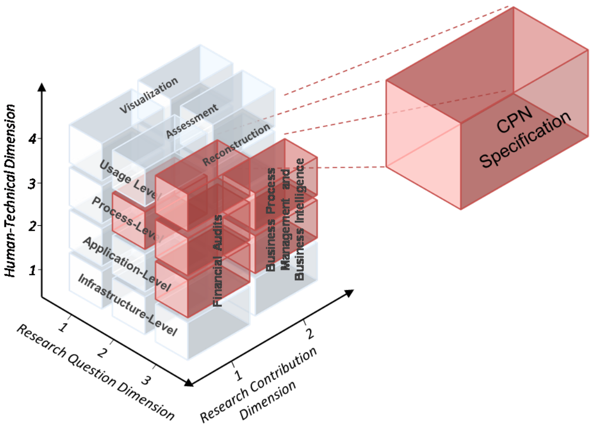
Method Engineering	Positivist	Qualitative	(Brinkkemper, 1996; Harmsen et al., 1994; Österle et al., 2010; Wilde and Hess, 2007, 2006)
Prototyping	Positivist	Qualitative	(Naumann and Jenkins, 1982; Wilde and Hess, 2007, 2006)
Simulation	Positivist	Quantitative	(Riege et al., 2009; van der Aalst and Stahl, 2011; Wilde and Hess, 2007, 2006)
Laboratory Experiment	Positivist	Quantitative	(Bhattacherjee, 2012; Österle et al., 2010; Riege et al., 2009; Wilde and Hess, 2007, 2006)
Quantitative Analysis	Positivist	Quantitative	(Bhattacherjee, 2012; Gujarati and Porter, 2009; Österle et al., 2010; Recker, 2012)

Table 2 lists the different artifacts, related research methods and diffusion types in the left column. The diagrams in the right column illustrate how the artifacts relate to the research segments described in chapter 3.2.2.

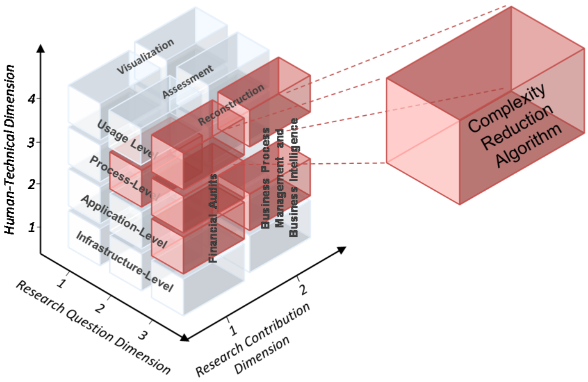
Table 2 Overview of Research Artifacts, Research Methods, Diffusion Types and Research Segments

<table>
<tr><th colspan="2">Research Artifacts, Research Methods and Diffusion Types</th><th>Related Research Segments</th></tr>
<tr><td>Artifact</td><td>Conceptual Models</td><td rowspan="8">
</td></tr>
<tr><td>Segments</td><td>All</td></tr>
<tr><td colspan="2">Research Methods</td></tr>
<tr><td>Analysis</td><td>Document and Data Analysis
Literature Review</td></tr>
<tr><td>Design</td><td>Modeling</td></tr>
<tr><td>Evaluation</td><td>-</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Diffusion Type</td><td>Book Chapters 3 and 4
Conference Paper 5</td></tr>
</table>

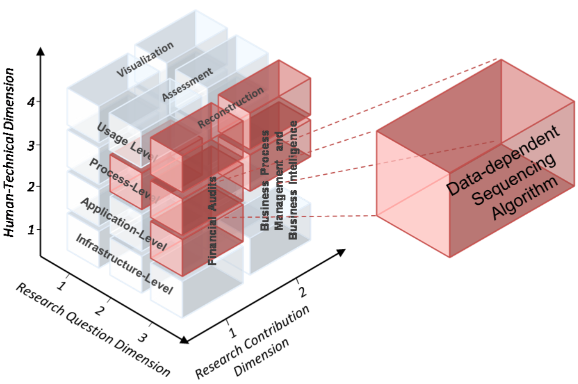
Artifact	CPN Specification
Segments	(3,4,2)
Research Methods	
Analysis	Document and Data Analysis Literature Review
Design	Modeling Prototyping
Evaluation	Laboratory Experiment Simulation
Diffusion Type	Conference Paper 6



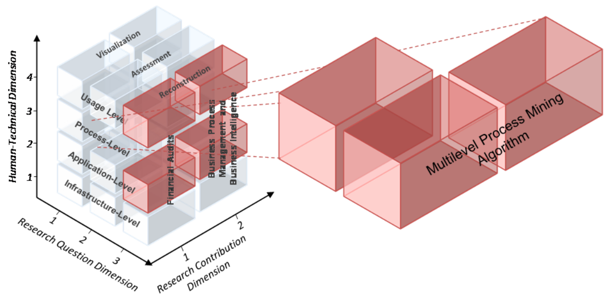
Artifacts	Complexity Reduction Algo- rithm
Segments	(3,3,2)
Research Methods	
Analysis	Document and Data Analysis
Design	Method Engineering Prototyping
Evaluation	Laboratory Experiment Simulation Quantitative Analysis
Diffusion Type	Conference Paper 7



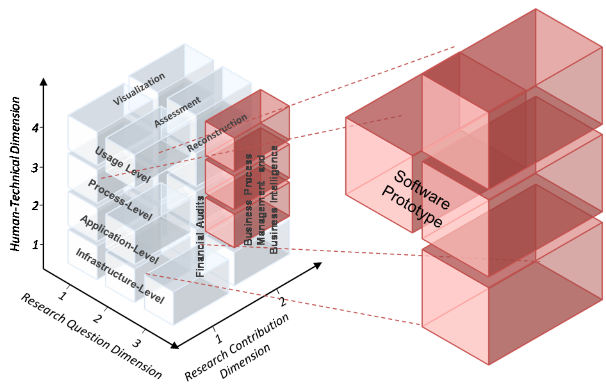
Artifacts	Data-dependent Sequencing Algorithm
Segments	(3,2,2)
Research Methods	
Analysis	Document and Data Analysis
Design	Method Engineering Prototyping
Evaluation	Laboratory Experiment Simulation Quantitative Analysis
Diffusion Type	Conference Paper 8



Artifact	Multilevel Process Mining Algorithm
Segments	(3,3,1),(3,3,2), (2,3,1)
Research Methods	
Analysis	Document and Data Analysis
Design	Method Engineering Prototyping
Evaluation	Laboratory Experiment Simulation Quantitative Analysis
Diffusion Type	Conference Paper 9 Journal Paper 10



Artifact	Software Prototype
Segments	(3,2,1),(3,3,1), (2,3,1),(3,4,1)
Research Methods	
Analysis	-
Design	Prototyping
Evaluation	Laboratory Experiment Simulation
Diffusion Type	Conference Paper 9 Journal Paper 10



3.5 Cumulative Publications

This thesis has been prepared as a cumulative dissertation. The research results that are presented in this thesis have been published in different scientific articles. Table 3 provides information on the different publications and how they relate to the different chapters in this thesis. The first column shows the number of the paper that is used for reference purposes in this thesis. The second column provides a reference to the appendix that contains the complete originally published papers.[5] The third column shows to which chapter the publication primarily refers to. The papers generally follow the publication guidelines for DSR (Gregor and Hevner, 2013). Each paper is a complete scientific publication and does not refer exclusively to a single chapter but also to aspects that are mentioned in various chapters of this thesis. The assignment to specific chapters is only intended to illustrate which core aspects are described in the different publications and how they generally refer to the structure of this thesis.

[5] The papers have been formatted in a consistent way. References to page numbers in this thesis refer to the originally published papers.

Column four shows the full title of each paper and the fifth column provides information on the publication outlet. Two papers were published as book chapters, six in conference proceedings and two in journals.[6] The sixth column provides the reference to the entry in the bibliography. Column seven lists the acceptance rate for scientific papers that were provided by the conference chairs. Columns eight to eleven provide information on the ranking of the publication outlets according to the VHB Jourqual 2.1 (Verband der Hochschullehrer für Betriebswirtschaft e.V., 2011), WKWI (WKWI, 2008), ERA 2010 (Australian Research Council, 2014) and CORE 2013 (CORE, 2014) rankings.

Column twelve describes the review procedure and lists how many different reviews from fellow researchers were received during the review process. The used abbreviations have the following meaning:

DB: Double blinded

B: Blinded

The number of authors that contributed to each paper is listed in column 13. Column 14 shows how many dissertation points can be assigned to each paper using the formula *2 / (number of authors +1)*. Column 15 shows the level of authorship from the author of this thesis that was commonly agreed among all co-authors. Two papers were prepared in single authorship. The average authorship for the remaining papers is 91 %. None of the listed papers is part of any other dissertation thesis.

[6] Paper ten was initially submitted for the IEEE TSC journal in August 2013. It passed the first review round in December 2013, the second in July 2015 before being published in December 2015. The included version refers to the preprint version at the time of submission of this dissertation.

Table 3 Overview of Cumulative Publications

#	Appendix	Primary Related Chapters	Title	Publication Outlet	Reference	Acceptance Rate	Rankings				Review Procedure / Reviews	Number of Authors	Dissertation Points	Authorship	Part of other Dissertations
							VHB JQ 2.1	WKWI 2008	ERA 2010	CORE 2013					
1	10.1	3	Who Is Afraid of the Big Bad Wolf - Structuring Large Design Science Research Projects	22nd European Conference on Information Systems (ECIS)	(Werner et al., 2014)	34 %	B (7.37)	A	A	A	DB (4)	4	0.40	85%	No
2	10.2	4.1	Process Mining	wisu - das wirtschaftsstudium	(Gehrke and Werner, 2013)	[7)]	E (2.86)	-	-	-	B (-)	2	0.67	95%	No
3	10.3	4.2 5.1	Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen	Forschung für die Wirtschaft	(Werner and Gehrke, 2011)	[7)]	-	-	-	-	B (-)	2	0.67	95%	No
4	10.4	4.2 5.1	Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung	Forschung für die Wirtschaft	(Werner, 2012)	[7)]	-	-	-	-	B (-)	1	1.00	100%	No
5	10.5	5.1	Business Process Mining and Reconstruction for Financial Audits	45th Hawaii International Conference on System Sciences (HICSS)	(Werner et al., 2012a)	54 %	C (6.44)	B	A	A	DB (6)	3	0.50	90%	No
6	10.6	5.2	Colored Petri Nets for Integrating the Data Perspective in Process Audits	32nd International Conference on Conceptual Modeling (ER)	(Werner, 2013)	32 %[8)]	B[8)] (7.59)	B[8)]	A[8)]	A[8)]	DB (3)	1	1.00	100%	No
7	10.7	5.3	Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits	33rd International Conference on Information Systems (ICIS)	(Werner et al., 2012b)	29 %[9)]	A[10)] (8.48)	A[10)]	A[10)]	A*[10)]	DB (3)	5	0.33	80%	No
8	10.8	5.4	Improving Structure: Logical Sequencing of Process Models	47th Hawaii International Conference on System Sciences (HICSS)	(Werner and Nüttgens, 2014)	56 %	C (6.44)	B	A	A	DB (4)	2	0.67	95%	No
9	10.9	6	Towards Automated Analysis of Business Processes for Financial Audits	11th International Conference on Wirtschaftsinformatik (WI)	(Werner et al., 2013)	26 %[11)]	C (6.73)	A	C	C	DB (5)	3	0.50	90%	No
10	10.10	5.5 6	Multilevel Process Mining for Financial Audits	IEEE Transactions on Services Computing	(Werner and Gehrke, 2015)	[12)]	-[13)]	A	-[13)]	-[13)]	B (3)[14)]	2	0.67	95%	No

[7] Invited for publication

[8] Submitted as full paper (14 pages), accepted as short paper (8 pages), presented at the ER 2013 and published in the conference proceedings, overall acceptance rate 32 %

[9] Acceptance rate for Completed Research Paper 28.92 %, Research-in-Progress 29.2 %

[10] Submitted and accepted as Research-in-Progress paper (12 pages), presented at the ICIS 2014 and published in the conference proceedings

[11] Acceptance rate for the relevant track 21.6 %

[12] The journal does not communicate acceptance rates for special or regular issues.

[13] Not explicitly listed in the VHB JQ2.1, ERA 2010 or CORE 2013 rankings. The ISI impact factor for this journal in 2014 was 3.049.

[14] Refers to the first review round, three more reviews were received during the second review round.

4 Analysis

The analysis phase is the first phase of the DSR cycle (Österle et al., 2010). Figure 7 shows a framework that provides guidance on how to carry out design science-oriented research (Hevner and Chatterjee, 2010; Hevner et al., 2004). It is useful to identify which aspects are important for the analysis phase. A key aspect is the analysis of the application domain for the identification and confinement of the research question. Taking requirements from the application domain into account ensures that the investigated question is indeed relevant for a specific practical purpose. A second aspect is the analysis and consideration of already existing knowledge from the scientific knowledge base which is important to ensure the research rigor by relying on knowledge that has already been evaluated and accepted in the scientific community. The following sub-chapters illustrate how knowledge from the knowledge base (chapter 4.1) and requirements from the application domain have been taken into account (chapter 4.2).

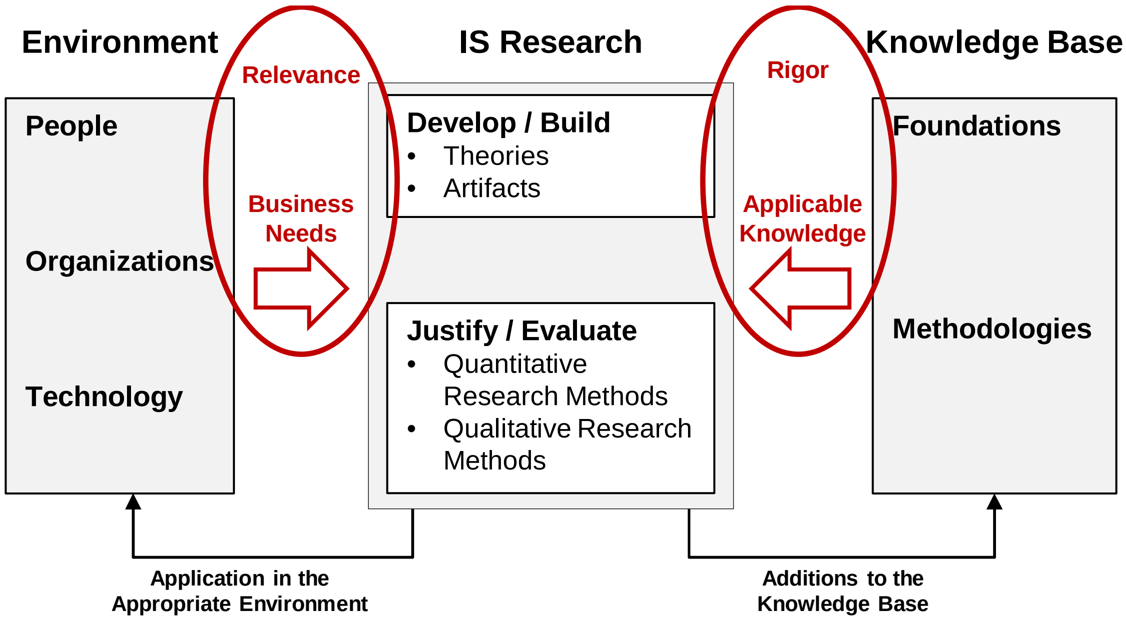


Figure 7 Integrating the Application Domain and Knowledge Base in DSR (adapted from Hevner et al., 2004, p. 80)

4.1 Related Scientific Work

4.1.1 Business Process Management

Business process management (BPM) is an important and mature research area that is relevant to the presented research.

> "Business process management includes concepts, methods, and techniques to support the design, administration, configuration, enactment, and analysis of business processes" (Weske, 2012, p. 5).
>
> It "is the discipline that combines knowledge from information technology and knowledge from management sciences and applies this to operational business processes" (van der Aalst, 2013, p. 1)

Several up-to-date textbooks provide fundamental and extensive knowledge on BPM (Dumas et al., 2013; Weske, 2012). Of particular interest are those publications related to the modeling of business processes using Petri Nets (van der Aalst and Stahl, 2011) because Petri Nets represent the most frequently used modeling language in the field of process mining (Tiwari et al., 2008). Van der Aalst provides a comprehensive survey on business process management and identifies 20 different use cases and six key concerns (van der Aalst, 2013). His survey relies on the analysis of 289 scientific publications published at the International Conference on Business Process Management (van der Aalst, 2012). The most relevant use case for the research at hand is the *'discover model from event data'* which strongly relates to the research domain of process mining.

4.1.2 Process Mining

Process mining is a relatively new research area that emerged in the late 1990s initiated by the work from Cook and Wolf about discovering software processes (Cook and Wolf, 1998a, 1998b, 1999). The discovery of process models was also already subject of research in various established scientific disciplines like concurrency theory, inductive inference, stochastic, data mining, machine learning or computational intelligences (van der Aalst, 2011a). Agrawal et al. and Maxeiner et al. were among the first who used workflow logs to create process models (Agrawal et al., 1998; Maxeiner et al., 2001). Schimm, Herbst and Karagiannis first used process mining in the context of workflow and business process management (Herbst, 2003, 2000a, 2000b; Herbst and Karagiannis, 1999, 1998; Schimm, 2001a, 2001b). Substantial and extensive research work on process mining has been published by researchers in the past decade. Van der Aalst provides a comprehensive summary of basic and advanced concepts in process mining that have been discovered in recent years (van der Aalst, 2011a). The Process Mining Manifesto is an important publication for the research in this area. It illustrates contemporary chances and challenges for process mining (van der Aalst et al., 2012). An overview of the state-of-the-art in process mining until 2008 is summarized in (Tiwari et al., 2008).

Figure 8 shows the distribution of scientific publications in the world until the end of 2013 that relate to process mining. The publications were identified by conducting an extensive literature review following the guidelines suggested by several scholars (Fettke, 2006; Rowley and Slack, 2004; vom Brocke et al., 2009; Webster and Watson, 2002). Scientific publications were identified by using the search term *'process mining'* for the following databases:

- ProQuest
- AISel
- EBSCO
- Springer Link
- ScienceDirect
- ACM

The search was restricted to the title and peer-reviewed publications when possible. The search also included a review of related articles in international top information systems journals (AIS Senior Scholars' Basket of Journals (AIS, 2014)):

- European Journal of Information Systems
- Information Systems Journal
- Information Systems Research
- Journal of the Association for Information Systems
- Journal of Information Technology
- Journal of Management Information Systems
- Journal of Strategic Information Systems
- Management Information Systems Quarterly

These journals were searched for hits in the title, abstract and keywords. The literature review finally covered 236 distinct articles[15]. The mapping to countries was achieved by investigating the heritage of the authors.[16] To prevent duplicate mappings only the first author of a publication was used to assign a publication to a specific country. Table 4 lists the publications per country. It can be seen that most publications were prepared by authors from the Netherlands followed by China and Germany.

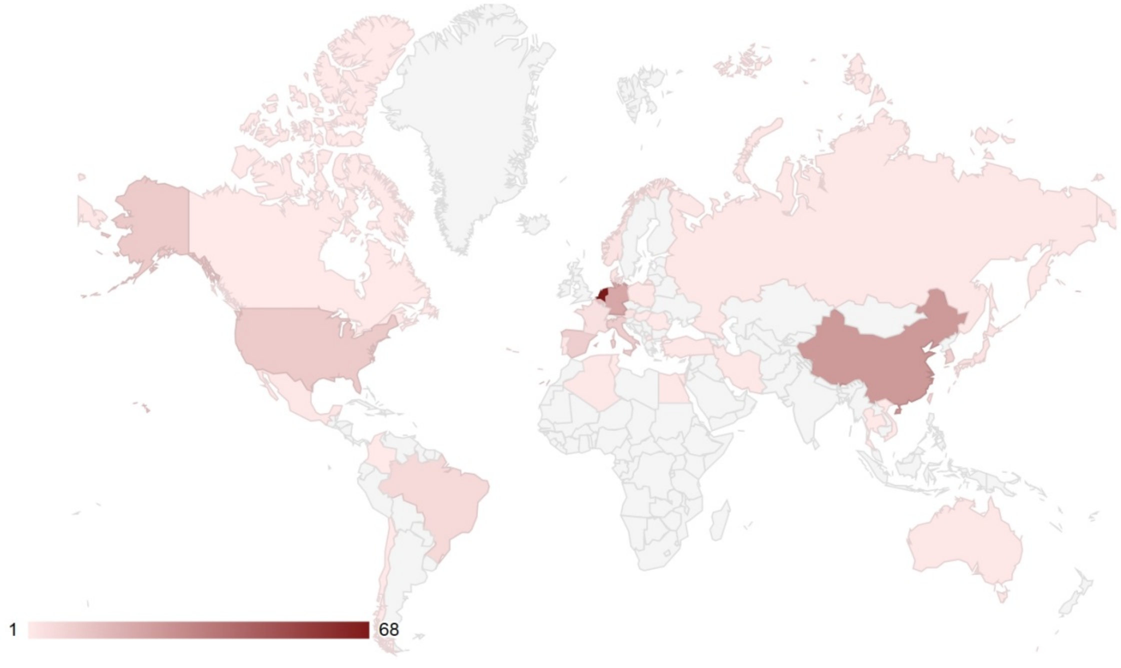


Figure 8 Distribution of Worldwide Process Mining Related Scientific Publications[17]

[15] The complete list of publications from this literature review is provided in appendix B.

[16] For the case that different information on the heritage of the author was found possibly because the author had switched his or her location in the academic career the most current information was used.

[17] The diagrams were prepared using Google Geocharts (Google, 2014)

Table 4 Distribution of Process Mining Related Scientific Publications in the World

Country	Publications
Netherlands	68
China	27
Germany	22
Belgium	14
USA	11
Spain	11
Italy	10
South Korea	9
Great Britain	8
Taiwan	7
Brazil	6
France	4
Japan	3
Norway	3
Austria	3
Egypt	2
Denmark	2
Russia	2

Country	Publications
Iran	2
Turkey	2
Australia	2
Algeria	2
Mexico	2
Poland	2
Portugal	2
Thailand	1
Canada	1
Colombia	1
Chile	1
Macedonia	1
Hungary	1
Tunisia	1
Vietnam	1
Rumania	1
Israel	1

Figure 9 and Table 5 show the distribution of process mining in Europe on the city level. It shows that Eindhoven can be considered the epicenter of research on process mining with 64 publications which represent 43 % of the publications in Europe and 27 % worldwide.

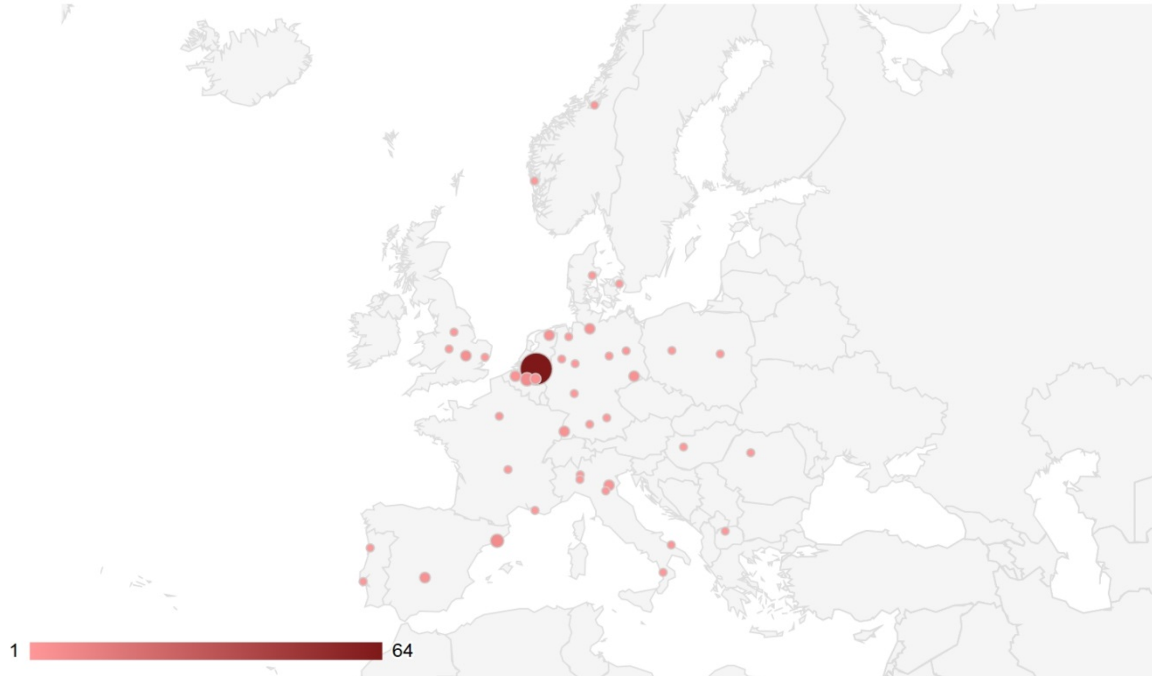


Figure 9 Distribution of Process Mining Related Scientific Publications in Europe

Table 5 Distribution of Process Mining Related Scientific Publications in Europe

Country	City	Publications
Belgium	Ghent	3
Belgium	Diepenbeek	4
Belgium	Leuven	7
Germany	Ingolstadt	1
Germany	Potsdam	2
Germany	Ulm	1
Germany	Paderborn	1
Germany	Dresden	4
Germany	Oldenburg	1
Germany	Magdeburg	1
Germany	Hamburg	4
Germany	Frankfurt	1
Germany	Münster	2
Germany	Freiburg	4
Denmark	Lyngby	1
Denmark	Aarhus	1
Great Britain	Birmingham	1
Great Britain	Sheffield	1
Great Britain	Bedford	5
Great Britain	Ipswich	1
France	Paris	2
France	Marseilles	1

Country	City	Publications
France	Clermont-Ferrand	1
Italy	Milan	1
Italy	Bologna	2
Italy	Rende	1
Italy	Bari	2
Italy	Pavia	1
Italy	Ferrara	3
Macedonia	Skopje	1
Netherlands	Eindhoven	64
Norway	Groningen	4
Norway	Trondheim	2
Norway	Bergen	1
Poland	Poznan	1
Poland	Warsaw	1
Portugal	Lisbon	1
Portugal	Santa Maria da Feira	1
Romania	Cluj-Napoca	1
Spain	Barcelona	7
Spain	Ciudad Real	4
Hungary	Veszprem	1

The identified articles represented the starting point for the scanning of the scientific knowledge base for the research at hand. They are just a subset of scientific articles that relate to the field of process mining. Many relevant articles do not include the search term *'process mining'* in the title and were therefore not identified in the initial review. The scientific knowledge base was extended during the iterative research cycles. Relevant related scientific work is referenced explicitly in the individual publications (compare appendix A). The following paragraphs provide an overview of the most important publications for the presented research. Fundamental aspects in regard of process mining have been published in (Gehrke and Werner, 2013). Process mining is used for process discovery, process enhancement, conformance and compliance checking (Gehrke and Werner, 2013; van der Aalst et al., 2012). Process discovery, conformance and compliance checking are especially relevant in the context of process audits.

Several powerful general purpose mining algorithms have been developed in recent years that use simple deterministic (van der Aalst et al., 2004), heuristic (Günther and van der Aalst, 2007; Weijters et al., 2006) or genetic (de Medeiros, 2006) approaches. An important aspect for the evaluation of the appropriateness of mining algorithms are quality criteria that can be used to assess mined process models (Rozinat, 2007; Rozinat et al., 2008). Several publications deal with process mining in the context of compliance checking. Compliance checking refers to the question if the observed process behavior complies with relevant rules. Ramezani et al. identify 55 control flow oriented compliance rules formalized in terms of Petri Net patterns that can be used to check mined models if they comply with these patterns (Ramezani et al., 2012). Van der Werf et al. scrutinize how information from the organizational contexts can be connected to recorded data in event logs to check compliance rules that are independent from individual process instances (van der Werf et al., 2012). Accorsi and Lehmann incorporate the data perspective to identify information leaks in business process models (Accorsi and Lehmann, 2012). Caron et al. suggest a rule-based compliance checking and risk management approach (Caron et al., 2013). Van der Aalst et al. introduce a conceptual model for online auditing using process mining techniques (van der Aalst et al., 2011) whereas Jans et al. discuss opportunities, challenges and limitations for using process mining in the context of audits (Jans, 2012; Jans et al., 2010). They also present several case studies (Jans et al., 2011, 2008). Conformance checking aims to identify deviant behavior in a process. It requires the existence of a model that is used for comparison. Rozinat and van der Aalst introduce an approach to compare mined process models to a reference model (Rozinat and van der Aalst, 2008). They use concepts like fitness and appropriateness to identify deviations. A similar approach is used by Adriansyah et al. who introduce a cost-based fitness analysis (Adriansyah et al., 2011). Van der Aalst and de Medeiros present a two-step approach (van der Aalst and de Medeiros, 2005). A reference model is mined in a first step and subsequent executions recorded in the event log are then used to identify any deviations compared to the previously mined model. Bezerra and Wainer analyze the event log to detect anomalies (Bezerra and Wainer, 2013). Yang and Hwan present a framework and case study to identify fraud in the healthcare sector (Yang and Hwang, 2006). Van der Aalst presents a case study using data from a Dutch governmental institution (van der Aalst, 2005).

Several academic and commercial process mining software tools exist. The most common are listed in Table 6. A major academic and open-source tool is ProM (Process Mining Group, 2015). It provides plug-ins for many different mining algorithms, as well as analysis, conversion and export modules. Disco (fluxicon, 2015) is a commercial application that benefits from intuitive and easy usability. It also provides integrated functionality for the filtering and loading of event logs. ProM and Disco are the software tools that were primarily used for evaluation and comparison purposes in the presented research.

Table 6 Process Mining Tools

Product Name	Link
ARIS Process Performance Manager	www.softwareag.com
celonis business intelligence	www.celonis.de
Disco	www.fluxicom.com
Genet/Petrify	www.lsi.upc.edu
Interstage Business Process Manager	www.fujitsu.com
QPR ProcessAnalysizer	www.gqr.com
ProM	www.processmining.org
ProcessGold	www.processgold.de
Rbminer/Dbminer	www.lsi.upc.edu
ReflectOne	www.pallas-athena.com
ServiceMosaic	soc.cse.unsw.edu.au

4.2 Requirements from the Application Domain

4.2.1 The Role of Business Processes in Financial Audits

The aim of the research that is described in this thesis is the development of analysis techniques that can be used as a foundation to automate the audit of business processes in the context of financial audits. A prerequisite to assess the requirements from the application domain that should be taken into account is an understanding of the role of business processes in financial audits. National and international laws and audit standards are an important information source for this purpose. ISA 315, IDW-PS 261 (IDW, 2009) and the AS No. 12 (PCAOB, 2010) require a risk based approach that takes business processes, internal controls and related information systems into account. A financial audit is commonly made up of different phases that are illustrated in Figure 10.

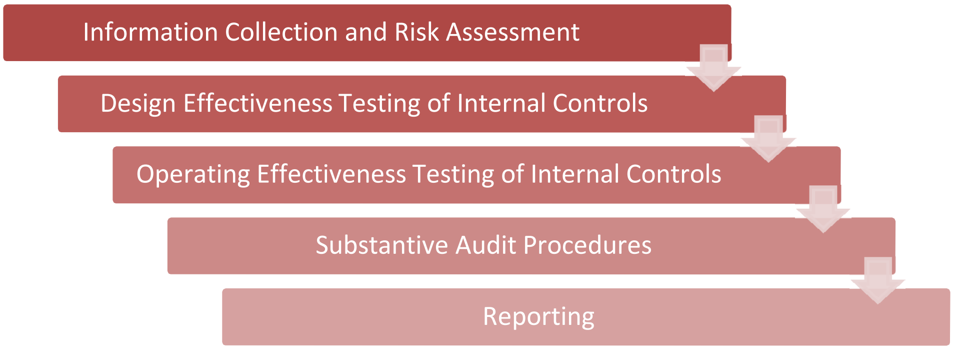


Figure 10 Audit Process (adapted from Werner and Gehrke, 2011, p. 105)

The process starts with the collection of necessary information on the audited entity and its business environment for the risk assessment, the identification of possible causes of risk and the determination of necessary materiality levels. It follows the testing of design effectiveness of internal controls. Internal controls are defined as

> "the process designed, implemented and maintained by those charged with governance, management and other personnel to provide reasonable assurance about the achievement of an entity's objectives with regard to reliability of financial reporting, effectiveness and efficiency of operations, and compliance with applicable laws and regulations." (IFAC, 2012, para. 4(c))

With regard to business process management and process modeling internal controls can commonly be considered as activities that influence the processing of business transactions. The objective of the test of the design effectiveness is to ensure that the internal controls are appropriately designed to achieve the intended control objectives. This assessment is only possible if the auditor has information about how the controls relate to the business transactions that actually create the postings on the financial accounts. The relationship between business processes and financial accounts is illustrated in Figure 11. The model represents a simple purchase process. The rectangles represent activities that are executed in an ERP system. They create postings on financial accounts. According to the double-entry bookkeeping system each posting consists of at least one credit and one debit entry that balance each other. Not every activity creates an entry in the financial accounts. If a company orders goods from a supplier this does not result in a posting, because the liability only occurs when the goods are received. The delivery of the ordered goods leads to entries in the raw materials and the goods received / invoices received account. When the invoice is received the goods received / invoices received is cleared and a corresponding entry is posted on the trade payables account. When the invoice is finally paid the entry on the trade payables account is cleared with a corresponding entry on the bank account. This example illustrates the relationship between business processes and financial accounts that are important from an audit perspective

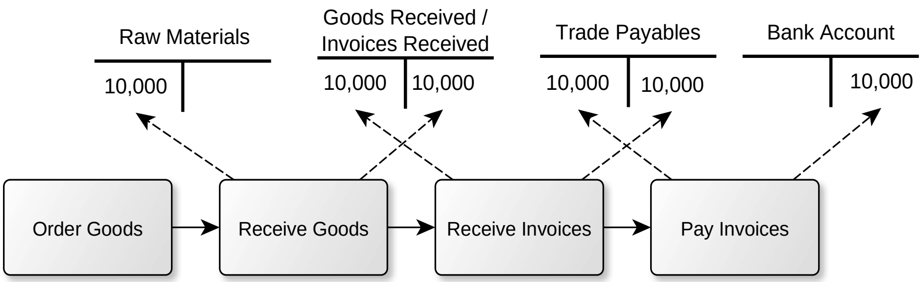


Figure 11 Example Model of a Purchase Process (Werner and Gehrke, 2015, p. 823)

The next step in a financial audit is the testing of the operating effectiveness of the identified controls to ensure that the controls were indeed effective in the audited period. The audit process continues with substantive testing procedures that include traditional analytical procedures, physical examinations of inventory or the assessment of confirmations of balances from suppliers. The audit process terminates with the final audit report.

Auditors use different types of audit procedures to test the design and operating effectiveness of internal controls. The primary audit procedures to test internal controls are inquiry, observation, inspection and re-performance (IFAC, 2012, para. A73). They differ in terms of audit effort and gained audit reliance as illustrated in Figure 12.

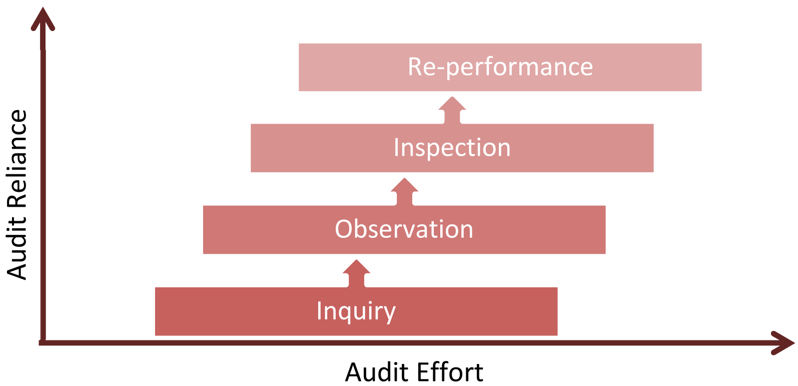


Figure 12 Audit Procedures (adapted from Werner and Gehrke, 2011, p. 105)

The audit procedures illustrated in Figure 12 are all manual procedures. They are time-consuming and error-prone and may become inefficient or even ineffective if information systems are used on the company's side to automate the operation of business processes and if very large amounts of data are processed (Werner and Gehrke, 2011). A key prerequisite for the effectiveness of inquiries is the assumption that the interview partner has sufficient information about relevant business processes and related internal controls and that the auditor is able to receive all necessary information in an interview. This might not be the case anymore if process activities are partially or completely processed by an information system without any human interaction. Observations, inspections and re-performances are carried out on the basis of sampling. With an increase of processed transactions the sample size has to be increased. When millions or even billions of transactions are processed during an audit period it can be assumed that they become very inefficient when the sample size is increased accordingly or that they even become ineffective if an appropriate sample size cannot be selected because of limited audit resources.

4.2.2 Data Structure

Manual audit procedures can be substituted by automatic audit procedures on the basis of process mining techniques if the data that is stored in information systems during the course of processing is used. The original source data can commonly not be used directly for process mining purposes because it does not exhibit the data structure that is required by established process mining algorithms (van der Aalst et al., 2012) .

Mining algorithms use event logs as input that exhibit a specific structure (Günther and Verbeek, 2012). An event log is basically a table which contains all recorded events that relate to executed business activities. Each event is mapped to a case that corresponds to a single execution of a business process which is called process instance. The sequence of recorded events in a case is called a trace. Cases and events are characterized by classifiers and attributes. Classifiers ensure the distinctness of cases and events by mapping unique names to each case and event. Attributes store additional information that can be used for analysis purposes (Gehrke and Werner, 2013). An example of an event log is given in Table 7.

Table 7 Event Log Structure (Gehrke and Werner, 2013, p. 935)

Case ID	Event ID	Timestamp	Activity
1	1000	01.01.2013	Order Goods
	1001	10.01.2013	Receive Goods
	1002	13.01.2013	Receive Invoice
	1003	20.01.2013	Pay Invoice
2	1004	02.01.2013	Order Goods
	1005	01.01.2013	Receive Goods
	…	…	…

The data entries that are recorded in ERP systems differ from the structure illustrated in Table 7. The necessary data is commonly stored in various database tables. This data exhibits a specific structure that relates to the systematic of double-entry bookkeeping. The Entity-Relationship-Model (ERM) in Figure 13 illustrates the relationship between transactions that are executed in ERP systems, data entries and financial accounts. The illustrated entity attributes follow the naming of data labels used in SAP ERP systems. The execution of a transaction that is labeled with a transaction code creates one or more posting documents.[18] These documents contain two or more journal entry items that are posted to a specific account. If the account is enabled for open-item-accounting each open item has to

[18] If the execution is not financially relevant no posting document is created.

be cleared by a clearing posting. If this is the case a cleared item carries a reference to a posting document that creates the clearing items.

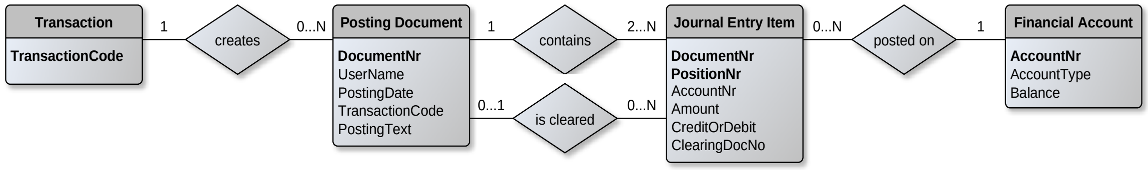


Figure 13 ERM for Accounting Data Structure (Werner, 2013, p. 389)

This data structure is different compared to the one presented in Table 7. Events in a traditional event log exhibit a strict linear order within each case. No concurrency is allowed on the process instance level. This strict order is a fundamental prerequisite for traditional process mining algorithms to infer the control flow. It is not given for financially relevant data as illustrated in Figure 13. Due to the 1 to 2…N cardinality of *'the contains'* relationship and the existence of the *'is cleared'* relationship with the 0…N to 0…1 cardinality the resulting data structure is not linear. The source data is also not labeled which means that events are not mapped to cases. Gehrke and Müller-Wickop introduce an algorithm that maps events to cases according to the data structure shown in Figure 13. Figure 14 shows an example of the source data structure of a process instance extracted from an SAP system using the approach and the notation introduced by (Gehrke and Müller-Wickop, 2010). A rounded rectangle represents an activity. A simple rectangle represents a cleared journal entry item that is involved in open-item-accounting. Hexagons represent those items that are not involved in open-item-accounting. Gray rectangles and hexagons represent debit and white ones represent credit postings. Each rectangle and hexagon displays values for the journal entry item number, the account it was posted on and the posted amount. The activity symbols display the transaction code that was used to create the posting documents. Activities and related journal entries belonging to the same document numbers are clustered in groups. The example shows that the recorded events which are represented by the activity symbols do not exhibit a strict linear order but show a concurrent behavior on the process instance level. It also becomes obvious that no direct information on the causal direction of activities can be derived from the presented model.

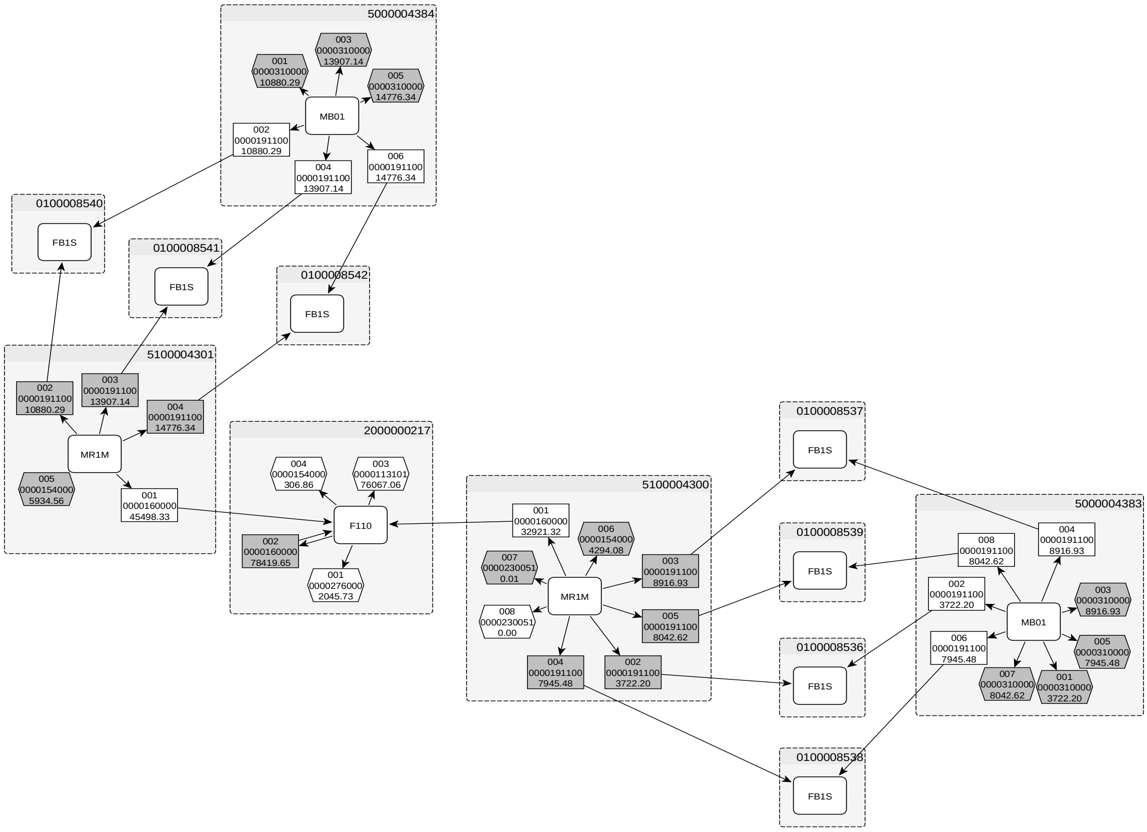


Figure 14 Source Data Structure for a Purchase Process Instance (adapted from Gehrke and Müller-Wickop, 2010, p. 7)

4.2.3 Requirements Summary

The previous two chapters illustrated the role of business processes in financial audits and the data structure of the source data that can be used for process mining purposes. A process mining algorithm that is used to discover and analyze process models in financial audits should meet the different criteria that stem from the application domain and the used ERP systems. Müller-Wickop and Schultz et al. conducted empirical investigations using expert interviews and surveys to identify key concepts and information requirements for process audits (Müller-Wickop and Schultz, 2013a; Schultz et al., 2012). They show that the process flow is a central concept and highly relevant from an audit perspective in practice. Two further important concepts for external auditors are financial statements and materiality[19]. Only those business transactions are inspected in a financial audit that can have a material effect on the financial statements. It is therefore necessary to receive information on the value flow that is created by the audited business process to decide if it needs to be audited from a materiality perspective or if it can be neglected. These requirements correspond to the logical considerations on the basis of the review of relevant audit standards in chapter 4.2.1. They can be summarized as follows:

[19] Materiality is defined in ISA 320 as followed: "Misstatements, including omissions, are considered to be material if they, individually or in the aggregate, could reasonably be expected to influence the economic decisions of users taken on the basis of the financial statements." (IFAC, 2009, para. 2).

Requirement I: The mined process models should provide information on the control flow of process activities.

Requirement II: The mined process models should provide information on the relationship of business processes and financial accounts.

Requirements I and II can be satisfied by integrating the control flow and the data flow perspective.

Another critical requirement in financial audits is the preservation of the audit trail. The audit trail is a fundamental concept in financial accounting. It is a path in an information system that allows tracing a transaction from the point of origin to the final output. It is used to verify the accuracy and validity of journal entries (Romney and Steinbart, 2008). Translating this requirement into the context of process mining implies that a mining algorithm may not alter the original data during the mining process.

Requirement III: The mining algorithm should not alter the original source data for creating process models.

If process mining is used in financial audits the generated models are used to discover incompliant behavior. The provided process models should therefore only represent information on process behavior that was actually observed in the source data. A fundamental challenge for process mining is the balancing between competing quality criteria (van der Aalst et al., 2012). Process mining is generally used to reduce complexity by visual representation and abstraction. Abstraction means that certain details are not represented in the created process models. This can lead to the phenomenon of miss-fitting process models. Models can be under- or over-fitting. A model is under-fitting if it allows execution paths in the process model that are not represented in the event log and over-fitting if they do not allow for any additional behavior that is not included in the event log. Rozinat et al. identify four quality criteria for the evaluation of mining algorithms: fitness, precision, generalization and structure (Rozinat et al., 2008). Fitness indicates if a model is able to represent all cases in the event log. Precision is the complementary criterion. It indicates if a process model does not allow additional behavior that was not observed in the log. Generalization addresses the capability of a model to express more behavior than recorded in the log. Structure refers to the graphical representation of a business process and depends on the graphical components of the target language. Other scholars use the closely related criterion of simplicity instead of structure (van der Aalst et al., 2012). Mined process models in financial audits should be as precise and fitting as possible. If the produced process models are over-fitting certain behavior recorded in the event log is not represented in the process model. The auditor would therefore assume that no incompliant behavior has occurred. But in reality it is just not represented in the miss-fitting process model which would eventually lead to false positive audit results. If the process model is too general, process behavior is illustrated that actually did not occur. This would lead to false negative audit results and unnecessary investigations by the auditor (Werner and Gehrke, 2015).

Requirement IV: The mined process models should be as precise as possible.

Requirement V: The mined process models should be as fitting as possible.

A suitable mining algorithm must also be able to present the detailed source data to the auditor for investigation purposes. Detailed information is necessary to inspect individual data values to identify or confirm compliance violations. But on the other hand the mining algorithm should also be able to present information at an adequate abstraction level to provide an overview of the control and data flow. It should therefore provide information on the investigated business process on different abstraction levels.

Requirement VI: The mining algorithm should be able to produce process models at different abstraction levels.

Additional requirements originate from the structure of the available source data as discussed in chapter 4.2.2. The source data from ERP systems is commonly not labeled and the activities on process instance level may exhibit concurrent behavior.

Requirement VII: The mining algorithm should be able to use unlabeled event logs as input.

Requirement VIII: The mining algorithm should be able to use event logs with non-linear traces as input.

5 Design

The second phase in the DSR cycle is the design of research artifacts such as constructs, models, methods, instantiations (March and Smith, 1995) and theories (Gregor, 2006). The design phase lies at the heart of design science-oriented research as illustrated in Figure 15. The aim is to use the results from the analysis phase concerning the requirements and business needs on the one hand and the scientific knowledge from the knowledge base on the other hand to develop novel solutions. They should be useful for the application domain and should also contribute to the general body of knowledge as shown with the feedback loops in the lower part of Figure 15. Chapter 3.3 has illustrated how the different artifacts relate to each other and that they have been developed in iterative cycles. The used design methods have been described in chapter 3.4. The following sub-chapters describe the different research artifacts as the main research output of the presented research.

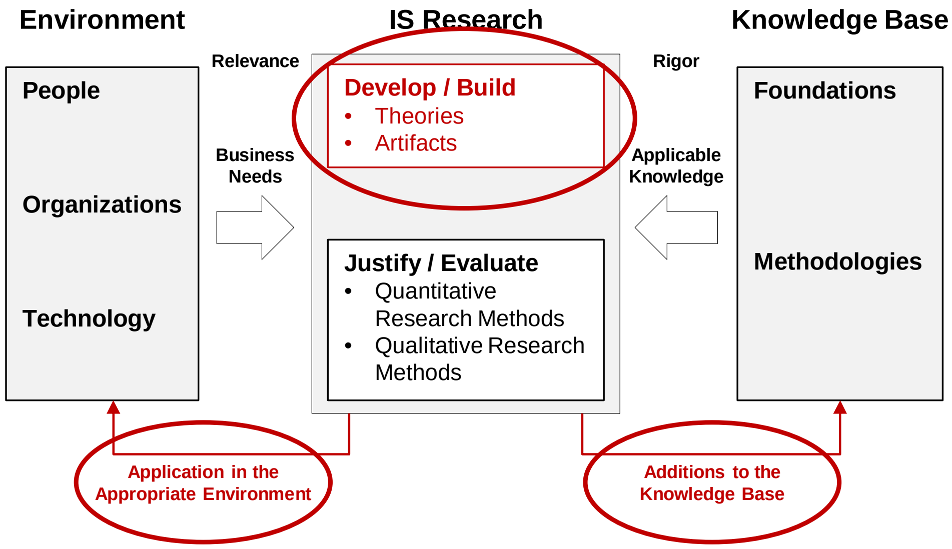


Figure 15 Design in the Information Systems Research Framework (adapted from Hevner et al., 2004, p. 80)

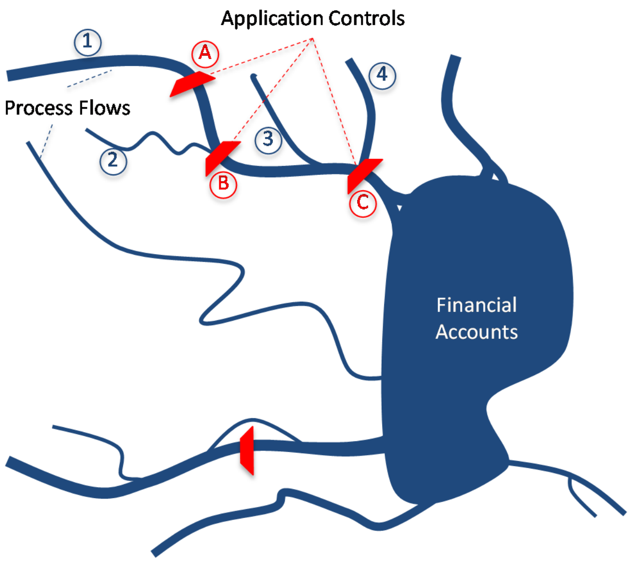
5.1 Conceptual Models for Specifying the Problem Domain and Solution

The key problem that motivates this research is the imbalance between automated transaction processing on the companies' side and manual audit procedures on the auditors' side leading to inefficient or even ineffective process audits (compare chapter 2.3). The idea to reduce this imbalance is the development and application of automated data analysis techniques such as process mining. But how should this work from a conceptual perspective?

We use a metaphor to describe the problem domain and possible solution. Business processes are a set of related activities to achieve a business goal (Weske, 2012). A business process can be interpreted as a flow or river. The execution of activities creates data entries that flow through the involved information systems. Financially relevant information finally ends up as journal entries on the financial accounts. This interpretation is conceptually illustrated in Figure 16. It shows different process flows that represent different business processes. These processes create data which is the water in our metaphor that feeds the financial accounts represented as the lake and final destination.

The figure also shows several application controls. Application controls are internal controls that are implemented in the information systems. They support or automate the processing of business transactions and regulate the transaction processing or respectively the flow of the water in our metaphor.

We illustrate the relationship between the process activities and application controls with an example. A typical purchase process starts with a purchase order (activity 1) that creates an order document. Purchase orders can usually only be issued up to a certain amount without further approval (control A). The ordered goods are eventually delivered (activity 2). A two-way-match (control B) ensures that the received goods can only be entered in the ERP system if the ordered goods match with the purchase order (Chuprunov, 2012). Finally the invoice for the delivered goods arrive (activity 4). A three-way-match ensures that the billed goods match with the purchase order and the delivered goods (control C) (Chuprunov, 2012). Activity 3 shows that not all application controls relate to all possible process flows or sub-processes. The two-way-match is usually not used for delivered services because they are not physically delivered and no corresponding delivery document exists.



Process Activities	
#	Description
1	Purchase order
2	Goods receipt for purchase order
3	Services receipt
4	Incoming invoice

Application Controls	
#	Description
A	System based approval for purchase order
B	Two-way-match
C	Three-way-match

Figure 16 Metaphor Illustrating the Relationship between Business Processes, Financial Accounts and Application Controls (adapted from Werner et al., 2012a, p. 5356)

A public accountant has to provide an opinion during a financial audit about whether the financial statements present a fair and true view of the financial situation of the company. If we rely on the presented metaphor this would mean that the auditor has to ensure that the lake only contains water from specific sources with a defined quality. Following contemporary audit approaches (compare chapter 2.2) the auditor would manually take samples from different places in the lake to verify the water quality and origin. This would be the equivalent of manual audit procedures like the inspection of selected documents or interviewing individuals. Based on experience and professional judgment he would maybe also choose several rivers for inspection and verify manually if control mechanisms are in place which regulate the flow and quality of the water that would be the equivalent of manual internal controls tests. An auditor has to rely on the information received from interviews and in best cases from available process descriptions but without actually knowing precisely which rivers and concurrent flows indeed exist and how much water they carry into the lake.

One of the most important criteria for an auditor is to know the relationship between the business processes and financial accounts (compare chapter 4.2.3.). The first step to support the auditor is the provision of data analysis techniques that provide information on the interaction of processes and accounts. Using process mining techniques enables the auditor to get a visual representation of relevant business processes. This is the foundation for the auditor to receive an all-embracing understanding of the relevant business processes and how they affect the financial accounts. When this information is available application controls can be taken into account in a second step. Application controls can be viewed as filters from an audit perspective. If the design and operating effectiveness of all necessary application controls for a specific business process are verified such a process does not need any further attention because the controls ensure that only valid and correct financial entries end up in the financial accounts from the controlled business processes. Audit resources can then be spent on those processes that are not controlled by application controls and that usually exhibit a higher audit risk than standard processes that are well-controlled.

The fundamental idea for automating process audits is the combination of process mining and automated application controls testing techniques. This is conceptually illustrated in Figure 17. The automation of the process discovery and analysis is the aim of the research presented in this thesis (compare chapter 3.2.2). It is the first step and a fundamental prerequisite for developing automated audit procedures.

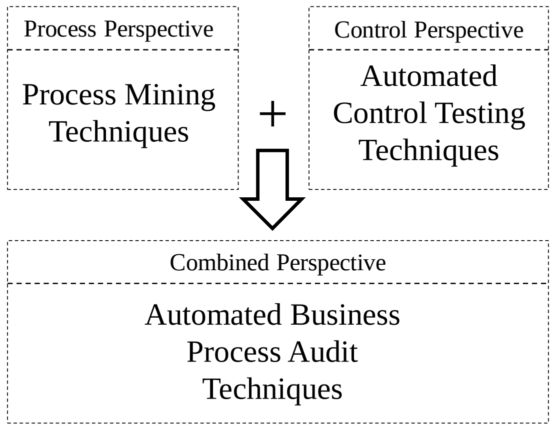


Figure 17 Conceptual Solution

5.2 CPN Specification for Integrating the Data Perspective

The expressive power of a mining algorithm partially depends on the used modeling language. The majority of process mining algorithms use Petri Nets for modeling mined process models (Tiwari et al., 2008). The usage of Petri nets allows the application of mature mining and analysis methods that are already implemented in academic software tools such as ProM (Process Mining Group, 2015). Petri nets are suitable for the modeling of business processes and offer a formal but also graphical notation which is also comprehensible to non-experts (van der Aalst and Stahl, 2011). They provide a sound mathematical foundation which allows the simulation and verification of Petri net models (Jensen and Kristensen, 2009).

It is possible to satisfy the requirement I from chapter 4.2.3 that postulates that the mined models should provide information on the control flow of process activities by modeling mined process models as Petri Nets. Petri Nets support all common control flow patterns (van der Aalst et al., 2003) and are therefore suitable to model the control flows in business processes.

Requirement II from chapter 4.2.3 postulates that the mined process models should provide information on the relationship of business processes and financial accounts which can be achieved by integrating the data perspective in process models. The data perspective has generally not been considered extensively yet for process mining (de Leoni and van der Aalst, 2013; Stocker, 2012). The majority of mining algorithm use low-level Petri Nets. Data objects can be included by using Colored Petri Nets (CPN). Such objects can be modeled as colored tokens and places (Werner, 2013).

A Colored Petri Net can formally be expressed as a tuple CPN = $(T, P, A, \Sigma, V, C, G, E, I)$ (Jensen and Kristensen, 2009). Table 8 provides a specification for integrating the data perspective using CPN for process mining in financial audits.[20]

Table 8 CPN Specification for Process Mining in Financial Audits (Werner, 2013, p. 390)

***T* is a finite set of transitions**	
Transactioncode Activity Name	The transitions represent the activities that were executed in the process. They display the name of the activity. Further information, for example the transaction code name, can be added.
***P* is a finite set of places**	
Places in the FPN represent financial accounts and control places.	
Control Places Source Place, Sequence Place, Sink Place	Control places determine the control flow in a process model. For every process model one source place is modeled that connects to the start transactions. The sink place marks the termination of the process. The control places between the start and end transition determine the execution sequence of the process model. A control place belongs to the set of control places CP.
Account Places Account: debit side of a balance sheet account Account: credit side of balance sheet account Account: credit side of a profit and loss account Account: debit side of a profit and loss account	The account places represent financial accounts that are affected by the execution of activities in a process. The symbol color indicates the meaning of an account. Account places belong to the set of account places AP.
$A \in P \times T \cup T \times P$ is a set of arcs also called flow relation	
⟶ Control Arc	Control arcs connect control places with transitions. They model the control flow in the model.
- - - -> Posting Arc	Posting arcs illustrate the relationship between activities represented as transitions in the model and financial accounts that are modeled as account places.
<- - - -> Clearing Arc	Clearing arcs are used to model that an activity cleared an entry on the corresponding account. Clearing arcs are double-headed arcs and are used as an syntactical abbreviation for two arcs (*p,t*) and (*t,p*).

[20] CPN following this specification are called Financial Petri Nets (FPN) in the related publication.

Σ is a set of non-empty color sets

colset *VALUES* = double	The color set contains the possible values that are posted or cleared.
colset $ACCOUNTS$ = string	The color set contains all account numbers.
colset $ACOUNTTYPE$ = boolean	The color set contains {1,0} indicating if the represented account is a balance sheet or a profit and loss account.
colset $CREDorDEB$ = boolean	The color contains {1,0} indicating if the account place is a representation of the debit or credit side of an account.
colset $EXECUTIONS$ = int	The color contains {1,…,n} indicating how often a path was chosen in the FPN.
colset $ACCOUNTPLACES$	$ACCOUNTPLACES$ is the color set as a product of $VALUES * ACCOUNTS * ACCOUNTTYPE * CREDorDEB$

$\boldsymbol{V}$ is a finite set of typed variables such that Type$[\boldsymbol{v}] \in$ Σ for all variables $\boldsymbol{v} \in \boldsymbol{V}$.

In FPN models arc inscriptions are modeled as constants. Variables are therefore not necessary and $V = \{\}$.

$\boldsymbol{C}: \boldsymbol{P} \rightarrow \boldsymbol{\Sigma}$ is a color set function that assigns a color set to each place.

The color set function in FPN assigns different color sets to places depending if they belong to the group of control or account places:

$$C(p)\begin{cases} ACCOUNTPLACES\ if\ p \in AP \\ EXECUTIONS\ if\ p\ \in CP \end{cases}$$

$\boldsymbol{G}: \boldsymbol{T} \rightarrow \boldsymbol{EXPR_V}$ is a guard function that assigns a guard to each transition $\boldsymbol{t}$ such that $\boldsymbol{Type}\ [\boldsymbol{G(t)}] = \boldsymbol{boolean}$.

$EXPR$ is the set of expressions that is provided by the used inscription language. FPN do not explicitly include guards because they do not model dynamic behavior of transitions that depends on specific input but illustrate the processing of already executed processes. The guard function for FPN is therefore defined as $G(t) = true\ for\ all\ t\ \in T$.

$\boldsymbol{E}: \boldsymbol{A} \rightarrow \boldsymbol{EXPR_V}$ is an arc expression function that assigns an arc expression to each arc $\boldsymbol{a}$ such that $\boldsymbol{Type}\ [\boldsymbol{E(a)}] = \boldsymbol{C(p)_{MS}}$, where p is the place connected to the arc $\boldsymbol{a}$.

The arc expressions in a FPN are constants. The arc expression function assigns to each posting and clearing arc a set of constants that denote the posted or cleared value, the account number, account type and an indicator if it is a credit or debit posting. For each control flow arc the number of execution times is assigned indicating how often this path was chosen in the process model.

$$E(a)\begin{cases} \{val \in VALUES, acc \in ACCOUNTS, acctyp \in ACCOUNTTYPE, cord \in CREDorDEB\} \\ \quad with\ Type\ [E(a)]\ = C(p)_{MS}\ \ = ACCOUNTPLACES\ if\ p \in AP \\ \\ \{ex \in EXECUTIONS\ with\ Type\ [E(a)] =\ C(p)_{MS}\ \ = EXECUTIONS\ if\ p\ \in CP \end{cases}$$

$\boldsymbol{I}: \boldsymbol{P} \rightarrow \boldsymbol{EXPR_{\emptyset}}$ is an initialization function that assigns an initialization expression to each place $\boldsymbol{p}$ such that $\boldsymbol{Type}\ [\boldsymbol{I(p)}] = \boldsymbol{C(p)_{MS}}$

The initialization function of a FPN assigns initialization expressions to each place as follows:

$$I(p)\begin{cases} n'ex \in EXECUTIONS\ with\ Type\ [I(p)] =\ C(p)_{MS}\ \ = EXECUTIONS\ if\ p = source \\ \\ \emptyset_{MS}\ \ otherwise \end{cases}$$

Only the source place is initialized in a FPN. The initialization expression for $p = source$ generates n tokens in the initial marking $M_0(p)$, one for each connected start transition. The inscription of each token is a member of the set $EXECUTIONS$.

Figure 18 shows an example of a simple purchase process that can be modeled by using the specification illustrated in Table 8.

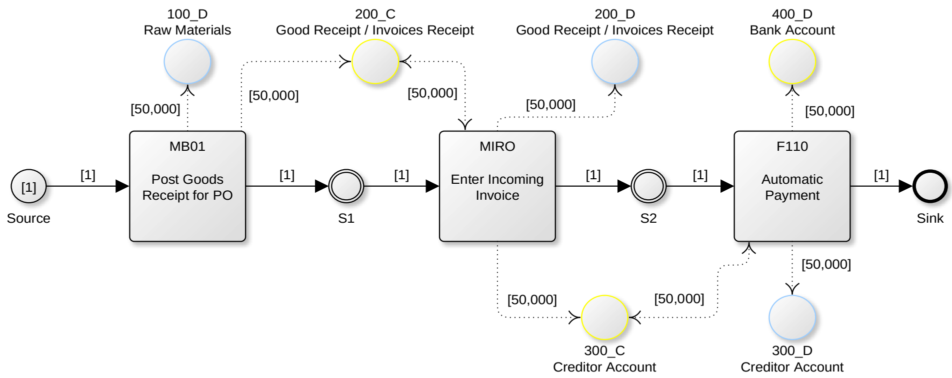


Figure 18 Simple Example of a Purchase Process Using CPN (Werner, 2013, p. 392)

5.3 Complexity Reduction Algorithm

A common challenge in process mining is the complexity of mined process models. Extremely complex models are called spaghetti and lasagna models due to their graphical appearance (van der Aalst, 2011c). Source data derived from ERP systems related to financially relevant transactions is complex (Werner et al., 2013, 2012b). Figure 19 shows a typical process instance. It consists of 12 transitions and 115 net elements[21]. The average process instance in the used data sets from three different companies consisted of 7 to 9 net elements (Werner et al., 2013). But the vast majority of the mined models just represented trivial instances that consist of one transition. The instance in Figure 19 can therefore be perceived as a typical process instance for a purchase process. Figure 20 and Figure 21 show two additional examples of process instance models that were derived from data provided by a company operating in the manufacturing industry. The model in Figure 19 can still be interpreted by simple observation. But this is not the case anymore for the models from Figure 20 and Figure 21. These represent instances that consist of 358 (example B) and 2,966 (example C)[22] transitions. Table 9 shows some characteristics of the three example instances.

[21] The sum of transitions, places and arcs in the CPN.

[22] We called them monster instances due to their large size and necessary computation time for mining, and the fact that they do not show the same graphical characteristics as spaghetti or lasagna processes but exhibit a more organic shape.

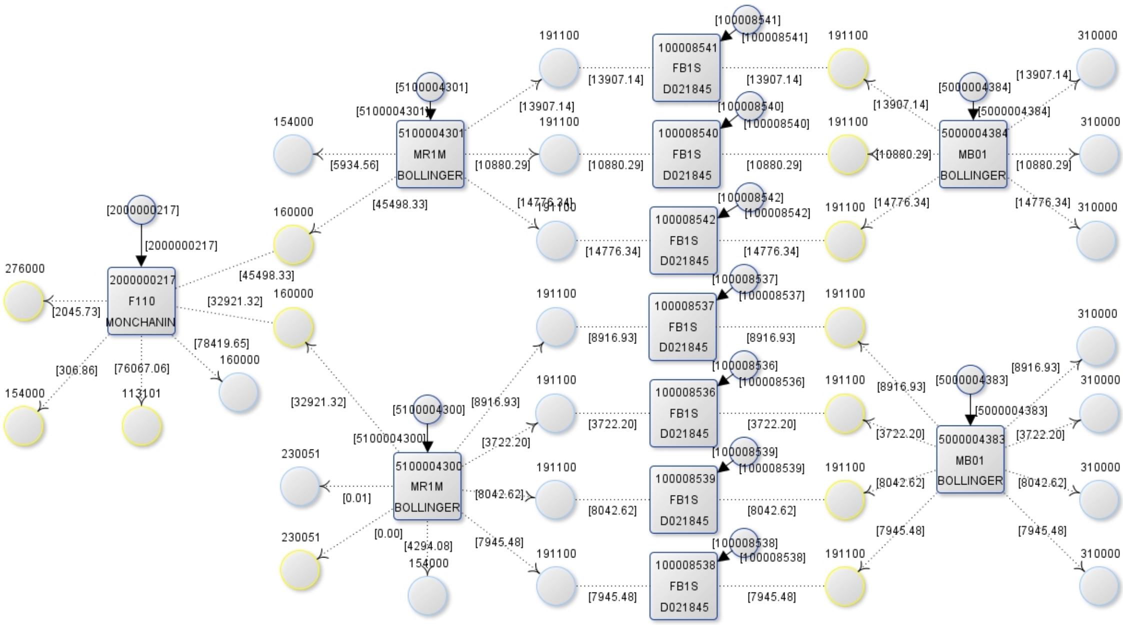


Figure 19 Example A: Normal Process Instance (Werner et al., 2012b, p. 4)

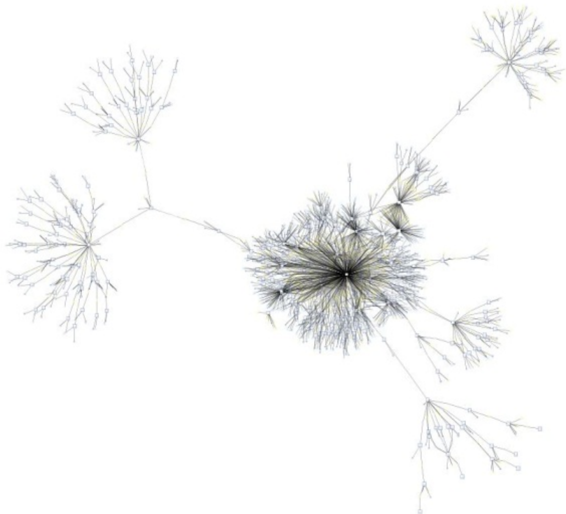


Figure 20 Example B: Large Process Instance (Werner et al., 2012b, p. 5)

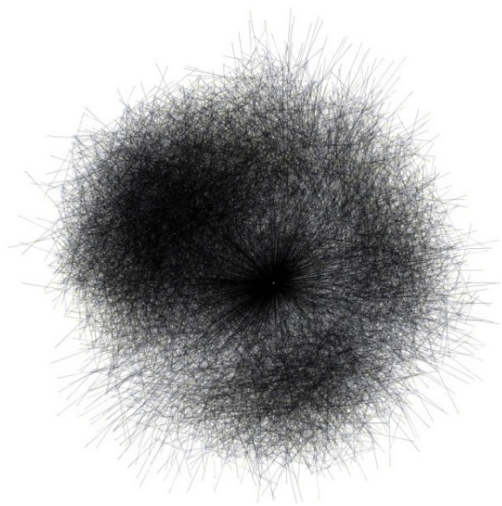


Figure 21 Example C: Monster Process Instance (Werner, 2012, p. 211)[22]

Table 9 Net Characteristics for Process Instance Examples

Example Instance	A	B	C
Number of transitions	12	358	2,966
Number of places	43	1,804	8,936
Number of arcs	60	2,549	14,867
Sum of net elements	115	4,711	29,735

The complexity of mined models can be reduced by using graph transformation techniques (Werner et al., 2012b). The complexity of a graph can be defined in many different ways depending on the relevant perspective und purpose (Neel and Orrison, 2006). A graph is generally defined as a pair of disjoint sets G = (V,E) .V is the set of vertices and E is the set of edges with E ⊆ V2 (Diestel, 2010). We consider the complexity of the graph simply as a function of its number of edges and vertices which is the same as the number of net elements as shown for example in Table 9.

The mined models consist of transitions that represent the process activities and account places (compare chapter 5.2). The account places represent the financial accounts. An account place is modeled for each journal entry item that was created by an activity. Although the complexity of mined models can be extremely high it is striking that the number of different accounts in such models is comparatively small. The mean value of different accounts per instance ranged from 2.33 to 2.95 (Werner et al., 2013). To reduce the complexity of mined models it would therefore be a promising approach to aggregate account places that carry the same account number using graph transformation techniques (Heckel, 2006; Rozenberg, 1997).

An important requirement described in chapter 4.2.3 is the preservation of the audit trail and the necessity to keep the source data unchanged during the mining process (requirement III). This requirement can be articulated in more detail in regard to the intended graph transformation (Werner et al., 2012b):

Requirement III a): The set of firing sequences has to stay constant.

Requirement III b): Different arc types may not be merged.

Requirement III c): The arc inscriptions representing the value of posted journal entries have to be preserved.

The aggregation algorithm shown in Listing 1 uses CPN with tuple N = (P, T, F, C, cd, W, m_0)[23] as input and aggregates places in accordance with requirements III a) to c).

Listing 1 Aggregation Algorithm (Werner et al., 2012b, p. 7)

```
P_Item ⊆ P      set of all places representing journal entry items in the net
P_ItemAgg= ∅    initially empty set for aggregated places
F_Arcs          set of all arcs in the net
Aggregate Places
While P_Item≠ ∅
      Take p_i ∈ P_Item
      Select all p_j ∈ P_Item with cd(p_i)=cd(p_j)
      Merge arcs for each p_i and p_j
      Add p_i to P_ItemAgg
      Remove p_i and p_j from P_Item
Set P_Item=P_ItemAgg
Merge Arcs
Get incoming arcs F_IncArcsI ⊆ F_Arcs for p_i
Get incoming arcs F_IncArcsJ ⊆ F_Arcs for p_j
For each arc a_i(t_m,p_i) ∈ F_IncArcsI and arc a_j(t_n,p_j) ∈ F_IncArcsJ
      If the arc type of a_i = arc type of a_j
            And if t_m = t_n then add W(a_j) to W(a_i)and  /Case 1
            remove a_j from F_Arcs
      Else set a_j(t_m,p_j)                            /Case 2,3 and 4
```

[23] The used CPN tuple elements refer to the tuple elements specified in chapter 5.2 where F is equivalent to A, C to Σ, cd to C, W to E and m_0 to $M_0(p)$.

```
Get outgoing arcs F_OutArcsI ⊆ F_Arcs for p_i
Get outgoing arcs F_OutArcsJ ⊆ F_Arcs for p_j
For each arc a_i(p_i,t_m) ∈ F_OutArcsI and arc a_j(p_j,t_n) ∈ F_OutArcsJ
      If the arc type of a_i = arc type of a_j
            And if t_m = t_n then add W(a_j) to W(a_i)and  /Case 1
            remove a_j from F_Arcs
      Else set a_j(p_j,t_m)                              /Case 2,3 and 4
```

Figure 22 shows the result when the algorithm is used to reduce the complexity of example A (Figure 19). The average of the number of net elements per instance could be reduced by 23.2 % and 23.4 % for two different data sets originating from a company operating in the retail business and a test SAP system (Werner et al., 2012b).

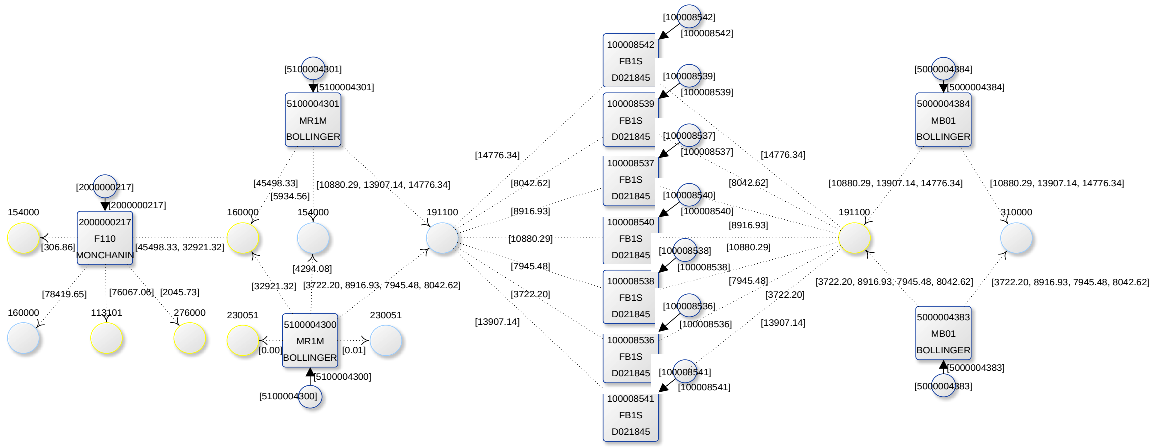


Figure 22 Aggregated Process Instance A (Werner et al., 2012b, p. 8)

5.4 Data-dependent Sequencing Algorithm

The source data in ERP systems that can be used to mine business processes that relate to financial accounts is unlabeled and recorded events do not exhibit a strict linear order on the process instance level (compare requirements VII and VIII from chapter 4.2.3). Gehrke and Müller-Wickop present an algorithm that can be used to map events to cases satisfying requirement VII (Gehrke and Müller-Wickop, 2010).

Traditional mining algorithms use the temporal ordering of events to infer the control flow. This is not possible if the events are not strictly ordered. Instead of relying on the temporal ordering it is possible to derive the control flow by analyzing the data dependency between events in a case. The control flow can be determined by analyzing which activity cleared items that were created by other activities. This is illustrated in Figure 23.

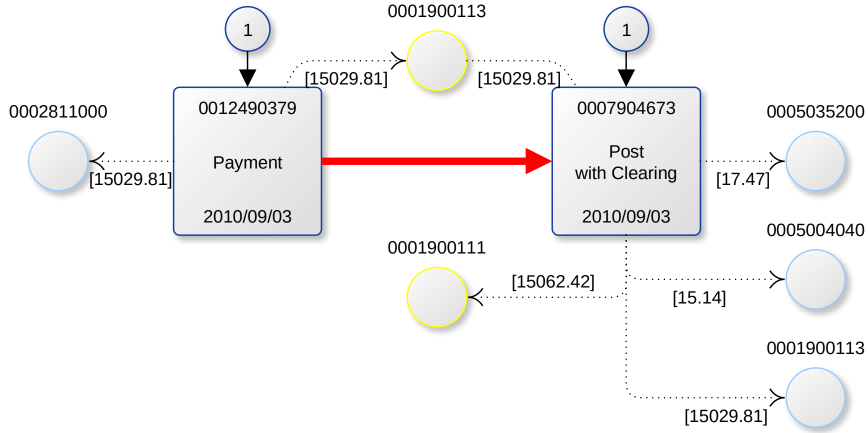


Figure 23 Logical Dependency (Werner and Nüttgens, 2014, p. 3892)

The activity *Payment* in Figure 23 posted an item with the value of *15,029.81* on the account *0001900113*. This item was cleared by the activity *Post with Clearing*. The logical sequence for these activities can therefore be derived as *Payment* → *Post with Clearing*. This approach can be used for all activities in a process instance to infer the complete control flow. Start nodes can be determined by identifying those activities that do not have any incoming control arcs. End nodes do not have any outgoing control arcs (Werner and Nüttgens, 2014).

A problem with this approach is the occurrence of clearing deadlocks. This constellation is shown in Figure 24. The shown *Clear Postings* activity did not create any other posted item and would therefore be defined as an end node. This is a problem because the clearing activity is actually not an end node but just an intermediary node in the process model.

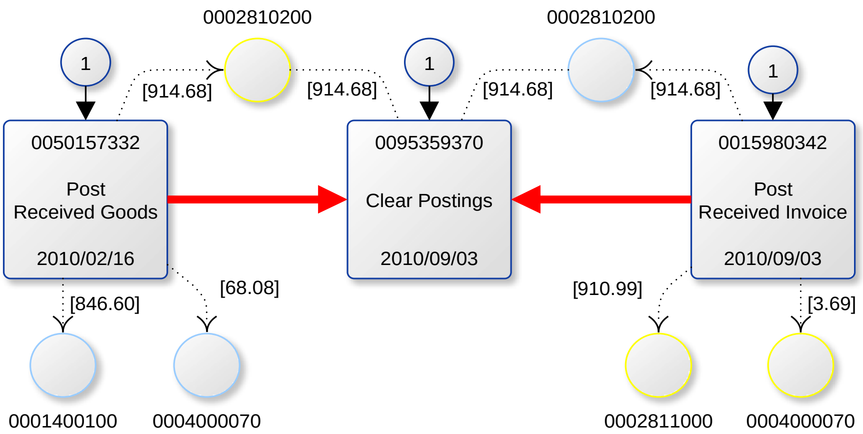
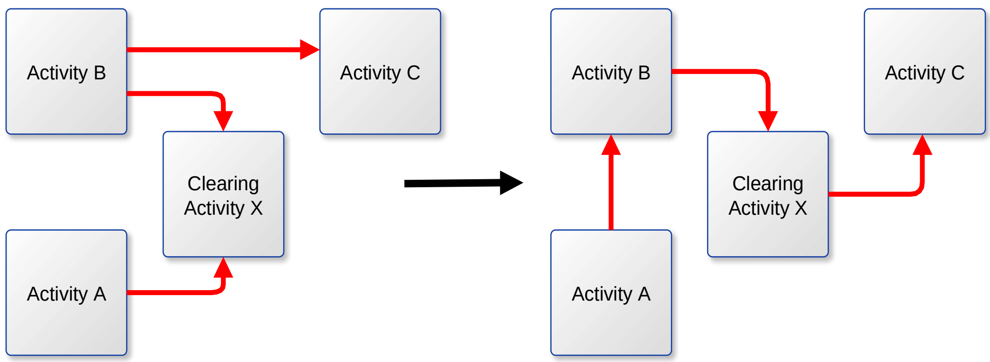


Figure 24 Clearing Deadlock (Werner and Nüttgens, 2014, p. 3892)

Clearing deadlocks can be removed by using graph transformation techniques (Rozenberg, 1997). Clearing activities can be identified by selecting all activities that did not post any items but only cleared items from other activities. Clearing activities commonly clear items from two or more other activities. At least one of these activities must have posted an additional item that was cleared by another activity different from the clearing activity for a deadlock to appear. Otherwise the clearing activity would represent a valid end node and such a constellation would not be considered a clearing deadlock (Werner and Nüttgens, 2014). The procedure to remove a clearing deadlock is graphically illustrated in Figure 25. Removing clearing deadlocks reduces the complexity and improves the readability of mined models but it violates requirement III described in chapter 4.2.3 because it alters the original data (Werner and Nüttgens, 2014). Applications in practice are necessary to evaluate if the alteration is acceptable or if it may not be used for specific purposes.



a) Deadlock Constellation

b) Deadlock Resolution

Figure 25 Deadlock Resolution (Werner and Nüttgens, 2014, p. 3893)

In summary the control flow of non-linear event logs can be inferred by:

(1) Defining causal dependencies

(2) Removing clearing deadlocks

(3) Defining start and end nodes

5.5 Multilevel Process Mining Algorithm

Chapters 5.1 to 5.4 showed different constructs, models and methods that solve specific problems for mining business process models in the context of financial audits. Listing 2 shows the final Multilevel Process Mining Algorithm (MLPM) that was designed by integrating the results from iterative research cycles. The individual research results have been described in the previous chapters and form the foundation for the algorithm laid out in the Listing 2.

Listing 2 Multilevel Process Mining Algorithm (Werner and Gehrke, 2015, p. 824)

```
Mine Cases (Section 1)
   D      set of all posting document numbers
   J      set of all journal entry item numbers
   ID     set of all case IDs
   D_i = Ø initially empty set of document numbers belonging to case i∈ID
   J_i = Ø initially empty set of journal entry item numbers for case i∈ID
  While D≠Ø
     Remove d ∈ D from D and insert d into D_i
        Insert all j ∈ J posted by d into J_i and remove j from J
        Insert all d ∈ D that cleared j ∈ J_i into D_i and remove d from D
        Repeat 54. and 55. for all d ∈ D_i and j ∈ J_i
```

```
Reconstruct Instance Graphs (Section 2)
  IG = ∅   initially empty set of instances graphs
  IG_i     instance graph (N_i, E_i, L_i, l_i) for case i with the set of nodes
           N_i ≠ ∅, E_i ⊆ N_i × N_i is the set of arcs, L_i the set of task labels
           and l_i: N_i → L_i is a labeling function mapping nodes onto L_i
 For all i ∈ ID
    Create n ∈ N_i for each d ∈ D_i with l_i(n)= transaction code of d
    Create e(n_j,n_k) ∈ E_i for each d_j and d_k ∈ D_i if d_k cleared an item
       j ∈ J_i that was posted by d_j
    Insert IG_i into IG

Reconstruct Instance Models (Section 3)
  IM = ∅   initially empty set of instances models
  IM_i     instance model (T_i, P_i, A_i, Σ_i, V_i, C_i, G_i, E_i, I_i) for case i ∈ ID
 For all i ∈ ID
    Set T_i = N_i
    For each e(n_j,n_k) ∈ E_i create p ∈ P_i, a(t_j,p), a(p,t_k)
    For each d ∈ D_i
       Create p ∈ P_i for each j ∈ J_i that was posted by d
       Create a(t,p) ∈ A_i for each j ∈ J_i that was posted by d and
         a(t,p), a(p,t) ∈ A_i for each j ∈ J_i that was cleared by d
    Aggregate all places p_k and p_j ∈ P_i if C_i(p_k) = C_i(p_j)
    Aggregate all transitions t_k and t_j ∈ T_i if l_i(t_k) = l_i(t_j)
    Insert IM_i into IM

Mine Process Models (Section 4)
  PM = ∅   initially empty set of process models
 Compute causal matrix CM(IM_i) for all i ∈ ID
 While IM ≠ ∅
    Remove IM_j from IM and insert IM_j into PM
    For each IM_k ∈ IM
       If CM(IM_j) = CM(IM_k)
         Merge IM_j and IM_k with T_jk=T_j∪T_k, P_jk=P_j∪P_k, A_jk=A_j∪A_k, Σ_jk=Σ_j∪Σ_k
         Aggregate all places p_l and p_m ∈ P_jk if C_jk(p_l) = C_jk(p_m)
         Aggregate all t_l and t_m ∈ T_jk if l_jk(t_l) = l_jk(t_m)
         Remove IM_k from IM
```

A complete description of the algorithm is available in (Werner and Gehrke, 2015). The algorithm is divided into the sections *Mine Cases, Reconstruct Instance Graphs, Reconstruct Instance Models*, and *Mine Process Models*. The algorithm produces models at different levels of abstraction (compare chapter 3.2.3) and therefore fulfills requirement VI described in 4.2.3. It takes unlabeled and non-linear event logs as input and produces CPN as specified in chapter 5.2. It integrates method fragments to label the event log by matching events to cases in *Section 1*, infers the control flow based on data dependency in *Section 2*, integrates the data perspective in *Section 3* and reduces the complexity of mined models in *Section 3 and 4*.

Figure 26 shows an example of a mined process model using the MLPM algorithm. It illustrates the control flow by showing the sequence of activities and the data flow by illustrating the amounts that were posted by the different activities on the financial accounts. The process model represents eight process instances. The MLPM algorithm can also be used to create the related instance models to enable the auditor to inspect individual instances on a detailed level.

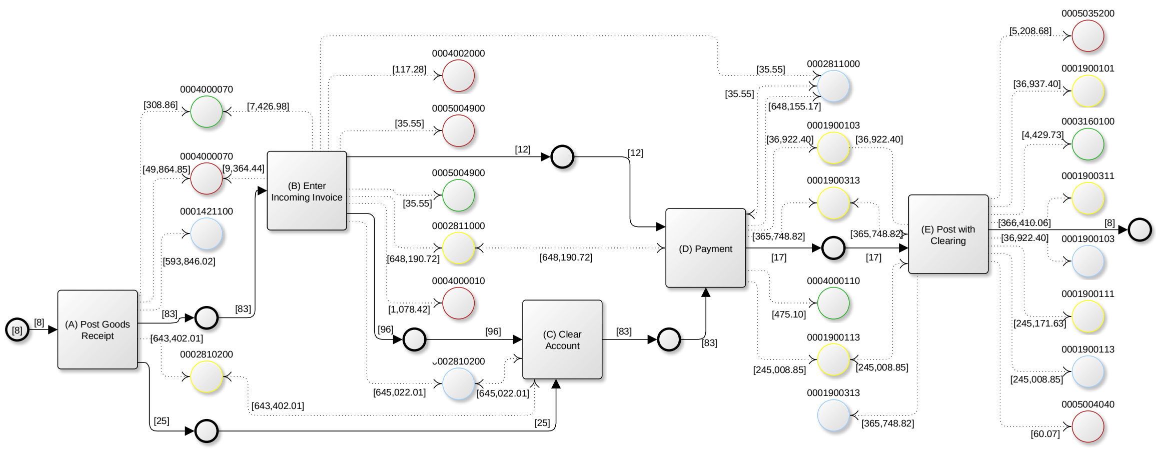


Figure 26 Example of a Mined Process Model (adapted from Werner and Gehrke, 2015, p. 828)[24]

5.6 Software Prototype

The research artifacts described in 5.2 to 5.5 were implemented in a software prototype. The conceptual structure of the prototype is shown in Figure 27. The prototype was implemented using the JAVA programming language (Oracle, 2014a) and the integrated development environment (IDE) NetBeans (Oracle, 2014b). The software prototype uses event logs as input. The event logs are provided by a separate extraction module that extracts relevant data tables from the source ERP systems. These are stored in a relational SQL database. The software consists of 10 software packages as listed in Table 10. The core package provides the functionality for graph modeling. Separate modules can be used for case matching and labeling, reconstruction of models based on the matched and labeled event data and aggregation for complexity reduction. Additional modules are used for managing the connection to the databases, the creation of output files and auxiliary purposes. The modular structure facilitates the integration of additional modules. The prototype uses two different kinds of databases. The original event log data is stored in a traditional relational SQL database (H2 Database Engine, 2014). The mined graphs are stored in a graph database Neo4j (Neo Technology Inc., 2014) for performance reasons.

Table 10 Prototype Modules

Module	Functionality
Core	Graph Modeling
M1	Case Matching and Labeling
M2	Reconstruction
M3	Aggregation
M4	Graphical User Interface
M5	SQL Database Management
M6	Graph Database Management
M7	File Writing
M8	Utilities
M9	Configuration

The prototype is able to create a variety of different output files that can be used as input for other software tools. Several independent software applications were used for specific purposes in the research cycle. The software yEd (yWorks GmbH, 2015) provides powerful layout algorithms and was used to graphically represent the mined models. Renew is a CPN

[24] The arc inscriptions for posting and clearing arcs only display the assigned constant for the posted or cleared value. The inscriptions for the account type, account number and credit or debit indicator are omitted for better readability. The same is the case for the inscriptions of the connected account places that only show the account number.

tool (University of Hamburg, 2015) that was used to simulate the execution of mined models to verify if the created models behaved like expected. Statistical analyses were carried out using Stata (StataCorp LP, 2014). ProM (Process Mining Group, 2015) and Disco (fluxicon, 2015) were used for evaluation purpose to compare mining results created by the prototype with those created by general purpose mining algorithms.

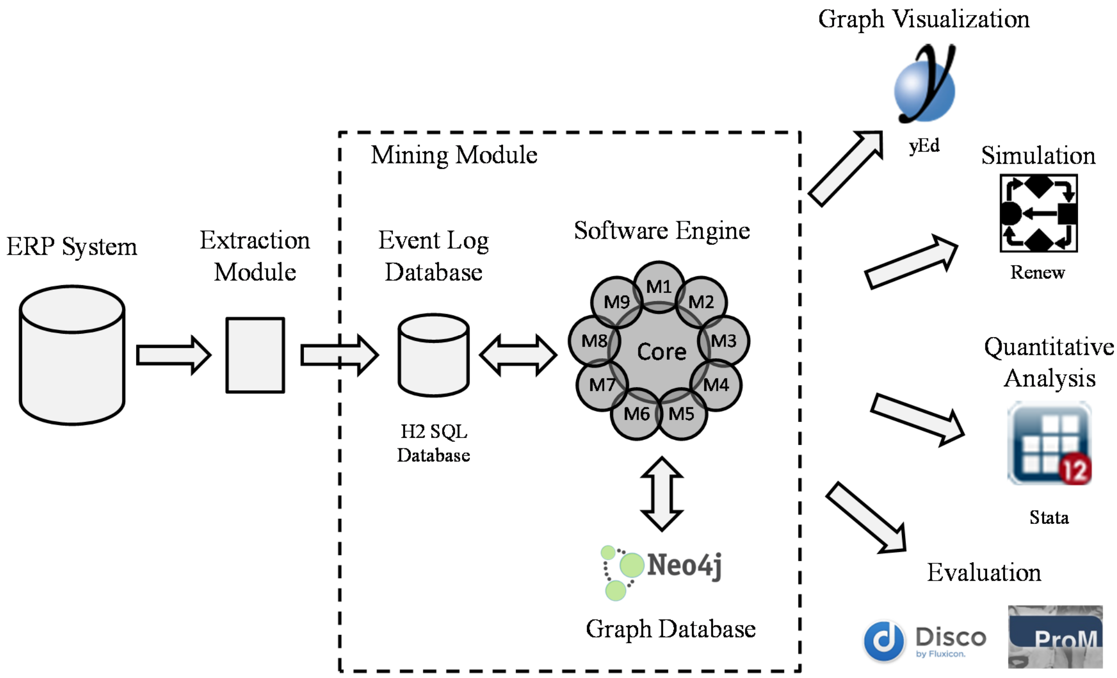


Figure 27 Prototype Structure

Figure 28 and Figure 29 show the prototype's graphical user interface (GUI). It is split into two screens. The main screen can be used to mine processes and to produce the different kind of output files. The created files are shown on the *Visualization* panel and can be opened using the yEd, Renew, ProM or Disco software. Two *Display* panels provide information on selected entries. The configuration screen provides configuration options on basic and advanced parameters.

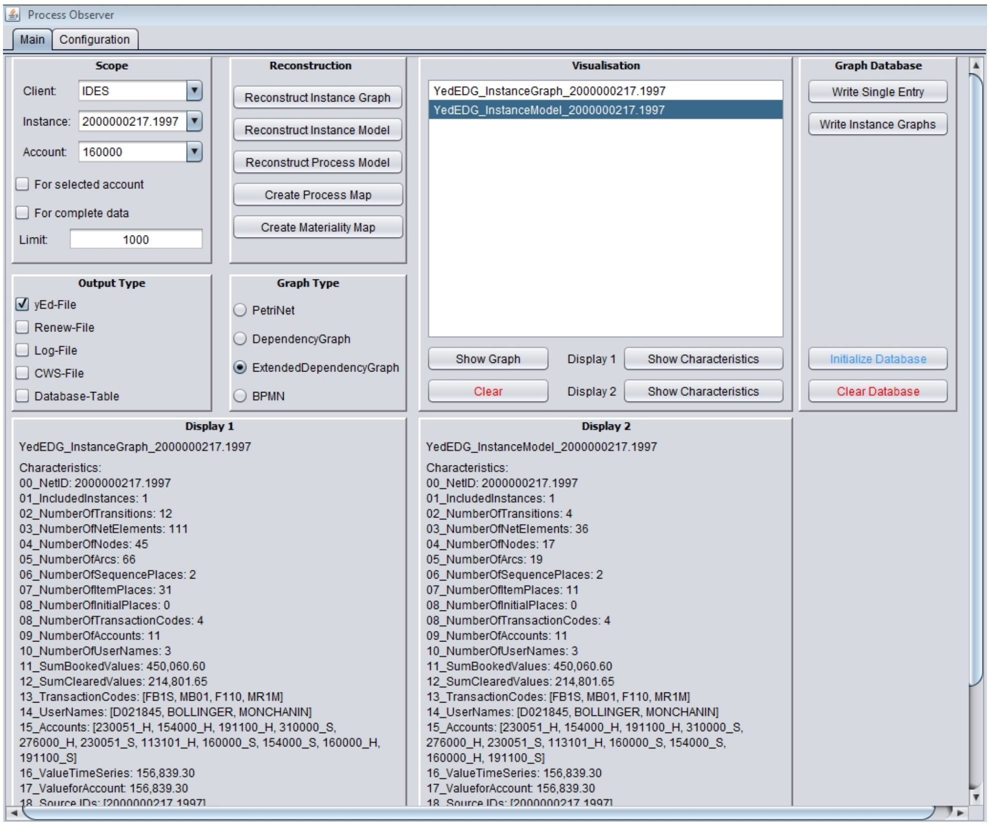


Figure 28 Prototype GUI Main Screen

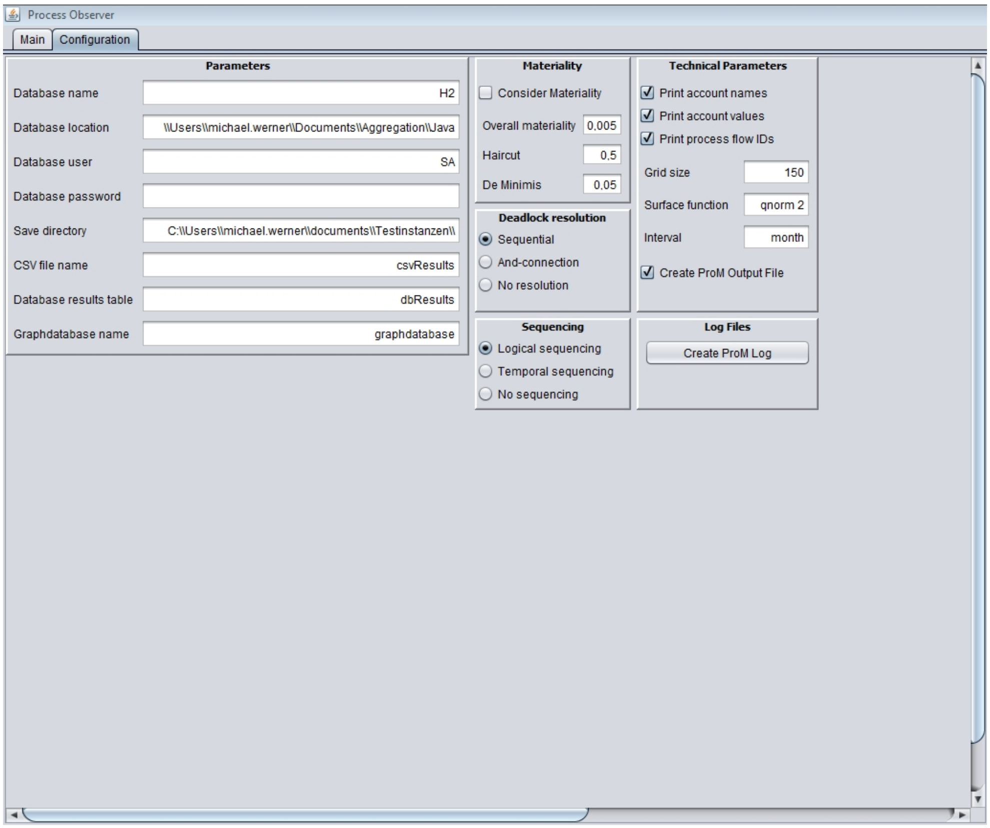


Figure 29 Prototype GUI Configuration Screen

6 Evaluation

Design science-oriented research is often criticized for its lack of scientific rigor due to insufficient evaluation efforts (Österle et al., 2010). Evaluation is an essential part of DSR as highlighted in Figure 30. The research presented in this paper was set up to ensure that sufficient evaluation can be carried out. The instantiation of the designed artifacts in a software prototype set the foundation to observe the behavior of the designed artifacts when exposed to test and real life data and to analyze the output that was created by the prototype. The following sub-chapters describe the evaluation efforts and results that could be achieved using different evaluation methods (compare chapter 3.4).

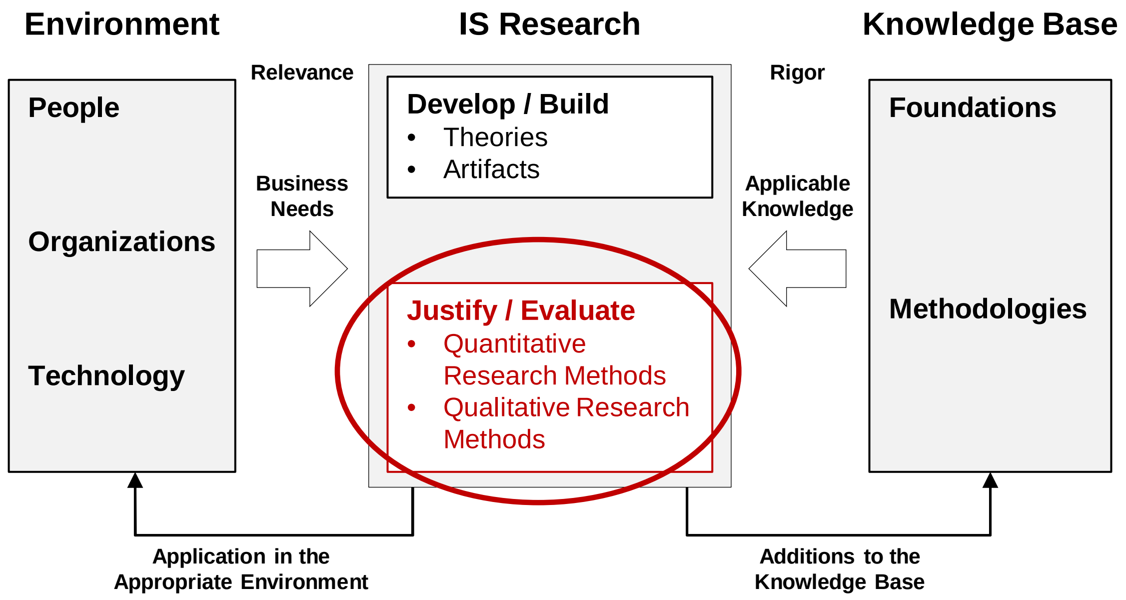


Figure 30 Evaluation in the Information Systems Research Framework (adapted from Hevner et al., 2004, p. 80)

6.1 Experimental Setup

A laboratory experiment was used to test the designed artifacts and to create output for observation and analysis purposes. Several experiments were conducted during the iterative research cycles but their basic structure was similar. It is illustrated in Figure 31. The first step was the extraction of the event data from the source systems using the extraction module. The extracted data was checked for first and second order defects (Kemper et al., 2010, pp. 29–32). The event log data served as input for the mining module. The implemented methods were used to create process models. The created models were observed using the yEd software to check if they fulfilled the desired requirements. The manual inspection was especially important in the first research cycles to check the proper operation of the implemented methods. The created models were further checked in respect to certain characteristics in simulations by using the Renew CPN tool (compare chapter 6.2). The output from the prototype was additionally compared to the results that could be achieved by using general purpose mining algorithms in ProM and Disco (Werner and Nüttgens, 2014). The created models and selected model characteristics were finally used for quantitative analyses (compare chapter 6.3).

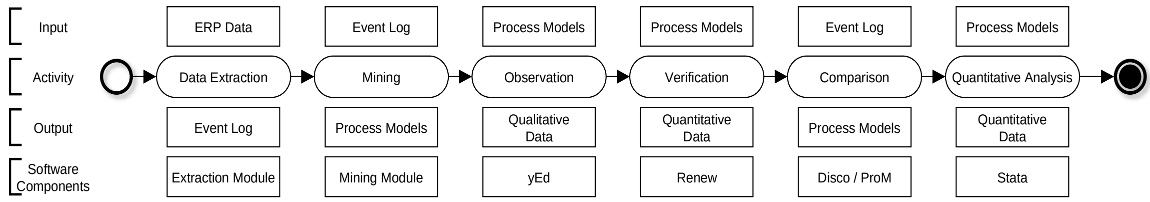


Figure 31 Experimental Setup

6.2 Simulation

Simulations can be used to gain information if it is difficult to represent the subject of investigation in a formal mathematical model due to its inherent complexity. The examples from chapter 5.3 illustrate that the complexity of the mined models tended to be very high. The Renew software tool was used to test on a sample basis if the prototype created sound process models according the definition used by (van der Aalst, 2011a, p. 39; Weske, 2012, pp. 326–329) with regard to:

- proper completion
- option to complete
- absence of dead transitions
- safeness

These criteria are formulated for low-level Petri Nets. The colored account places as defined in the CPN specification in chapter 5.2 were excluded for completion and safeness testing due to the fact that account places can hold multiple tokens that represent the journal entry items and that by definition remain on the account places even after the process is finished.

Figure 32 shows the CPN model from Figure 19 in Renew. Figure 32 shows the model in the initial state and Figure 33 the same model after the simulation has been executed. Every transition has fired and produced the tokens on the account places that represent the posted journal items.

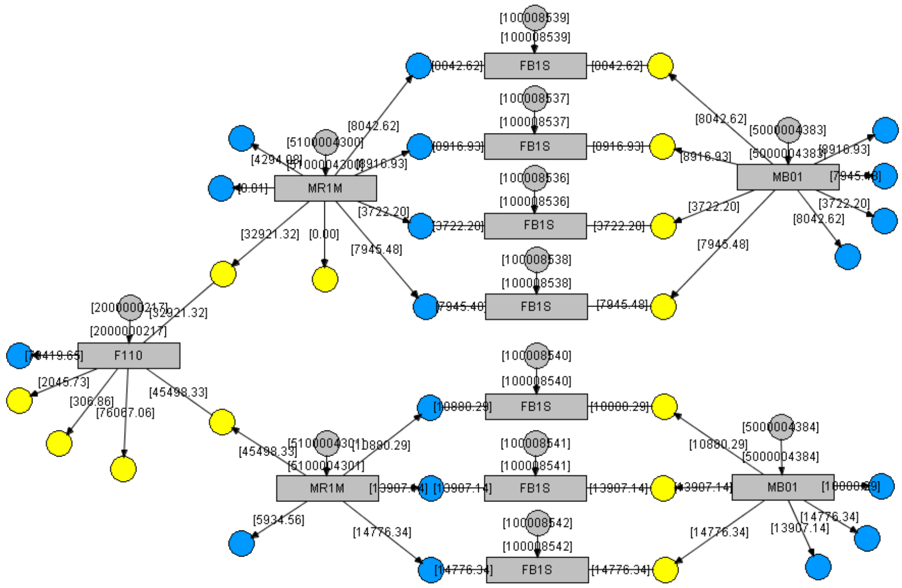


Figure 32 Example Process Instance before Simulation

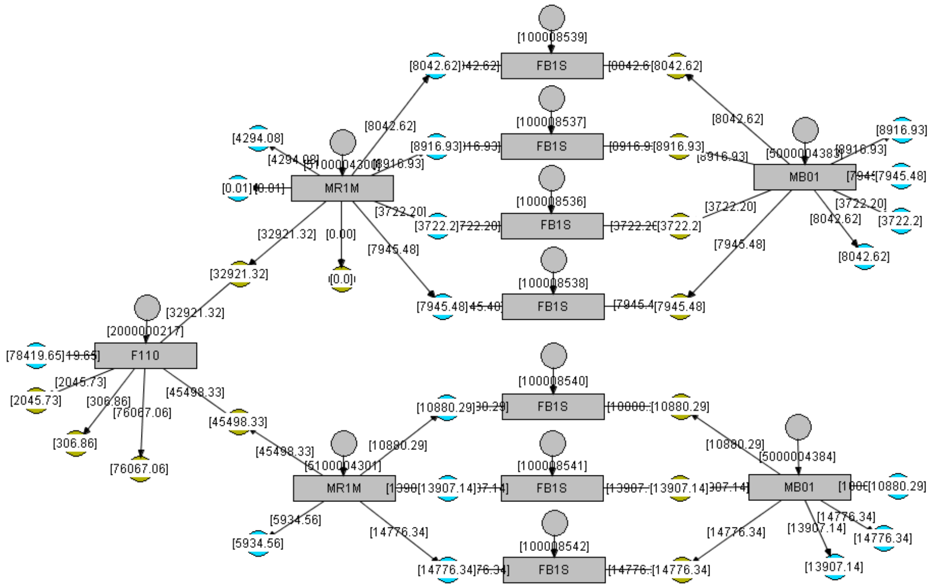


Figure 33 Example Process Instance after Simulation

6.3 Results Analysis

The produced process instances were used as a source data for descriptive analyses. Selected model characteristics like the number of transitions, model complexity or the number of represented instances in a process model were computed and used as input for statistical analyses (Werner et al., 2013). Four different source data sets were used during the overall research process. The first set was extracted from the SAP IDES test system. The test system is available for universities participating in the SAP University Alliance Program (SAP, 2015). Three data sets were gathered from companies operating in the retail, manufacturing and media industries. Table 11 provides an overview of the volume of the used data sets.

Table 11 Evaluation Data Sets (adapted from Werner et al., 2013, p. 381)

Data Set	1 SAP IDES	2 Retail	3 Manufacturing	4 Media
Number of journal entries	115,060	92,487	1,764,773	156,604
Number of journal entry items	419,106	222,901	7,395,434	559,506
Number of process instances	81,171	40,130	1,035,805	18,975
Number of processes	361	307	841	516
Covered period	17 years	1 year	1 year	1 year

The source data was used to create process instance and process models. Figure 34 to Figure 36 show the distributions of the number of process instances over the number of net elements with logarithmic scaling on the x- and y-axes. They illustrate that only very few instances consist of very many net elements. The vast majority of instances consist of relatively few net elements. Table 12 provides an overview of specific characteristic values of the distributions.

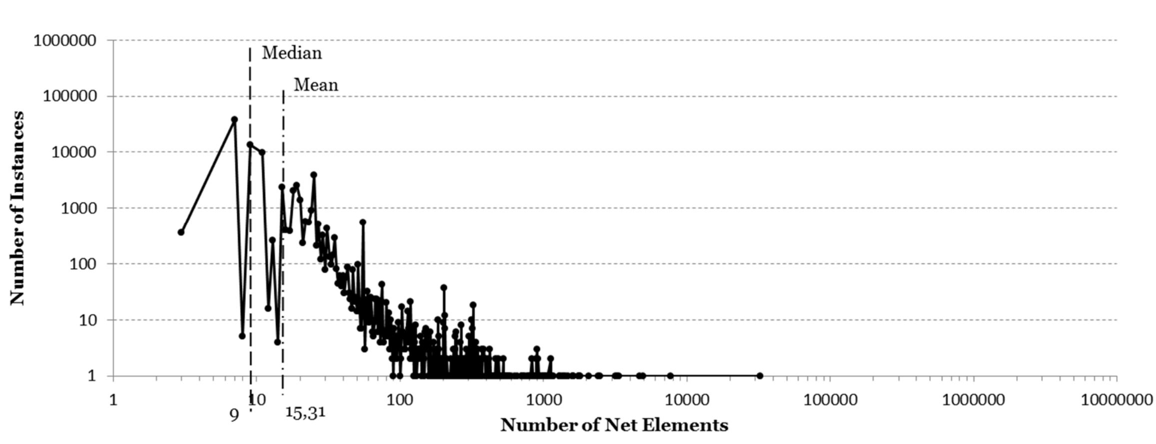


Figure 34 Data Set 1 Distribution of Number of Net Elements over the Number of Instances (Werner et al., 2013, p. 382)

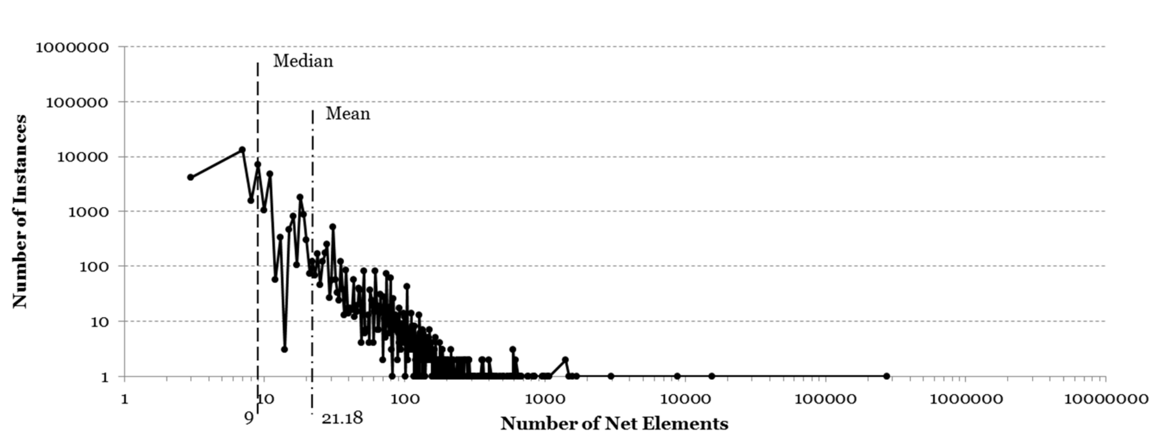


Figure 35 Data Set 2 Distribution of Number of Net Elements over the Number of Instances (Werner et al., 2013, p. 382)

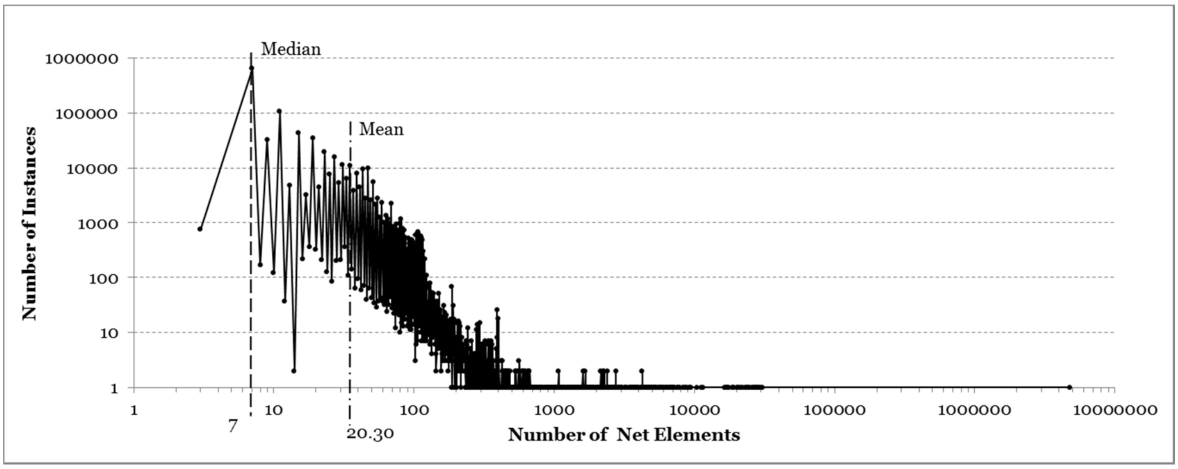


Figure 36 Data Set 3 Distribution of Number of Net Elements over the Number of Instances (Werner et al., 2013, p. 382)

Table 12 Overview of Net Size Distribution Characteristics (Werner et al., 2013, p. 383)

Data Set	#1 SAP IDES	#2 Retail	#3 Manufacturing
Mean value of net elements per instance	15.31	21.28	20.30
Median value of net elements per instance	9	9	7
Maximum of net elements per instance	32,519	275,870	4,769,379
Standard deviation of number of net elements	127.83	1,380.41	4,688.30

Figure 37 to Figure 39 show the distributions of transaction code combinations over the number of instances. The y-axes follow a logarithmic scaling. Each number on the x-axis represents a transaction code combination. The majority of instances only contain very few combinations. In combination with the results from analyzing the distribution of net sizes it can be assumed that the majority of instances are very limited in size and reveal the same transaction code combinations.

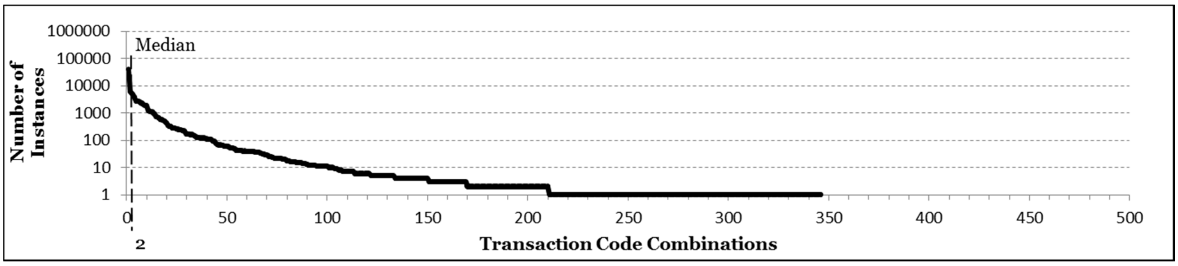


Figure 37 Data Set 1 Distribution of the Number of Instances for Different Transaction Code Combinations (Werner et al., 2013, p. 383)

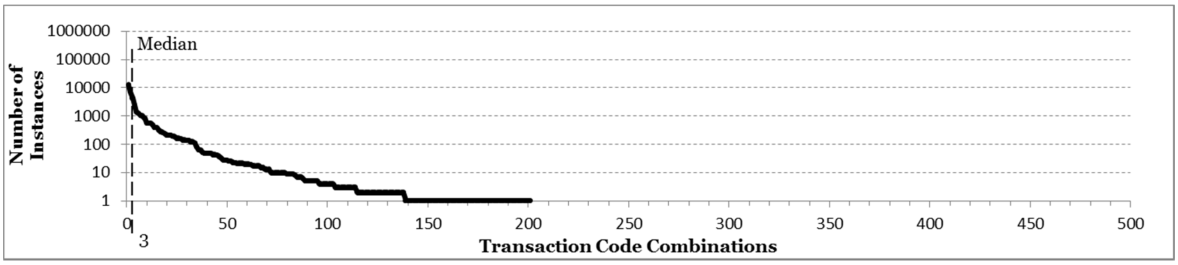


Figure 38 Data Set 2 Distribution of the Number of Instances for Different Transaction Code Combinations (Werner et al., 2013, p. 384)

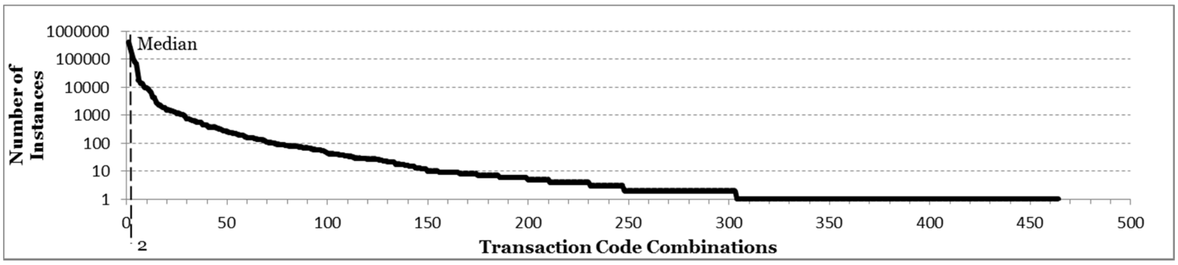


Figure 39 Data Set 3 Distribution of the Number of Instances for Different Transaction Code Combinations (Werner et al., 2013, p. 384)

Figure 40 to Figure 42 show the distributions of accounts in the instance models. Table 3 provides an overview of characteristic values of these distributions. The maximum number of used accounts in a process instance model is relatively low compared to the maximum of possible net sizes listed in Table 12. For the interpretation of the illustrated distributions and their characteristic values it is helpful to take into account that only one instance in data set two uses the maximum number of 596 accounts and one in data set three the maximum of 399 accounts. All other instances do not use more than 63 accounts in data set two and no more accounts than 45 in data set three. The observation of the distributions reveals that journal entry items are posted to relatively few accounts in a specific process instance. This is reasonable because a specific process generally only uses a subset of the available set of accounts.

Table 13 Overview of Account Distribution Characteristics (Werner et al., 2013, p. 385)

Data Set	1 SAP IDES	2 Retail	3 Manufacturing
Mean value	2.95	2.30	2.33
Median value	2	2	2
Maximum value	36	596	399
Standard deviation	1.61	3.37	0.92

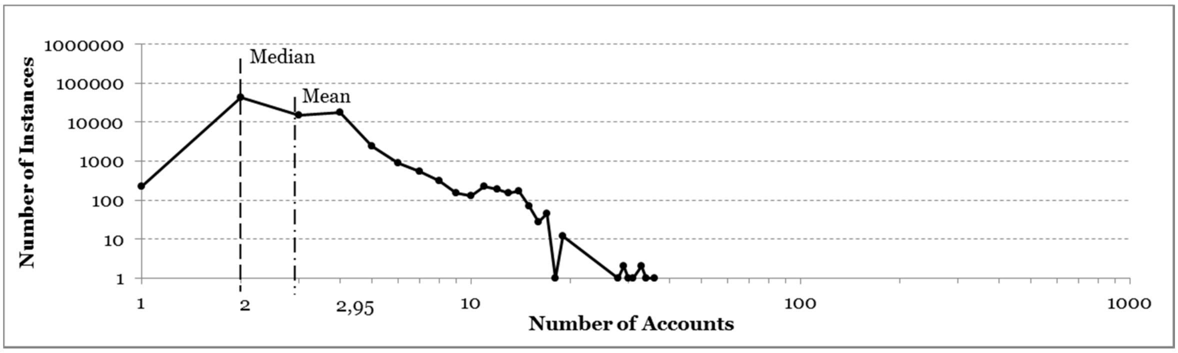


Figure 40 Data Set 1 Distribution of the Number of Instances over Number of Accounts (Werner et al., 2013, p. 386)

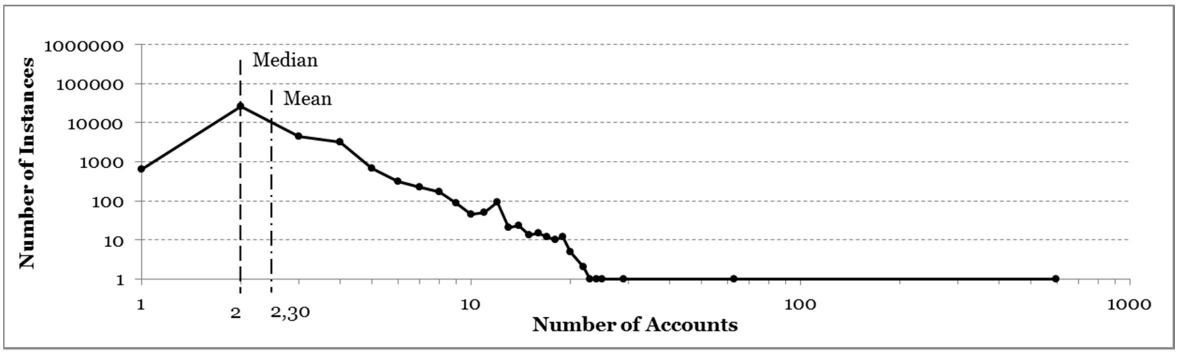


Figure 41 Data Set 2 Distribution of the Number of Instances over Number of Accounts (Werner et al., 2013, p. 386)

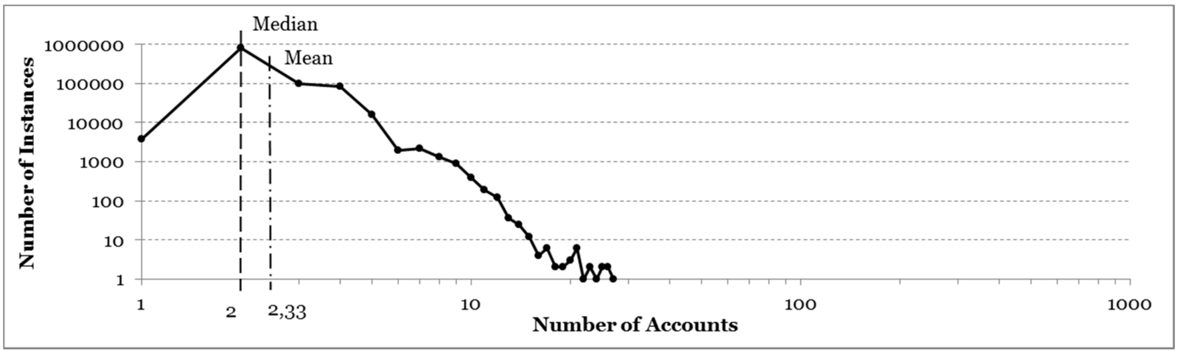


Figure 42 Data Set 3 Distribution of the Number of Instances over Number of Accounts (Werner et al., 2013, p. 386)

The mined process models from data sets 2 to 4 were further inspected in terms of achieved fitness and precision. The metrics f_{Path} and p_{Path} were used to measure fitness and precision (Werner and Gehrke, 2015) with:

$$f_{Path} = \frac{|\{p \mid p \in P_{PM} \land \in P_{IG}\}|}{|P_{IG}|} \qquad p_{Path} = \frac{|P_{PM}| - |\{p \mid p \in P_{PM} \land \notin P_{IG}\}|}{|P_{PM}|}$$

The average values over the data sets 2 to 4 for these metrics were:

f_{Path} = 1

p_{Path} = 0.81[25]

Figure 43 and Figure 45 show the frequency distributions of the mined process models depending on their model complexity measured as the number of included transitions. They show that the process models are distributed similarly to a normal distribution. Figure 44 and Figure 46 present scatter diagrams for data sets 3 and 4. They illustrate the distribution of the number of represented instance models in a process model depending on the model size. The diagrams show that the vast majority of process instances actually belong to very simple process models that contain only a few transitions.

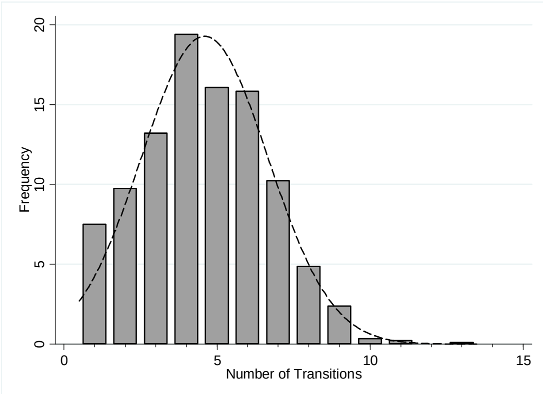


Figure 43 Frequency Distribution for Data Set 3 (Werner and Gehrke, 2015, p. 829)

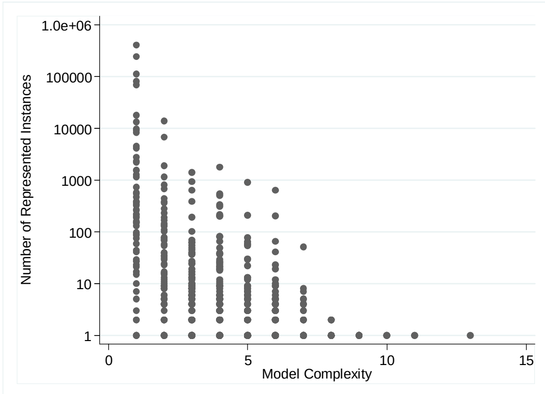


Figure 44 Scatter Diagram for Data Set 3[26] (Werner and Gehrke, 2015, p. 829)

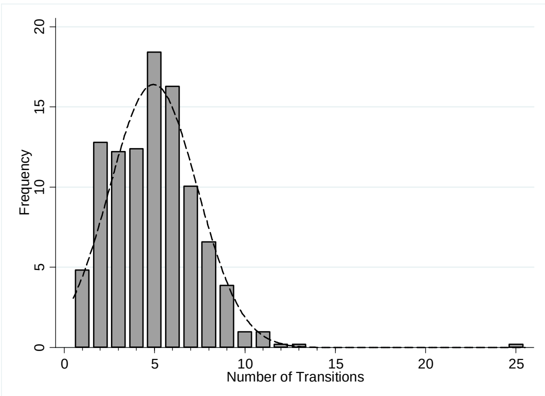


Figure 45 Frequency Distribution for Data Set 4 (Werner and Gehrke, 2015, p. 829)

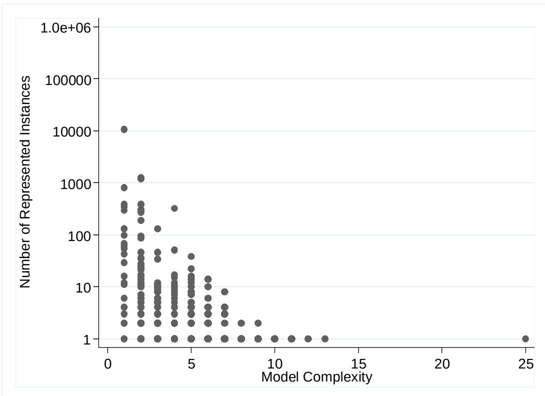


Figure 46 Scatter Diagram for Data Set 4[26] (Werner and Gehrke, 2015, p. 829)

[25] Process models including loops were excluded from the calculation.

[26] The dependent variables in Figure 44 and Figure 46 use a logarithmic scaling.

6.4 Requirements Fulfillment

Several requirements have been identified as being relevant for the development of a process mining algorithm for financial audits (compare chapter 4.2.3). Table 14 provides an overview of the different requirements and how they were met.

Table 14 Requirements Overview

Requirement		Accounted for by	Fulfillment
I	The mined process models should provide information on the control flow of process activities.	Using CPN to model the control flow.	Yes
II	The mined process models should provide information on the relationship of business processes and financial accounts.	Using CPN to integrate the data flow perspective.	Yes
III	The mining algorithm should not alter the original source data for creating process models.	Simulation results show that the algorithm does not alter the original source data and that it creates sound process models.	Yes[27]
IV	The mined process models should be as precise as possible.	Analysis results show that a high but not perfect precision is achieved.	Partially
V	The mined process models should be as fitting as possible.	Analysis results show that a perfect fitness is achieved.	Yes
VI	The mining algorithm should be able to produce process models at different abstraction levels.	The algorithm is able to create models on different abstraction levels (instance graphs representing the source data structure, process instance models and process models).	Yes
VII	The mining algorithm should be able to use unlabeled event logs as input.	The algorithm preprocesses the input data and assigns a case to each event.	Yes
VIII	The mining algorithm should be able to use event logs with non-linear traces as input.	The algorithm is able to take event logs with non-linear traces as input and infers the control flow by using data dependencies.	Yes

7 Diffusion

The last phase in DSR is the diffusion of the achieved research results into the application domain and the scientific knowledge base. Scientists use a variety of communication types to distribute achieved research results amongst scholarly colleagues and practitioners. Publication types encompass conference articles, presentations, scientific- or practice-oriented

[27] The MLPM algorithm can be used with our without enabled clearing deadlock resolution. If the resolution is disabled no alteration of the original source data takes place.

reports, dissertations or habilitations, theses, textbooks, websites, guidelines, standards, lectures, seminars, research proposals, implementations, spin-offs and so forth.

The publication of scientific articles as book chapters, in conference proceedings or journals in combination with the presentation of these articles in scientific conferences constitute the primary diffusion types that were used in the course of the presented research. Ten different publications were published that are listed in chapter 3.5. The publications followed the iterative research cycle (compare Figure 5 in chapter 3.3). All publications went through a rigorous review process. 28 independent and formal reviews from scientists were received for the listed publications in the course of submission, revision and acceptance of the published papers. The review and feedback from scientific experts in the relevant research area ensured a high quality of the research outputs and provided an external evaluation of the achieved results that could be used to improve the research outputs.

The research from the various publications was presented at scientific conferences. Table 15 lists the conferences and presented papers.

Table 15 Overview Conferences and Presented Papers

EMISA 2011	Discussion and presentation of the research exposé at the doctorial consortium of the 4th International Workshop on Enterprise Modeling and Information Systems Architectures in Hamburg
HICSS 2012	Presentation of a the research paper *Business Process Mining and Reconstruction for Financial Audits* (Werner et al., 2012a) at the 45th Hawaii International Conference on System Sciences in Wailea
BPM 2012	Discussion and presentation of the research exposé at the doctorial consortium of the 10th Business Process Management conference in Tallinn
ICIS 2012	Presentation of the Research-in-Progress paper *Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits* (Werner et al., 2012b) at the 33rd International Conference on Information Systems in Orlando sponsored by a scholarship of the German Academic Exchange Service (DAAD)
WI 2013	Presentation of the paper *Towards Automated Analysis of Business Processes for Financial Audits* (Werner et al., 2013) at the 11th International Conference on Wirtschaftsinformatik in Leipzig
ER 2013	Presentation of the paper *Colored Petri Nets for Integrating the Data Perspective in Process Audits* (Werner, 2013) at the 32nd International Conference on Conceptual Modeling (ER 2013) in Hong Kong
HICSS 2014	Presentation of the paper *Improving Structure: Logical Sequencing of Process Models* (Werner and Nüttgens, 2014) at the 47th Hawaii International Conference on System Sciences in Waikoloa sponsored by a scholarship of the German Academic Exchange Service (DAAD)
ECIS 2014	Presentation of the paper *Who is afraid of the Big Bad Wolf - Structuring Large Design Science Research Projects* (Werner et al., 2014) at the 22nd European Conference on Information Systems in Tel Aviv

Research results were also continuously published on the research project website (University of Hamburg, 2014) and presented in workshops to the project partners from the industry to achieve a diffusion into the application domain. The presented research results also served as input for the development of a commercial software application that is currently under development. The application of the presented academic prototype in field experiments is planned for future research.

The positive feedback from scientists in the form of reviews for published papers and discussions at several conferences has shown that the aim has been achieved to distribute the research results into the academic knowledge base. Distribution into the application domain was initiated via the cooperation with project partners from industry and contacts with commercial software companies such as fluxicon, the software development company for the process mining tool Disco (fluxicon, 2015).

The presented research results have further been used as a scientific foundation to prepare and submit two research proposals to the German Research Foundation (DFG).

8 Summary and Outlook

8.1 Summary

Financial audits play a significant role in our economic system as an important control mechanism to safeguard the correctness and reliability of published financial information. Process audits are an important part of financial audits but public accountants face difficulties in conducting these audits with the ongoing integration of information systems for the operation of business processes, increased process complexities and growing amounts of processed data. Manual audit procedures that are traditionally used to carry out process audits become inefficient or even ineffective in environments that exhibit such characteristics.

The aim of the research that is presented in this thesis is the development of data analysis techniques that can be used to support auditors in process audits and that are able to produce reliable process models by using the data which is stored in information systems that process financially relevant transactions. It has been carried out using a design science-oriented research approach due to the proximity of the research question to practical problems and the objective to create research artifacts that contribute both to the scientific knowledge base and the application domain. This aim has been achieved by the development of several research artifacts and in particular a special purpose mining algorithm that takes the requirements from the application domain and knowledge from the scientific knowledge base into account.

The research followed the design science phases analysis, design, evaluation and diffusion. These phases were iterated in several cycles leading to different developed methods that have been formalized as algorithms as well as constructs and models for describing the problem domain and solution space. The main output of the research is a Multilevel Process Mining algorithm that produces a variety of models. The algorithm integrates the control flow and data flow perspective. It operates on different abstraction levels, creates precise and fitting process models, accepts unlabeled and non-linear event logs from ERP systems as input, and considers data relationships to infer the control flow. It is an innovative solution especially designed for the application in financial audits. Gregor and Hevner provide a framework that can be used to categorize the knowledge contribution of designed artifacts in DSR (Gregor and Hevner, 2013). They use the four domains routine design, improvement, exaptation and invention to categorize the knowledge contribution. The presented research work can be assigned to the exaptation quadrant because the main objective is to provide a solution for a new application area by partly using already existing knowledge. But it also affects the improvement quadrant by introducing new methods to model the control flow and data flow simultaneously in mined models, to label event logs from ERP systems and to infer the control flow relying on data dependencies between events recorded in non-linear event logs. Gregor and Hevner further differentiate between three levels of contribution types that range from abstract, complete and mature knowledge on the highest level to more specific, limited and less mature knowledge on the lowest level. The research results presented in this thesis are mainly located on the second level providing constructs and methods for the mining of process models and on the first level presenting an instantiated software artifact. The results that can be achieved by analyzing the mining outcomes can also be input for the third and highest knowledge contribution level.

Although the MLPM algorithm is meant to be a special purpose mining algorithm many research results that were achieved on the route to the final algorithm are also generally applicable. The integration of the data perspective is an important aspect for process mining and it has not been addressed yet sufficiently in the academic arena (de Leoni and van der Aalst, 2013; Stocker, 2012). The provided CPN specification shows a solution how the data perspective can be integrated which can also be adapted for other application areas. Contemporary process mining algorithms require labeled event logs and strict linearly ordered events. The MLPM algorithm accepts unlabeled and non-linear event logs as input. Labeling events is an important aspect in process mining and not well researched (Ferreira and Gillblad, 2009). Non-linear event logs are also present in other application areas like process mining in online discussion forums (Wang et al., 2014).

The research artifacts have been implemented in an academic software prototype that can be used in real scenarios and that is able to take test and real life data from SAP ERP systems as input. The researched artifacts have been evaluated extensively in simulations with test and real data. The produced models have been analyzed using descriptive statistics. They provide a novel empirical data base for observing real business processes in organizations. Knowledge on business processes gained on the basis of data that is created by using process mining techniques is still scarce. The provided information on the mined models can be seen as a first step to broaden this knowledge base.

The MLPM algorithm is supposed to be used to support public accountants in process audits. The application of automated data analysis techniques alone will most likely not prevent accounting scandals. But it can be used as a tool to improve process audits. Its application enables auditors to receive all-embracing and reliable information on the audited business processes and their relationships to the financial accounts. It can be used as a foundation to automate the analysis and audit of standard processes that generally exhibit a lower risk than non-standard process that usually exhibit a higher risk. Efficient and effective automated analyses of standard processes set free audit resources that can then be spent on non-standard transactions.

8.2 Limitations

The MLPM algorithm is able to discover process models in accordance with the identified requirements to a large extent as discussed in chapter 6.4. Requirement IV could just partially be met. The mined models are not absolutely precise. It has to be validated in further research if the achieved level of precision is sufficient in practice.

Several additional limitations have to be taken into account when discussing the achieved research results. The mined process models do not represent sound workflow nets according to commonly used definitions (van der Aalst, 2011a, p. 39; Weske, 2012, pp. 326–329). The MLPM algorithm can be used to mine precise and fitting process models based on the available source data from ERP systems. Formally well-structured process models are not critical from this point of view. Sound process models could be achieved by neglecting the data perspective by not modeling the account places. The remaining models would then represent sound workflow nets but without representing the data flow perspective.

All test and real data sets were extracted from SAP ERP systems. It can therefore not be concluded that the research results also hold true for other data sources. But the MLPM algorithm exploits the general structure of accounting entries as described in chapter 4.2.2. This structure is independent from the implemented data structures of a particular ERP system.

The resolution of deadlocks as discussed in chapter 5.4 changes the original source data and therefore violates requirement III described in chapter 4.2.3. The deadlock resolution can be enabled or disabled in the implemented prototype. Field experiments could provide further information on the feasibility of the deadlock resolution in real settings.

Some mined process models showed loops. These loops can occur when a transaction has cleared a journal item that was posted by the same transaction or by a transaction located in the subsequent execution path. This constellation leads to a deadlock in the process model. Such a deadlock is not critical for the interpretation of the model from an audit perspective but generally not desired for the modeling of correct process models. A solution could be the prevention of aggregating transitions carrying the same label if this would result in a loop.

The mining algorithm produces precise and fitting process models at the cost of lacking generalization. It is therefore not applicable for scenarios with highly variable business processes. In the worst case scenario all process instances show a different behavior. The mining algorithm would then produce a process model for each process instance. The data presented in Table 11 shows that this will most likely not occur in the context of financial audits. Business processes are usually standardized to a certain degree when they are supported by ERP systems. The data shows that the number of process models ranges from 307 for the smallest data set to 841 models for the largest. This may still seem to be a big number, but many of the process models represent trivial processes. 63 process models in data set 3 only consist of one transition. The activities in these processes were mostly carried out by using a single general purpose transaction. They are of little interest from a process perspective. The process models that reflect the major business processes are those that contain many transitions and represent a high number of process instances. Data set 3 contains 135 process models consisting of 5 transitions. But just two of them already represent 62% of the instances of this category. It can therefore be assumed that the majority of models for more complex processes only represent very infrequent behavior and can be tested traditionally by inspecting individual journal entries.

8.3 Outlook and Future Research

The presented research results support the automated discovery and analysis of process models in process audits and lay the foundation for the development of completely automated audit procedures as conceptually illustrated in chapter 5.1. The next research step would be the integration of application controls.

Several aspects that are important from an audit perspective have not been considered sufficiently yet. Materiality is a key concept in financial audits. Research on the role of business processes from a materiality perspective is currently pending. The presented MLPM algorithm is able to produce process models. But it does not provide an overall process map as used as a metaphor in chapter 5.1. Methods for mining process maps are currently under development but have not been evaluated yet.

The prototype uses data from SAP systems. Data from other source systems have not been considered extensively yet. The development of extensions to integrate other data sources is intended for future research. The used event log data can be used as input for academic and commercial software tools like ProM and Disco. But due to the non-linear ordering of events in these logs the non-linear traces have to be sliced into linear traces as described in (Müller-Wickop and Schultz, 2013b). But this slicing leads to undesired side effects like the duplication of events and related data values. The integration of the designed solutions into academic software tools like ProM is planned for future research.

9 Bibliography

Accorsi, R., Lehmann, A., 2012. Automatic Information Flow Analysis of Business Process Models, in: Barros, A., Gal, A., Kindler, E. (Eds.), Business Process Management, Lecture Notes in Computer Science. Springer, pp. 172–187.

Adriansyah, A., van Dongen, B.F., van der Aalst, W.M.P., 2011. Conformance Checking Using Cost-based Fitness Analysis, in: 15th IEEE International Enterprise Distributed Object Computing Conference. pp. 55–64.

Agrawal, R., Gunopulos, D., Leymann, F., 1998. Mining Process Models from Workflow Logs, in: Proc. Sixth Int'l Conf. Extending Database Technology. pp. 469–483.

AIS, 2014. Senior Scholars' Basket of Journals [WWW Document]. URL http://ais.site-ym.com/?SeniorScholarBasket (accessed 7.16.14).

Alturki, A., Gable, G.G., Bandara, W., 2011. A Design Science Research Roadmap, in: Service-Oriented Perspectives in Design Science Research. Springer, pp. 107–123.

Archer, L.B., 1984. Systematic Method for Designers, in: Developments in Design Methodology. John Wiley, London, pp. 57–82.

Australian Research Council, 2014. The Excellence in Research for Australia (ERA) - Australian Research Council (ARC) [WWW Document]. URL http://www.arc.gov.au/era/ (accessed 8.21.14).

Baskerville, R., Lyytinen, K., Sambamurthy, V., Straub, D., 2010. A Response to the Design-oriented Information Systems Research Memorandum. European Journal of Information Systems 20, 11–15.

Becker, J., Probandt, W., Vering, O., 2012. Grundsätze ordnungsmäßiger Modellierung Konzeption und Praxisbeispiel für ein effizientes Prozessmanagement. Springer Gabler, Berlin; Heidelberg.

Bezerra, F., Wainer, J., 2013. Algorithms for Anomaly Detection of Traces in Logs of Process Aware Information Systems. Information Systems 38, 33–44.

Bhattacherjee, A., 2012. Social Science Research: Principles, Methods, and Practices. A. Bhattacherjee, Tampa, Fla.

Bierstaker, J., Janvrin, D., Lowe, D.J., 2014. What Factors Influence Auditors' Use of Computer-assisted Audit Techniques. Advances in Accounting 30, 67–74.

BPM, 2014. BPM 2014 [WWW Document]. URL http://bpm2014.haifa.ac.il/ (accessed 10.29.13).

Braun, R.L., Davis, H.E., 2003. Computer-assisted Audit Tools and Techniques: Analysis and Perspectives. Managerial Auditing Journal 18, 725–731.

Brinkkemper, S., 1996. Method Engineering: Engineering of Information Systems Development Methods and Tools. Information and Software Technology 38, 275–280.

Caron, F., Vanthienen, J., Baesens, B., 2013. Comprehensive Rule-Based Compliance Checking and Risk Management with Process Mining. Decision Support Systems 54, 1357–1369.

Chen, H., Chiang, R.H.L., Storey, V.C., 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36, 1165–1188.

Chuprunov, M., 2012. Handbuch SAP-Revision: internes Kontrollsystem und GRC. Galileo Press, Bonn.

Cole, R., Purao, S., Rossi, M., Sein, M.K., 2005. Being Proactive: Where Action Research Meets Design Research, in: Proceedings of the 26th International Conference on Information Systems.

Cook, J.E., Wolf, A.L., 1999. Software Process Validation: Quantitatively Measuring the Correspondence of a Process to a Model. ACM Transactions on Software Engineering and Methodology 8, 147–176.

Cook, J.E., Wolf, A.L., 1998a. Discovering Models of Software Processes from Event-based Data. ACM Transactions on Software Engineering and Methodology 7, 215–249.

Cook, J.E., Wolf, A.L., 1998b. Event-based Detection of Concurrency. ACM SIGSOFT Software Engineering Notes 23, 35–45.

CORE, 2014. Computing Research & Education & Conference Rankings [WWW Document]. URL http://core.edu.au/index.php/conference-rankings (accessed 8.21.14).

de Leoni, M., van der Aalst, W.M.P., 2013. Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments, in: 28th Annual ACM Symposium on Applied Computing. Coimbra, Portugal, pp. 1454–1461.

de Medeiros, A.K.A., 2006. Genetic Process Mining. Eindhoven University of Technology, Eindhoven.

Deutscher Bundestag, 2013. Handelsgesetzbuch.

Diestel, R., 2010. Graph Theory, 4th ed. Springer, Heidelberg; New York.

Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A., 2013. Fundamentals of Business Process Management. Springer.

Eekels, J., Roozenburg, N.F.M., 1991. A Methodological Comparison of the Structures of Scientific Research and Engineering Design: Their Similarities and Differences. Design Studies 12, 197–203.

Ferreira, D., Gillblad, D., 2009. Discovering Process Models from Unlabelled Event Logs. Business Process Management 143–158.

Fettke, P., 2006. State-of-the-Art des State-of-the-Art. Wirtschaftsinformatik 48, 257–266.

fluxicon, 2015. Process Mining and Process Analysis - Fluxicon [WWW Document]. URL www.fluxicon.com (accessed 2.1.13).

Fowler, J.F.J., 1984. Survey Research Methods, Auflage: 4th. ed. SAGE Publications, Inc.

Gehrke, N., Müller-Wickop, N., 2010. Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems, in: Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru.

Gehrke, N., Werner, M., 2013. Process Mining. wisu - das wirtschaftsstudium 934–943.

Google, 2014. Visualization: Geochart - Google Charts — Google Developers [WWW Document]. URL https://developers.google.com/chart/interactive/docs/gallery/geochart?hl=de (accessed 7.16.14).

Gregor, S., 2006. The nature of theory in information systems. MIS Quarterly 30, 611–642.

Gregor, S., Baskerville, R., 2012. The Fusion of Design Science and Social Science Research. ISF 2012.

Gregor, S., Hevner, A.R., 2013. Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37, 337–355.

Gubrium, J.F., Holstein, J.A., 2002. Handbook of Interview Research: Context and Method. SAGE.

Gujarati, D.N., Porter, D.C., 2009. Basic Econometrics. McGraw-Hill Irwin, Boston.

Günther, C., van der Aalst, W.M.P., 2007. Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. Business Process Management 328–343.

Günther, C.W., Verbeek, E.H., 2012. XES Standard Definition. Eindhoven University of Technology, Eindhoven.

H2 Database Engine, 2014. H2 Database Engine [WWW Document]. URL http://www.h2database.com/html/main.html (accessed 7.25.14).

Hansen, H.R., Neumann, G., 2009. Wirtschaftsinformatik 1 Grundlagen und Anwendungen, 10th ed. Lucius & Lucius, Stuttgart.

Harmsen, A.F., Brinkkemper, J.N., Oei, H., 1994. Situational Method Engineering for Information System Project Approaches. University of Twente, Department of Computer Science.

Heckel, R., 2006. Graph Transformation in a Nutshell. Electronic Notes in Theoretical Computer Science 148, 187–198.

Heinrich, L.J., Heinzl, A., Roithmayr, F., 2004. Wirtschaftsinformatik-Lexikon. Oldenbourg, München.

Herbst, J., 2003. Ein induktiver Ansatz zur Akquisition und Adaption von Workflow-Modellen. Tenea Verlag Ltd.

Herbst, J., 2000a. Dealing with Concurrency in Workflow Induction, in: European Concurrent Engineering Conference. SCS Europe.

Herbst, J., 2000b. A Machine Learning Approach to Workflow Management, in: López de Mántaras, R., Plaza, E. (Eds.), Proceedings ofMachine Learning: 11th European Conference Machine Learning. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 183–194.

Herbst, J., Karagiannis, D., 1999. An Inductive Approach to the Acquisition and Adaptation of Workflow Models, in: Proceedings of the IJCAI. pp. 52–57.

Herbst, J., Karagiannis, D., 1998. Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models, in: Proceedings

9th International Workshop on Database and Expert Systems Applications (DEXA'98). pp. 745–752.

Hevner, A., Chatterjee, S., 2010. Design Science Research in Information Systems, in: Design Research in Information Systems. Springer US, Boston, MA, pp. 9–22.

Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design Science in Information Systems Research. MIS Quarterly 28, 75–105.

IDW, 2009. IDW PS 261 Feststellung und Beurteilung von Fehlerrisiken und Reaktionen des Abschlussprüfers auf die beurteilten Fehlerrisiken.

IFAC, 2012. ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment.

IFAC, 2009. ISA 320 Materiality in Planning and Performing an Audit.

Jans, M., 2012. Process Mining in Auditing: From Current Limitations to Future Challenges, in: Daniel, F., Barkaoui, K., Dustdar, S. (Eds.), Business Process Management Workshops. Springer, Berlin, Heidelberg, pp. 394–397.

Jans, M., Alles, M., Vasarhelyi, M., 2011. Process Mining of Event Logs in Internal Auditing: A Case Study, in: 2nd International Symposium on Accounting Information Systems.

Jans, M., Alles, M., Vasarhelyi, M., 2010. Process Mining of Event Logs in Auditing: Opportunities and Challenges. Working paper. Hasselt University. Belgium.

Jans, M., Lybaert, N., Vanhoof, K., Van Der Werf, J.M., 2008. Business Process Mining for Internal Fraud Risk Reduction: Results of a Case Study, in: Proceedings of the International Research Symposium on Accounting Information Systems, Paris.

Jensen, K., Kristensen, L.M., 2009. Coloured Petri Nets. Springer.

Jick, T.D., 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. Administrative Science Quarterly 24, 602–611.

Kemper, H.-G., Mehanna, W., Baars, H., 2010. Business Intelligence - Grundlagen und praktische Anwendungen. Vieweg + Teubner, Wiesbaden.

Konradin Mediengruppe, 2011. ERP Studie 2011.

Krcmar, H., 2010. Informationsmanagement. Springer, Berlin; Heidelberg.

Küting, K., Reuter, M., 2004. Bilanzierung im Spannungsfeld unterschiedlicher Adressaten. DSWR 9, 230–233.

March, S.T., Smith, G.F., 1995. Design and Natural Science Research on Information Technology. Decision Support Systems 15, 251–266.

Maxeiner, M.K., Küspert, K., Leymann, F., 2001. Data Mining von Workflow-Protokollen zur teilautomatisierten Konstruktion von Prozessmodellen, in: Datenbanksysteme in Büro, Technik Und Wissenschaft (BTW), 9. GI-Fachtagung,. Springer-Verlag, pp. 75–84.

Mingers, J., 2001. Combining IS Research Methods: Towards a Pluralist Methodology. Information Systems Research 12, 240–259.

Müller, R.M., Lenz, H.-J., 2013. Business Intelligence, eXamen.press. Springer, Berlin, Heidelberg.

Müller-Wickop, N., 2014. Integration von Wertflüssen in Geschäftsprozessmodellierungssprachen: Ein gestaltungsorientierter Ansatz zur Unterstützung von Revisoren bei der wertflussorientierten Planung und Durchführung von Prozessprüfungen. Hamburg.

Müller-Wickop, N., Schultz, M., 2013a. Modelling Concepts For Process Audits - Empirically Grounded Extension Of BPMN, in: Proceedings of the 21st European Conference on Information Systems, Utrecht.

Müller-Wickop, N., Schultz, M., 2013b. ERP Event Log Preprocessing: Timestamps vs. Accounting Logic, in: Proceedings of the 8th International Conference on Design Science Research in Information Systems and Technology, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 105–119.

Müller-Wickop, N., Schultz, M., Peris, M., 2013. Towards Key Concepts for Process Audits – A Multi-Method Research Approach, in: Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics, Utrecht.

Naumann, J.D., Jenkins, A.M., 1982. Prototyping: the New Paradigm for Systems Development. MIS Quarterly 29–44.

Neel, D.L., Orrison, M.E., 2006. The Linear Complexity of a Graph. The Electronic Journal of Combinatorics 13, 1–19.

Neo Technology Inc., 2014. Neo4j - The World's Leading Graph Database [WWW Document]. URL http://www.neo4j.org/ (accessed 7.25.14).

Niehaves, B., 2007. On Epistemological Diversity in Design Science: New Vistas for a Design-oriented IS Research, in: Proceedings of the 28th International Conference on Information Systems, Montreal, pp. 1–13.

Nunamaker, J.F., Chen, M., Purdin, T.D.M., 1991. Systems Development in Information Systems Research. Journal of Management Information Systems 7, 89–106.

Oracle, 2014a. Java [WWW Document]. URL https://www.java.com/de/ (accessed 7.25.14).

Oracle, 2014b. Welcome to NetBeans [WWW Document]. URL https://netbeans.org/ (accessed 7.25.14).

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J., 2010. Memorandum on design-oriented information systems research. European Journal of Information Systems 20, 7–10.

Palvia, P., Leary, D., Mao, E., Midha, V., Pinjani, P., Salam, A.F., 2004. Research methodologies in MIS: an update. Communications of the Association for Information Systems (Volume 14, 2004) 526, 542.

Palvia, P., Mao, E., Salam, A.F., Soliman, K.S., 2003. Management Information System Research: What's There in a Methodology? Communications of the Association for Information Systems 11, 289–309.

PCAOB, 2010. Auditing Standard No. 12 Identifying and Assessing Risks of Material Misstatement.

Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V., Bragge, J., 2006. The design science research process: a model for producing and presenting information systems research, in: Proceedings of the 1st International Conference on Design Science Research in Information Systems and Technology (DESRIST), pp. 83–106.

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S., 2007. A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–77.

Process Mining Group, 2015. ProM [WWW Document]. URL http://www.processmining.org/prom/start (accessed 4.8.13).

Ramezani, E., Fahland, D., van der Aalst, W.M.P., 2012. Where Did I Misbehave? Diagnostic Information in Compliance Checking, in: Barros, A., Gal, A., Kindler, E. (Eds.), Business Process Management, Lecture Notes in Computer Science. Springer, pp. 262–278.

Recker, J., 2012. Scientific Research in Information Systems: a Beginner's Guide. Springer, New York.

Reichert, M., Weber, B., 2012. Enabling Flexibility in Process-aware Information Systems Challenges, Methods. Springer, Berlin; New York.

Riege, C., Saat, J., Bucher, T., 2009. Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik. Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik 69–86.

Romney, M.B., Steinbart, P.J., 2008. Accounting Information Systems, 11th Revised edition. ed. Prentice Hall.

Rowley, J., Slack, F., 2004. Conducting a Literature Review. Management Research News 27, 31–39.

Rozenberg, G., 1997. Handbook of Graph Grammars and Computing by Graph Transformation, illustrated edition. ed. World Scientific Pub Co, Singapore.

Rozinat, A., 2007. Towards an Evaluation Framework for Process Mining Algorithms (BPM Center Report). BPMcenter.org.

Rozinat, A., de Medeiros, A.K.A., Günther, C.W., Weijters, A., van der Aalst, W.M.P., 2008. The Need for a Process Mining Evaluation Framework in Research and Practice, in: Business Process Management Workshops. pp. 84–89.

Rozinat, A., van der Aalst, W.M.P., 2008. Conformance Checking of Processes Based on Monitoring Real Behavior. Information Systems 33, 64–95.

SAP, 2015. SAP-UCC [WWW Document]. URL http://www.sap-ucc.com/ (accessed 5.3.12).

Schauer, C., 2011. Die Wirtschaftsinformatik im internationalen Wettbewerb. Gabler, Wiesbaden.

Schimm, G., 2001a. Process Mining linearer Prozessmodelle - Ein Ansatz zur Automatisierten Akquisition von Prozesswissen, in: Proceedings 1. Konferenz Professionelles Wissensmanagement.

Schimm, G., 2001b. Process Mining Elektronischer Geschäftsprozesse, in: Proceedings Elektronische Geschäftsprozesse.

Schipper, K., 2005. The introduction of International Accounting Standards in Europe: Implications for International Convergence. European Accounting Review 14, 101–126.

Schira, J., 2009. Statistische Methoden der VWL und BWL Theorie und Praxis. Pearson Studium, München, Boston.

Schultz, M., 2015. Business Process Compliance from an Audit Perspective - A Design Science Research Approach for an Integrated Processing of Business Process Models and Internal Control in the Context of Process Audits. Hamburg.

Schultz, M., Müller-Wickop, N., Nüttgens, M., 2012. Key Information Requirements for Process Audits - an Expert Perspective, in: Proceedings of the 5th International Workshop on Enterprise Modelling and Information Systems Architectures, Vienna.

StataCorp LP, 2014. Stata | Data Analysis and Statistical Software [WWW Document]. URL http://www.stata.com/ (accessed 7.25.14).

Stocker, T., 2012. Data Flow-oriented Process Mining to Support Security Audits, in: Service-Oriented Computing-ICSOC 2011 Workshops. pp. 171–176.

Takeda, H., Veerkamp, P., Yoshikawa, H., 1990. Modeling Design Process. AI magazine 11, 37–48.

Tiwari, A., Turner, C.J., Majeed, B., 2008. A Review of Business Process Mining: State-of-the-art and Future Trends. Business Process Management Journal 14, 5–22.

Turban, E., Aronson, J.E., Liang, T.-P., Sharda, R., 2007. Decision Support and Business Intelligence Systems. Pearson Education International, Upper Saddle River, London.

United States Congress, 2012a. Securities Act of 1933.

United States Congress, 2012b. Securities Exchange Act of 1934.

United States Congress, 2002. Sarbanes-Oxley Act, Public Law 107–204.

University of Hamburg, 2015. Renew - The Reference Net Workshop [WWW Document]. URL http://www.renew.de/ (accessed 7.14.15).

University of Hamburg, 2014. Virtual Accounting Worlds [WWW Document]. URL http://www.virtualaccountingworlds.de/ (accessed 7.2.14).

Vaishnavi, V., Kuechler, W., 2004. Design Science Research in Information Systems [WWW Document]. URL http://www.desrist.org/design-research-in-information-systems/

van der Aalst, W.M.P., 2013. Business Process Management: A Comprehensive Survey. ISRN Software Engineering 2013.

van der Aalst, W.M.P., 2012. A Decade of Business Process Management Conferences: Personal Reflections on a Developing Discipline, in: Business Process Management. Springer, pp. 1–16.

van der Aalst, W.M.P., 2011a. Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st Edition. ed. Springer, Berlin Heidelberg.

van der Aalst, W.M.P., 2011b. Using Process Mining to Bridge the Gap Between BI and BPM. Computer 44, 77–80.

van der Aalst, W.M.P., 2011c. Process Mining: Discovering and Improving Spaghetti and Lasagna Processes, in: IEEE Symposium on Computational Intelligence and Data Mining (CIDM). pp. 1–7.

van der Aalst, W.M.P., 2005. Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing. Requirements Engineering Journal Vol. 10, pp. 198–211.

van der Aalst, W.M.P., Andriansyah, A., de Medeiros, A.K., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., 2012. Process Mining Manifesto, in: BPM 2011 Workshops Proceedings. pp. 169–194.

van der Aalst, W.M.P., de Medeiros, A.K.A., 2005. Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance. Electronic Notes in Theoretical Computer Science 121, 3–21.

van der Aalst, W.M.P., Hofstede, A.H.M. ter, Kiepuszewski, B., Barros, A.P., 2003. Workflow Patterns. Distributed and Parallel Databases 14, 5–51.

van der Aalst, W.M.P., Stahl, C., 2011. Modeling Business Processes : a Petri Net-oriented Approach. MIT Press, Cambridge, Mass.

van der Aalst, W.M.P., van Hee, K., van der Werf, J.M., Kumar, A., Verdonk, M., 2011. Conceptual Model for Online Auditing. Decision Support Systems 50, 636–647.

van der Aalst, W.M.P., Weijters, T., Maruster, L., 2004. Workflow Mining: Discovering Process Models from Event Logs. IEEE Transactions on Knowledge and Data Engineering 16, 1128–1142.

van der Werf, J.M.E.M., Verbeek, H.M.W., van der Aalst, W.M.P., 2012a. Context-Aware Compliance Checking, in: Barros, A., Gal, A., Kindler, E. (Eds.), Business Process Management, Lecture Notes in Computer Science. Springer, pp. 98–113.

van der Werf, J.M.E.M., Verbeek, H.M.W., van der Aalst, W.M.P., 2012b. Context-Aware Compliance Checking, in: Barros, A., Gal, A., Kindler, E. (Eds.), Business Process Management, Lecture Notes in Computer Science. Springer, pp. 98–113.

Venable, J., Pries-Heje, J., Baskerville, R., 2012. A Comprehensive Framework for Evaluation in Design Science Research. Design Science Research in Information Systems. Advances in Theory and Practice 423–438.

Venkatesh, V., Brown, S.A., Bala, H., 2013. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. MIS Quarterly 37, 21–54.

Verband der Hochschullehrer für Betriebswirtschaft e.V., 2011. VHB-Jourqual2.1.

Vercellis, C., 2009. Business Intelligence. Wiley, Chichester.

vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A., 2009. Reconstructing the Giant - on the Importance of Rigour in Documenting the Literature Search Process, in: Proceedings of the 17th European Conference on Information Systems, pp. 2–13.

Walls, J., Widmeyer, G., El Sawy, O., 1992. Building an Information System Design Theory for Vigilant EIS. Information Systems Research 3, 36–59.

Wang, G.A., Wang, H.J., Li, J., Fan, W., 2014. Mining Knowledge Sharing Processes in Online Discussion Forums, in: Proceedings of the 47th Hawaii International Conference on System Science, IEEE, pp. 3898–3907.

Webster, J., Watson, R.T., 2002. Analyzing the Past to Prepare for the Future. MIS Quarterly 26, xiii – xxiii.

Weijters, A., van der Aalst, W.M.P., de Medeiros, A.K.A., 2006. Process Mining with the Heuristics Miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP 166.

Werner, M., 2013. Colored Petri Nets for Integrating the Data Perspective in Process Audits, in: Proceedings of the 32nd International Conference on Conceptual Modeling (ER 2013), Springer-Verlag, Hong Kong, China, pp. 387–394.

Werner, M., 2012. Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung, in: Plate, G. (Ed.), Forschung Für Die Wirtschaft. Cuvillier Verlag, Göttingen, pp. 199–214.

Werner, M., Gehrke, N., 2015. Multilevel Process Mining for Financial Audits. IEEE Transactions on Services Computing 8, 820–832.

Werner, M., Gehrke, N., 2011. Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen, in: Plate, G. (Ed.), Forschung für die Wirtschaft. Shaker Verlag, Aachen, pp. 99–120.

Werner, M., Gehrke, N., Nüttgens, M., 2013. Towards Automated Analysis of Business Processes for Financial Audits, in: Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI 2013), Leipzig, pp. 375–389.

Werner, M., Gehrke, N., Nüttgens, M., 2012a. Business Process Mining and Reconstruction for Financial Audits, in: Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS 2012), Maui, pp. 5350–5359.

Werner, M., Nüttgens, M., 2014. Improving Structure - Logical Sequencing of Process Models, in: Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS 2014), Big Island, pp. 3888–3897.

Werner, M., Schultz, M., Müller-Wickop, N., Gehrke, N., Nüttgens, M., 2012b. Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits (Research in Progress), in: Proceedings of 33rd International Conference on Information Systems (ICIS 2012), Orlando, pp. 1–12.

Werner, M., Schultz, M., Müller-Wickop, N., Nüttgens, M., 2014. Who Is Afraid of the Big Bad Wolf - Structuring Large Design Science Research Projects, in: Proceedings of the 22nd European Conference on Information Systems (ECIS 2014), Tel Aviv, Israel, pp. 1–16.

Weske, M., 2012. Business Process Management Concepts, Languages, Architectures. Springer, Berlin; New York.

Wilde, T., Hess, T., 2007. Forschungsmethoden der Wirtschaftsinformatik. Wirtschaftsinformatik 49, 280–287.

Wilde, T., Hess, T., 2006. Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung. Arbeitspapiere des Instituts für Wirtschaftsinformatik und Neue Medien, LMU München 2.

WKWI, 2008. WI-Orientierungslisten. Wirtschaftsinformatik 50, 155–163.

Wollnik, M., 1988. Ein Referenzmodell des Informationsmanagements. Information Management 3, 34–43.

Workflow Management Coalition, 1999. Terminology & Glossary. Technical Report.

Yang, W.-S., Hwang, S.-Y., 2006. A Process-mining Framework for the Detection of Healthcare Fraud and Abuse. Expert Systems with Applications 31, 56–68.

yWorks GmbH, 2015. yEd - Graph Editor [WWW Document]. URL http://www.yworks.com/de/products_yed_about.html (accessed 7.15.15).

10 Appendix A: Publications

10.1 Who Is Afraid of the Big Bad Wolf - Structuring Large Design Science Research Projects

Number	**1**
Title	**Who Is Afraid of the Big Bad Wolf - Structuring Large Design Science Research Projects**
Appendix	10.1
Primary Related Chapters	3
Type	Conference Paper
Conference	22nd European Conference on Information Systems (ECIS 2014)
Reference	(Werner et al., 2014)
Acceptance Rate	34 %
VHB JQ 2.1 Ranking	B (7.37)
WKWI Ranking	A
ERA 2010	A
CORE 2013	A
Review Procedure	Double Blinded
Number of Reviews	4
Authors	1. Michael Werner 2. Martin Schultz 3. Niels Müller-Wickop 4. Markus Nüttgens
Dissertation Points	0.40
Authorship	
Overall	85%
Design	85%
Realization	85%
Writing	85%
Status	Published
Part of other Dissertations	No
Link	http://ecis2014.eu/E-poster/files/0690-file1.pdf

Who Is Afraid of the Big Bad Wolf - Structuring Large Design Science Research Projects

WERNER, MICHAEL

University of Hamburg, Germany
michael.werner@wiso.uni-hamburg.de

SCHULTZ, MARTIN

University of Hamburg, Germany
martin.schultz@wiso.uni-hamburg.de

MÜLLER-WICKOP, NIELS

University of Hamburg, Germany
niels.mueller-wickop@wiso.uni-hamburg.de

NÜTTGENS, MARKUS

University of Hamburg, Germany
markus.nuettgens@wiso.uni-hamburg.de

Abstract: Design Science Research is an important research approach that has gained increased attention over the past decade in the international scientific community. Recommendations and methodologies for design science-oriented research published in recent years provide good guidance for scientists to conduct and publish design science-oriented research. But little attention has yet been paid to the setup and structuring of large research projects. Such projects sometimes look like dangerous and greedy big bad wolves that threaten the innocent project participants. Large research projects are special because of the volume and complexity of the research scope and the interaction of participating project members. We look for solutions to tame the wolf and present a framework to confine and structure the content of large research projects into well-defined and individually manageable research segments. This framework is suitable for projects characterized by manifold research tasks and the involvement of many project participants. This article presents the designed framework and illustrates benefits, limitations and practical implications. Its application is illustrated by means of a research project case study that was carried out over a three year period with the participation of several research institutions and partner companies.

Keywords: Design Science, Research Process, Research Framework, Case Study

1 Introduction

Design Science Research (DSR) has gained increased attention over the last decade in the international scientific community as an important research approach (Vaishnavi and Kuechler, 2004). DSR is especially prevalent in German speaking countries (Wilde and Hess, 2007) where information systems research has traditionally been closely related to the natural and engineering sciences in contrast to research communities in Anglo-Saxon countries which show a tendency to positivist, behavioristic research methods (Schauer, 2011). Several publications on the role and interaction of design science, natural science (March and Smith, 1995) and social science (Gregor and Baskerville, 2012) have led to more clarity about the differences, relationships, and interactions between different research approaches in the scientific domain. Scientists have made valuable contributions on how to conduct DSR in a structured and rigorous manner (Hevner et al., 2004; Peffers et al., 2007; Österle et al., 2010; Hevner and Chatterjee, 2010) and for positioning DSR results in the academic arena (Gregor and Hevner, 2013). But little attention has yet been paid to the question of how large research projects can actually be structured and set up. Large research projects sometimes appear like hungry and dangerous wolves that threaten the frightened project participants because they do not know how to manage the voluminous research tasks appropriately to achieve the intended research goals. Projects get out of control, out of budget, out of time and consume valuable research resources without achieving the intended research objectives. The involved researchers feel eaten up, become frustrated and burn out. We describe a framework that can help to improve the research process by giving guidance as to how the research content and tasks can be structured and divided into manageable components.

Design science-oriented research in information systems research is defined by different steps (Peffers et al., 2007; Gregor and Baskerville, 2012) that commonly at least include the phases analysis, design, evaluation and diffusion (Österle et al., 2010). The completion of all phases is a resource- and time-consuming task. In comparison to purely descriptive science, DSR is characterized by the duality of the epistemological and the design objective (Riege et al., 2009). The objective of creating artifacts that are valuable for practical purposes (Hevner et al., 2004) requires the involvement and participation of project members from the application domain. Evaluation is an essential component in design science-oriented research projects (Riege et al., 2009; Venable et al., 2012). Many evaluation methods like simulations, lab or field experiments require instantiated artifacts that can be used in natural or artificial evaluation environments. The instantiation of designed artifacts is commonly a time-consuming task that requires different skills than the design phase. The necessity to include all necessary research phases and to conduct extensive evaluation for rigorous research results as well as the aim to create valuable research artifacts that on the one hand contribute to the scientific knowledge base but that can on the other hand also be applied in practice, commonly lead to large research projects that include the participation of multiple researchers, research assistants, different scientific institutions and partner companies from the application domain. The interaction of multiple researchers and participation of different interest groups contribute to the complexity of these projects.

Our research is guided by the question as to how the research content and tasks in large research projects can be structured to facilitate parallel research work by simultaneously encouraging mutual research efforts and synergies. We introduce a framework that can be used to structure large design science-oriented research projects. The framework allows the participants to divide the research content of a research project into small segments that can then be addressed individually to identify necessary research artifacts and appropriate research methods. We focus on design science-oriented research projects in the information systems science discipline because of the idiosyncratic characteristics of design science-oriented research. The duality of the epistemological and design objective in these projects leads to research activities that are different compared to other research approaches.

The overall objective of presenting our framework is to enable researchers to confine and structure the content of research projects in such a way that research segments are well separated from each other to allow distributed and parallel research work on different segments but also to enable the research group to keep track of the whole research process and to benefit from the interrelations of participating researchers and the mutual research efforts. The framework focusses on setting the scope of a research project and for dividing its content into small and well-defined parts that can be addressed by different researchers. It should not be regarded as a project management framework. Project management is a complex process and recommendations and guidance on project management is for example available in relevant national and international standards (International Organization for Standardization, 2012; Deutsches Institut für Normung, 2009; Project Management Institute, 2013; Great Britain and Office of Government Commerce, 2009). The presented framework should be seen as a useful tool in the setup and operational phase of a project to promote the assignment of research tasks to the involved researchers. Our research results are derived from a large design science-oriented research project that was carried out in cooperation with four private companies and two research institutions over a three year period. It included the participation of several senior and junior researchers, research assistants and representatives of the participating companies. The presented framework was developed and used during this research project. The application of the framework and the insights from its practical usage - which is included as a case study - illustrates the benefits and limitation of the presented framework.

2 Methodology and Research Structure

The work presented in this paper follows a DSR approach. DSR addresses research problems that originate from the application domain and its aim is to develop useful artifacts that can be applied in practice (Hevner et al., 2004). The application domain of our research work is the field of information systems research itself. We relied on a DSR approach in order to create an artifact that can improve the research process in large research projects and hence the proximity of the research question to practical problems in daily scientific practice. The presented research work follows the research phases described by Österle et al. which consist of analysis, design, evaluation and diffusion (Österle et al., 2010). The analysis of the problem domain is described in section four within the case study. The main focus of this paper is the presentation of the designed model for structuring and confining research projects which is addressed in section five. Section six continues with a discussion on lessons learned based on the described case study, identified benefits, limitations and evaluation aspects. The diffusion of our research results as the last phase in DSR is already addressed by the publication of this article. A major challenge in scientific endeavors is the selection of an adequate research method (Galliers and Land, 1987). A variety of well-established research methods is available for researchers in information systems science (Palvia et al., 2003; Palvia et al., 2004; Wilde and Hess, 2007). The selection of a research method should follow the intended research objective. The objective of our research lies in the development of a framework that can be used by scientists to structure and confine the research content in large research projects. We reviewed relevant frameworks in the information systems science discipline and amalgamated different models to construct a new framework useful for the purpose at hand. This approach is comparable to method engineering (Brinkkemper, 1996). Method engineering is used to construct methods based on already existing methods and method fragments. But instead of merging existing method fragments we used a model engineering approach by referring to model components that have already been proven useful in the information system research discipline for creating a novel solution.

Research projects are complex undertakings that include the interaction of many individuals with differing or sometimes even opposing motivations. The development of a framework for structuring research projects needs to consider the complex social settings in research projects. We used the developed framework in a real research project which is illustrated as a case study in this paper. Case study research is a common research method in information systems research (Chen and Hirschheim, 2004). It is a special form of qualitative-empirical research methods (Wilde and Hess, 2006) that involves the close examination of people, topics and issues (Hays, 2004). It is especially suited to investigate complex phenomena in their natural environments and can be used for behavioral or design-oriented research (Wilde and Hess, 2006). Case study research is commonly criticized for the lack of generalizability due to the uniqueness of the investigated case. But Yin points out that similar concerns can, for example, also be applied in the contexts of single experiments (Yin, 2008). Furthermore the objective of our research is of exploratory nature and not empirical evaluation.

3 Related Work

Scholars have highlighted the need for structured and commonly agreed research processes in the DSR community (Leist and Rosemann, 2011). Peffers et al. present a research methodology[28] for DSR in the information systems community (Peffers et al., 2006; Peffers et al., 2007). The authors present a research framework that consists of six phases: (1) identify problem and motivate, (2) define objectives of a solution, (3) design and development, (4) demonstration, (5) evaluation and (6) communication. Their framework is based on a review of existing scientific publications on the research process in the information systems and related research disciplines. The research framework is composed of process elements that have been identified by different scholars working in the information systems (Takeda et al., 1990; Nunamaker et al., 1991; Walls et al., 1992; Hevner et al., 2004; Cole et al., 2005) and engineering (Archer, 1984; Eekels and Roozenburg, 1991) discipline. It is interesting

[28] The term methodology is interpreted ambiguously in information systems science (Mingers, 2001). Peffers et al. refer to a methodology as a combination of methods independent of a single research project and intend to present a methodology that serves as a commonly accepted framework for carrying out design science research.

that the continental European research community is mostly neglected by the authors although the design science discipline has a long lasting tradition especially in German speaking countries (Winter, 2008). Österle et al. as representatives of this community suggest four phases for DSR: (1) analysis, (2) design, (3) evaluation and (4) diffusion (Österle et al., 2010). Gregor and Baskerville examine the research process from a philosophy of science perspective with the objective to provide a framework for the combination of design science and social science research. The presented research process consists of the phases (A) construct and test artefacts, (B) formulate prescriptive knowledge and theory, (C) study artefact(s) in use, (D) test knowledge of artefacts in use and (E) formulate descriptive knowledge (Gregor and Baskerville, 2012). All authors explicitly emphasize the iterative relationship between the different research steps in each model. Alturki et al. present a more detailed model that consists of 14 research steps. The authors also present an extensive summary of relevant literature (Alturki et al., 2011).

It is interesting to note that researchers from the information systems discipline rarely refer to project management literature. Available knowledge on project management that has been standardized in international (International Organization for Standardization, 2012) or national (Project Management Institute, 2013; Deutsches Institut für Normung, 2009; Great Britain and Office of Government Commerce, 2009) guidelines has rarely been considered yet, although research work exhibits all the characteristics that are also associated with projects (vom Brocke and Lippe, 2010). Vom Brocke and Lippe are among the few authors that build the bridge between research processes and project management. They point out the need to tailor existing project management guidelines for research projects and identify eight characteristics that distinguish design science-oriented research projects from traditional project types (vom Brocke and Lippe, 2010). We are not aware of quantitative research on project management in the field of information systems science.

The aforementioned literature provides useful insights into the research processes in design science-oriented research. But little has yet been published that provides guidance as to the effective structure and set up of large DSR projects. To do exactly that is what we intend with the presentation of the research results in this article.

4 Case Study

The research project that is described as a case study in this article was concerned with research questions from the field of financial audits. Companies prepare financial statements to provide interested parties with financially relevant information. The correctness and reliability of this information are a key requirement for stakeholders to direct their decisions. National laws and regulations mandate the audit of financial statements by an independent third party to prevent the distribution of false financial information because of its paramount role for the well-functioning of economic markets. These audits are carried out by public accountants. Accounting scandals in recent years have shown that auditors have not been able to prevent these scandals or at least indicate any violations before the actual collapse. A common problem in financial audits is an imbalance between automated transaction processing of partly huge data volumes on the companies' side and traditional and manual audit procedures on the auditors' side (Werner et al., 2012). Companies use information systems to support and automate the operation of their business processes. Auditors primarily rely on traditional audit procedures like interviews and manual inspections of available documents to achieve the necessary audit comfort. But these procedures become inefficient or even ineffective in environments where the processing of transactions is highly automated and includes the handling of large data volumes (Werner and Gehrke, 2011). A solution to decrease this imbalance would be the application of automated audit procedures. Business processes play a significant role in financial audits. The audit of business processes and internal controls that affect the processing of transactions are an important part of financial audits (International Federation of Accountants, 2012). The rationale for considering business processes is the assumption that well-controlled processes will lead to complete and correct entries on the financial accounts. The objective of the described research project was the development of methods and tools that support the auditor by automating parts of the procedures that are necessary to conduct process audits, and to thereby make these audits more efficient and effective. The fundamental idea was to use innovative data analysis techniques for automating the discovery of process models and to automatically assess the design and operating effectiveness of internal controls by analyzing relevant control data. The overall research question was formulated as follows:

- How can data analysis techniques be used to automate the audit of business processes in the context of financial audits?

Process mining provides powerful methods and tools that can be used to reconstruct process models based on the analysis of recorded event logs (van der Aalst, 2011). For designing the desired tools and methods the following more detailed research questions had to be answered.

(1) How can reliable process models be automatically reconstructed by analyzing data stored in information systems that process financially relevant transactions?

(2) How can process models be automatically assessed from an audit perspective by integrating control data that is stored in the source systems?

(3) How can process models be graphically represented to display information that is relevant to auditors and that can be applied in real audit environments?

These questions needed to be answered in order to develop research artifacts that are able to close the research gap and that are also valuable for the application domain. A main component of the project was the development of a software prototype. The design of the prototype was on the one hand desired for the creation of a valuable artifact for the application domain but also for evaluation purposes in the research process. The project members consisted of two research institutions and four partner organizations. A software company was responsible for the programming and instantiation of the software tool. A large auditing company supported the requirement analysis, design and testing phase and provided necessary data. A small auditing and consulting company was included to reflect the requirements from small and medium sized companies. A public association for board members contributed as a project partner to diffuse the research results into the broader application domain and by providing information about aspects required by the board level. The project members consisted of three PhD students and two professors from the information systems area, three software developers, several research assistants and about 20 contact persons from partner companies who were contacted during the research project. Concerns arose at the beginning of the research work about how the overall project should be structured. Each project participant had a different motive to take part on the project. The partner companies needed a software artifact for practical use. The junior researchers were eager to advance in their PhD studies and the senior researchers were concerned about resources and had to keep in mind the overall progress of the involved scientific institutions. There was the risk that research work on specific tasks would be conducted redundantly and other research tasks be neglected due to uncoordinated research activities and deviating motivations. A framework was necessary to identify which research outputs would be critical for the success of the research project, how these related to each other, which scientific approach would be adequate for the development and evaluation of research artifacts and which researcher would fit best to accomplish different research tasks according to available expertise and skills. It was also necessary to decide on the roles and responsibilities for the communication with the project partners for specific research aspects such as the requirement analysis, software development and evaluation approaches.

5 Segmentation Framework

This section deals with the description of the framework for structuring large design science-oriented research projects that was developed in the project mentioned above. The objective of the presented research was the development of a tool that allows the confining and separation of the research content and tasks. The first step to design a useful framework is the identification of distinguishing features that can be used to categorize different research contents. The main subjects of interest in information systems research are information technologies and the man/machine interaction. Gregor points out that the distinguishing characteristic of the information systems area is not only the consideration of both worlds - technology and humans - but also the investigation of phenomena that emerge from their interaction in socio-technical systems (Gregor, 2006). A framework for the structuring of research content should therefore take into account the technological and human-interaction aspects. Categorizations for different levels of information technology and human interaction can be found in various models. A common model from the field of information management is presented by Krcmar (Krcmar, 2010). He distinguishes between three different levels of management tasks. These are accompanied by independent leadership tasks which are relevant for all levels. Figure 1 shows a graphical representation of the model. The lowest level contains tasks for the management of the technical infrastructure that is necessary for the use of information and communication technology at the higher levels. The second level deals with the management of information systems and includes the management of data, processes, applications and their life-cycles. At the highest level reside the tasks for the management of the information economy. The main objective of the tasks at this level is the management of the resource information, its supply, demand and usage. The lower two levels are mainly concerned with the technical aspects of information management. The human aspect is considered at the highest level where the requirements for the lower levels are defined on the basis of the needs of human information recipients and users of the applications, processes, data and technology that is provided by the lower levels. The model of information management addresses both elements that are subject to research in information system science: information systems and human interaction in socio-technical systems. It is therefore a useful starting point for the categorization of research content because each researched artifact in design science-oriented research can be characterized if it addresses one or more of the different levels, infrastructure, applications and usage. The original model represents applications, data and processes at the same level.

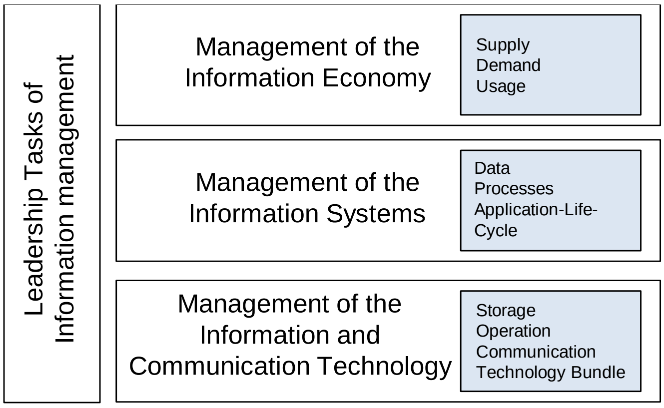


Figure 1 Model of Information Management (adapted from Krcmar 2010 p.50) [29]

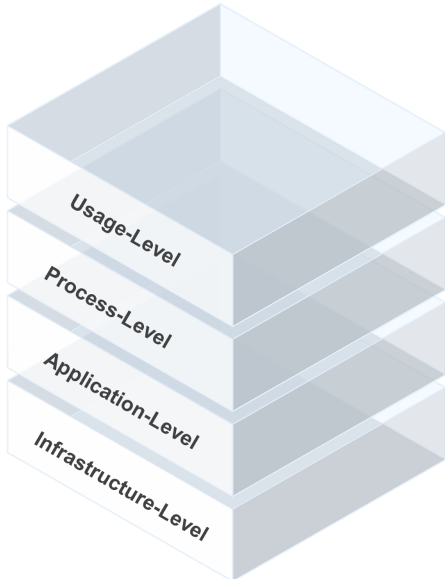


Figure 2 Human-Technical Dimension

Prominent conferences in the information systems science area like the Business Process Management conference (BPM, 2014), comprehensive publications (Weske, 2012) and extensive reviews (van der Aalst, 2012; van der Aalst, 2013) show that business processes are a key component in information systems research. We therefore considered it appropriate to divide the level of information systems into the two levels of software applications and processes. The process level is the connecting layer were process participants use components from the lower application level to satisfy information demands from the higher level. The requirement to include this additional level also became obvious in our research project because it was not sufficient to consider relevant software applications and how information was consumed at the usage-level but also how relevant information was created within business processes and how these relate to the financial statements that serve as information input for stakeholders at the highest level. The resulting four levels as shown in Figure 2 represent the first dimension for categorizing research content.

The presented levels are very broad. Although they can be used to distinguish research content on a technical vs. human-interaction dimension experiences from our case study showed that they are not sufficient to divide the research content into manageable segments. The usage of the four levels as the single separation criterion would mean that researchers would only be concerned about the assigned level which is not a suitable solution

[29] The model refers to the reference model that was originally introduced by Wollnik. The three levels of the information management model relate to the levels of information usage, information and communication systems and infrastructure of the information processing and communication in Wollnik's reference model (Wollnik, 1998).

because interdependencies between the different levels would be neglected. Furthermore it became obvious in our research project that such a broad separation was not adequate in practical settings when considering the motivation and skill sets of individual researchers.

Hevner et al. present a research framework for information systems research that provides an illustration how different concepts that are relevant for research projects relate to each other (Hevner et al., 2004). The framework is presented in Figure 3. It shows the relationships between the main research activities for design science (build and evaluate) and behavioral science (develop and justify) research, the environment and the knowledge base. The environment or application domain defines the problem space. The phenomena of interest for design science-oriented research should be derived from the environment. The knowledge base represents the already existing pool of research results that have been explored. Each research project should consider already existing knowledge to execute the research activities, provide useful artifacts to the application domain and add additional generalized knowledge to the knowledge base.

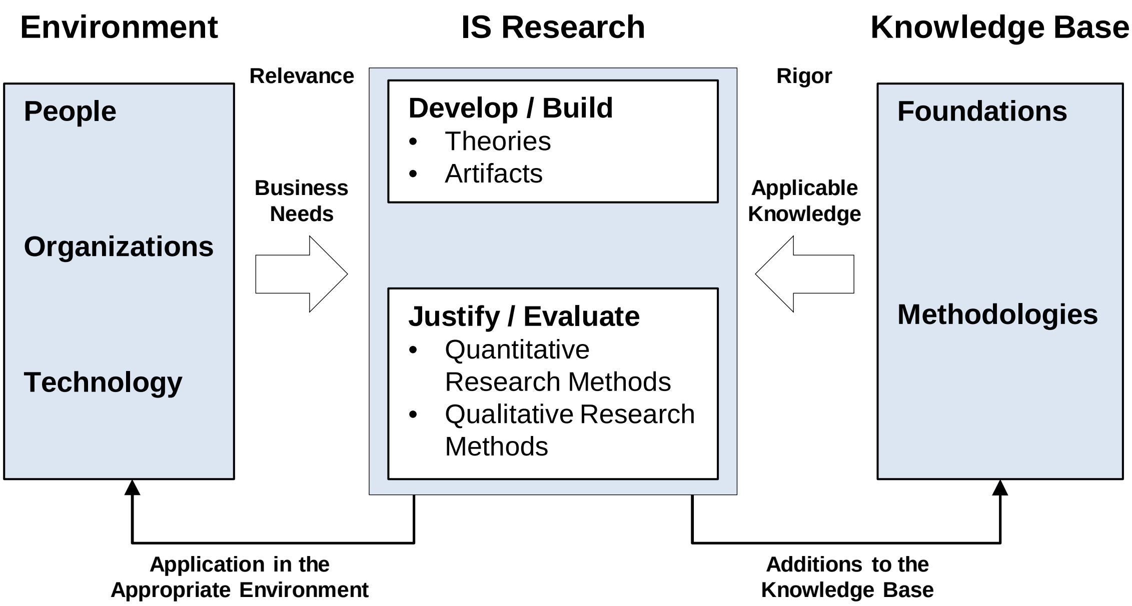


Figure 3 Information Systems Research Framework (adapted from Hevner et al. 2004 p. 80)

DSR projects are characterized by the duality of the epistemological and design objective which is illustrated in the model through the feedback loops from the research activities to the environment and the knowledge base. The distinction between the research contribution target domains can be used as a second dimension for the categorization of research content. Figure 4 shows the integration of the application domain and knowledge base as a second dimension.

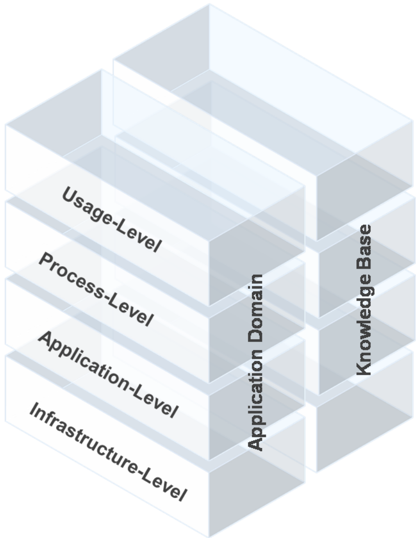


Figure 4 Integration of the Research Contribution Dimension

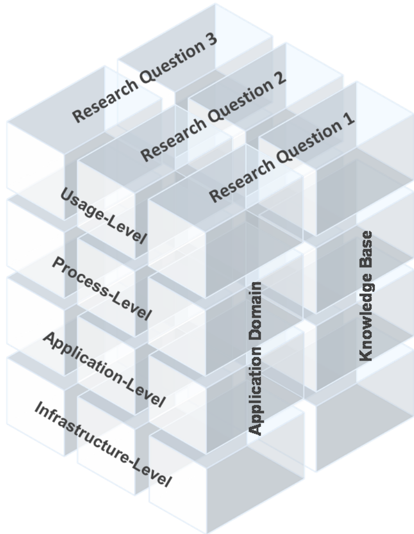


Figure 5 Integration of the Research Question Dimension

Each research project addresses an overall research question. The research questions in large research projects are, as a rule, complex. Otherwise it would be questionable if such a project indeed has the characteristics of a large project. Complex research questions can usually be divided into detailed lower-level research questions. These research questions can be used as a categorization criterion for a third dimension as shown in Figure 5. The application of all three dimensions as shown in Figure 5 illustrates how separate segments emerge based on the different dimension categories. Each segment can now be considered separately. And groups consisting of segments from different dimensions can be assigned to different researchers. This ensures that researchers do not just focus on an isolated segment but that they consider requirements from and interrelationships with different dimensions. The division into research segments can be seen as a 'divide and conquer' approach. The overall research project is broken into manageable components that can be conquered individually. Figure 6 illustrates a single segment of the overall model.

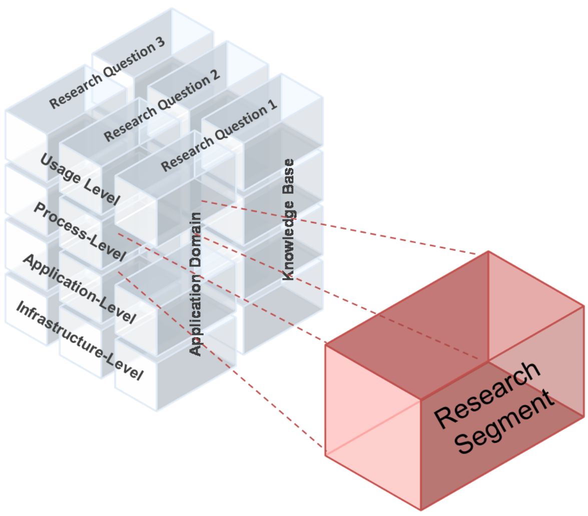


Figure. 6 Research Segment

Artifact
Research Methods
- Analysis
- Design
- Evaluation
Diffusion Type

Table 1 Segment Description

Artifacts in design science-oriented research should be developed by using suitable research methods. For each segment it is now possible to identify and describe the relevant artifact(s), the research methods for the analysis, design and evaluation as well as the diffusion type. A template for such a description is illustrated in Table 1. Evaluation methods for a single artifact can for example be chosen by relying on available frameworks (Venable et al., 2012), whereas the type of diffusion can for example be determined by referring to the knowledge contribution level of design science-oriented research (Gregor and Hevner, 2013).

Each segment should be associated with one (or more) artifact. March and Smith define four types of research artifacts: constructs, methods, models and instantiations (March and Smith, 1995). Gregor argues that design theories should also be regarded as an important outcome of design science-oriented research (Gregor, 2006). Many research methods exist for analysis, design and evaluation purposes. Table 2 illustrates exemplary research methods that have commonly been cited by renowned scholars. The superscripts a) to e) disclose the origin of the listed methods that are described in Footnote 30.

Artifacts	**Research Methods**	**Diffusion Types**
• Construct [a), e)] • Method [a), e)] • Model [a), e)] • Instantiation [a), e)] • Design theory [f)]	**Analysis** • Literature review [c)], survey [a)], expert interview [a), c)], case study [a)], data analysis [a)] **Design** • Modeling [a)] ,conceptual [c)] and reference modeling [a)], method engineering [a)], argument-based, concept-based and formal-deductive analysis [b)], case study [b), c)], prototyping [a), b)], qualitative analysis [b), c)], quantitative analysis [b)], action research [b)], lab experiment [b), c)], field experiment [b), c)], field study [c)], survey [c)], secondary data [c)] **Evaluation** • Action research [d)], focus groups [d)], case study [d)], participant observation [d)], ethnography [d)], survey [d)], mathematical or logical proof [d)], criteria-based evaluation [d)], lab experiment [a), d)], field experiment [a), d)], simulation [a), d)], pilot application [a)], expert reviews [a)]	• Conference presentation • Journal publication • Workshop presentation

Table 2 Overview of Research Artifacts, Methods and Diffusion Types[30]

The benefit of applying the segmentation framework becomes obvious when it is illustrated with examples. Figure 7 shows an instantiation of the reference framework for the described case study. The research question domain was separated into three categories that

[30] a) (Österle et al., 2010), b) (Wilde and Hess, 2007), c) (Palvia et al., 2003) d) (Venable et al., 2012), e) (March and Smith, 1995), f) (Gregor, 2006)

represent three different research questions. These were derived from the overall research question and objective (compare section 4). The primary application domain for the research project was financial audits. Business process management and business intelligence were perceived as being the most important scientific disciplines that form the knowledge base for conducting research in the project.

One segment in the framework is highlighted. It is located on the process-level because the objective of this segment was the creation of process models that are useful in the context of financial audits. The research task of this segment was the development of a mining algorithm that is able to reconstruct process models that comply with the requirements of the application domain. The designed algorithm exploits the specific structure of financially relevant transaction data to create process models (Gehrke and Müller-Wickop, 2010) and uses the data perspective to model the relationship between financial accounts and process activities (Werner, 2013). The requirements for developing this algorithm were derived from interviews with experts that showed that the data perspective is of utmost importance in the field of financial audits for illustrating the relationship between business processes and the financial accounts (Müller-Wickop et al., 2013). The technical requirements were investigated by analyzing the relevant data structure in ERP systems. The mining algorithm was designed by using components and research results from already existing mining algorithms following a method engineering approach (Brinkkemper, 1996).

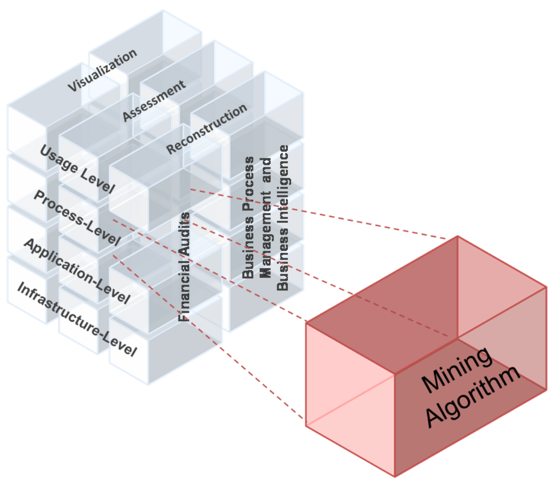


Figure 7 Example Research Segment

Artifact	Financial Process Mining Algorithm
Research Methods	
Analysis	Structured Interview Data Analysis
Design	Method Engineering Prototyping
Evaluation	Simulation Case Study
Diffusion Type	Journal

Table 3 Example Description

The mining algorithm was implemented as a prototype and evaluated by using test data for simulations (Werner et al., 2013). The evaluation also included a case study in a real world scenario. Table 3 summarizes the research methods and diffusion type for the designed artifact

Figure 8 shows a second example and illustrates two research segments. It demonstrates the relationship between segments that are related to the same level on the technical vs. human-interaction dimension but address different knowledge contribution categories on the knowledge contribution dimension. Table 4 lists the relevant artifacts and summarizes the research methods and diffusion types. A data extraction module was designed for the extraction of the event log data from the source ERP systems. The inspection of the extracted event logs revealed that they are not suitable for traditional mining algorithms. Financially relevant process instances do not exhibit strict linear control flows as assumed by traditional process mining algorithms, their execution behavior includes divergent and convergent behavior on the business process instance level (Werner and Nüttgens, 2014). It was necessary to develop a pre-processing algorithm that transforms non-linear event logs into linear event logs to be able to compare the mining results of the designed Financial Process Mining Algorithm with other mining algorithms (Mueller-Wickop and Schultz, 2013). The extraction module is an artifact relevant for the application domain because it allows the extraction of event log data for specific ERP systems in real world scenarios. The pre-processing algorithm is an artifact that is not specific to the application domain but can be applied in a variety of application scenarios for transforming non-linear event logs into linear event logs and can therefore be considered as a generalized contribution to the process mining knowledge base. Both artifacts needed to consider the ERP systems that are used in organizations for processing business transactions. The designed artifacts needed to be able to interact with these information systems and provided the input for the process mining algorithm described in the first example. It was therefore sensible to locate these artifacts on the application level of the human-technical dimension.

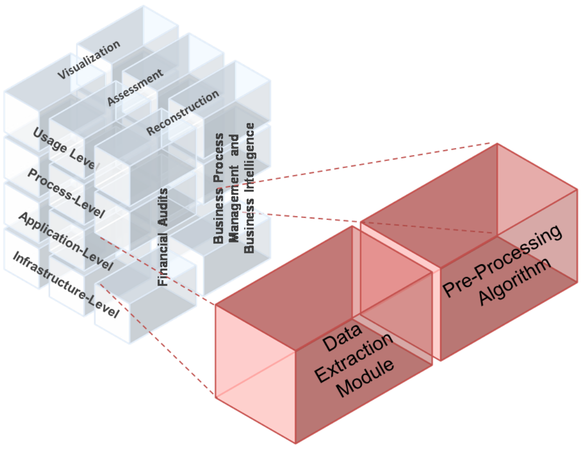


Figure 8 Example Multiple Research Segments

Artifact	Pre-Processing Algorithm
Research Methods	
Analysis	Experiment
Design	Method Engineering
Evaluation	Simulation
Diffusion Type	Conference

Artifact	Data Extraction Module
Research Methods	
Analysis	Data Analysis
Design	Prototyping
Evaluation	Case Study
Diffusion Type	Conference

Table 4 Example Description

The framework is further useful for the identification of dependencies between individual segments and the grouping of interdependent segments. Each group can be assigned to the involved research project participants based on motivational preferences and skill sets. Figure 9 shows the research scopes that were assigned to the three PhD students that were involved in the research project. The figure illustrates different aspects. First it is notable that not all possible segments are covered. This was intended and can be attributed to the research objective and scope of the research project. Research on the infrastructure-level was explicitly excluded from the project scope because it was not crucial for the achievement of the overall research objective. Not covered segments provide opportunities for additional and subsequent research. The illustration shows a second aspect. It can be seen that overlaps occurred in some segments. These overlaps were important. A basic requirement in large research projects is the clear separation of research tasks. But it is also important to have overlaps that create a common knowledge base that is fundamental for all research tasks, from preventing completely isolated research efforts and the emergence of 'Chinese Walls' within the project.

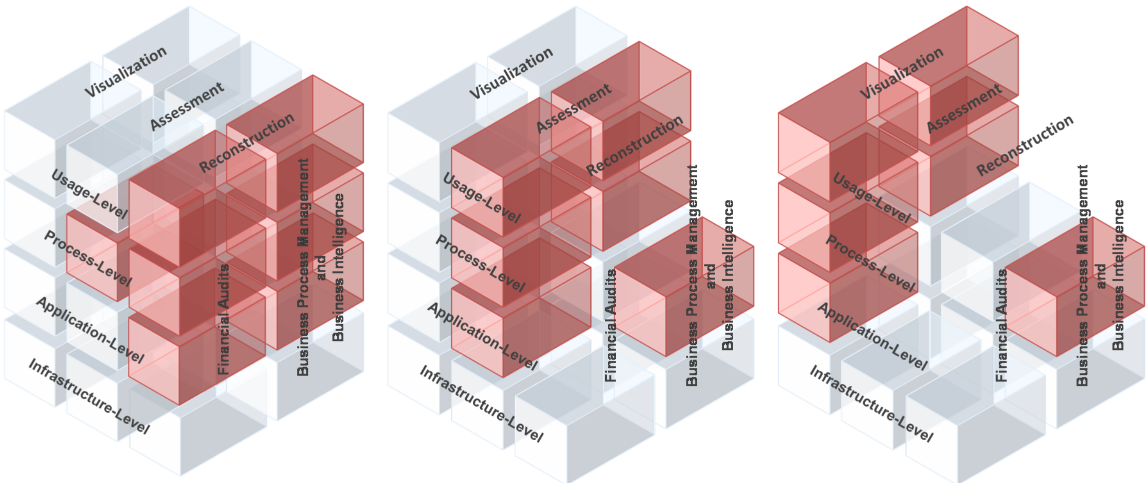


Figure 9 Research Scopes

A certain degree of overlap is necessary to encourage communication, mutual research effort and the creation of research synergies. Too much overlap leads to redundant and unproductive research work. Finding a balance between separation and overlap is difficult but crucial for the success of the research project.

6 Discussion

The previous section describes a framework for the structuring of large research projects. It was applied in a research project that is presented as a case study in this article. Its application proved to be useful for the described case. Not all research objectives that were identified in the initiation phase of the project could be achieved but the main research questions were answered with the design and evaluation of relevant research artifacts and their instantiations. The research results were published in 21 peer-reviewed publications and the project was finished on time and budget with an instantiated software prototype that included the majority of the developed methods. The provided framework played a significant role in the research process for the coordination of the research activities of the participating researchers. The application of the framework showed that several aspects are crucial for the successful application. A major obstacle was the development of a mutually agreed understanding of terms and definitions regarding the research artifacts and used research methods among all involved parties. The understanding of specific constructs deviated quite significantly between researchers and practitioners but also within the researcher group. The mutual research in specific research segments was beneficial to create a common agreement on fundamental concepts and terms. Another crucial success factor was the development of, agreement on and also implementation of the research and publication plans that can be developed on the basis of the applied framework. But if these plans are not followed strictly by all individuals the risk of rivalry and counterproductive activities increases. The implementation of designed artifacts in a software prototype was an important part of the project. A lack of commonly agreed documentation standards led to additional programming efforts when software implementations initially developed for one segment had to be used for another one. A suitable way to prevent such detrimental developments are mandatory continuous project meetings on a formal level for the discussion of the project progress but also on an informal level to encourage the mutual research efforts among different research segments and also to minimize communication barriers between the involved researchers. Dependencies between segments should be considered when planning the sequence of research activities and to identify synergies that can be realized if an analysis method for one segment can for example also be used for identifying requirements for another segment. A crucial outcome from our project was the insight that is counterproductive to assign segments to project participants that do not fit to the latters' motivation and skill profile of a participant. It is for example disadvantageous for the project progress to assign research tasks that require software implementation to project participants that lack relevant programming skills or the willingness to acquire these. (Brooks, 1987) stressed that a critical success factor in software development projects is the selection of top designers. A similar success factor for large DSR projects is the selection of project participants that fit to the identified research segments or segment groups in terms of motivation and skills.

Although we believe that the presented framework has the potential to facilitate the structuring and management of large research projects a major limitation is its limited evaluation. A case study has been chosen to demonstrate the applicability and usefulness of the presented framework. Such studies investigate cases for the purpose of illumination and understanding (Hays, 2004). But single case studies have the disadvantage that it is questionable if the creation of generalized knowledge is possible based just on one single observation. Extensive evaluation to strengthen the reliability of the created research results in the context of research processes is difficult due to the relative long run-time of research projects and the limited possibility to receive data for evaluation purposes. Although the limited evaluation might be seen as a constraint for the generalizability of the presented results it should be kept in mind that the purpose of this article is of exploratory nature. Our intention is to provide researchers a useful tool that assists in structuring research projects and therefore improving research processes and outcomes. The presented framework is a designed model that is not only applicable to an individual situation but that can be used as a reference model (Heinrich et al., 2004) to structure the content of any DSR project. The model can be seen as an improvement of existing tools for project management in the scientific area. Like in software engineering there is no single solution that fits all possible project scenarios. Defining and grouping research tasks and assigning them to suitable project participants is a big challenge. The provided framework can be used as a tool to facilitate this task.

7 Summary and Outlook

DSR is an important research approach in the international information systems science community. A variety of publications exist that provide guidance on how to conduct design science-oriented research. DSR differs from social science research due to its normative nature and the duality of the research objective. DSR does not only aim at generating generalized knowledge as a contribution to the scientific knowledge base but also intends to develop artifacts that are useful for the application domain. This duality leads to research activities like the instantiation of designed artifacts and evaluation types that are idiosyncratic to DSR compared to other research approaches. Large DSR projects can appear like dangerous wolves hungry to eat up the helpless project participants and research resources. This paper presents a framework that can be used to structure the content of research projects by dividing the overall research tasks into manageable segments and thereby tames the wolves. The use of the framework allows the coordination of parallel and mutual research work of the participating researchers necessary to conduct large and complex research projects successfully. The applicability of the framework and the benefits that can be gained by its application have been described by means of a case study. Research on the management of research projects in information systems science is still scarce. Further research efforts will be made to evaluate the presented framework in future research. Many aspects of project management in the information systems science discipline, e.g. concerning the portfolio management of research projects have not been investigated yet and empirical research is almost absent. We hope that this field of research will be addressed more closely also by other researchers.

8 References

Van der Aalst, W.M. (2012). A decade of business process management conferences: personal reflections on a developing discipline. In Business Process Management. Springer, pp. 1–16.

Van der Aalst, W.M. (2013). Business Process Management: A Comprehensive Survey. ISRN Software Engineering, 2013.

Van der Aalst, W.M.P. (2011). Process Mining: Discovery, Conformance and Enhancement of Business Processes. 1st Edition. Berlin Heidelberg: Springer.

Alturki, A., Gable, G.G. and Bandara, W. (2011). A design science research roadmap. In Service-Oriented Perspectives in Design Science Research. Springer, pp. 107–123.

Archer, L.B. (1984). Systematic method for designers. In Developments in Design Methodology. London: John Wiley, pp. 57–82.

BPM. (2014). BPM 2014. Available from: http://bpm2014.haifa.ac.il/

Brinkkemper, S. (1996). Method engineering: engineering of information systems development methods and tools. Information and Software Technology, 38(4), pp.275–280.

Vom Brocke, J. and Lippe, S. (2010). Taking a project management perspective on design science research. In Global Perspectives on Design Science Research. Springer, pp. 31–44.

Brooks. (1987). No Silver Bullet Essence and Accidents of Software Engineering. Computer, 20(4), pp.10–19.

Chen, W. and Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. Information Systems Journal, 14(3), pp.197–235.

Cole, R. et al. (2005). Being Proactive: Where Action Research Meets Design Research. In Proceedings of the 26h International Conference on Information Systems. ICIS.

Deutsches Institut für Normung. (2009). DIN SPEC 69901 - Projektmanagement. Berlin: Beuth Verlag GmbH.

Eekels, J. and Roozenburg, N.F.M. (1991). A methodological comparison of the structures of scientific research and engineering design: Their similarities and differences. Design Studies, 12(4), pp.197–203.

Galliers, R.D. and Land, F.F. (1987). Viewpoint: choosing appropriate information systems research methodologies. Communications of the ACM, 30(11), pp.901–902.

Gehrke, N. and Müller-Wickop, N. (2010). Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems. In Proceedings of the 16th Americas Conference on Information Systems. Americas Conference on Information Systems. Lima, Peru.

Great Britain and Office of Government Commerce. (2009). Managing successful projects with PRINCE2. London: TSO.

Gregor, S. (2006). The nature of theory in information systems. MIS Quarterly, 30(3), pp.611–642.

Gregor, S. and Baskerville, R. (2012). The Fusion of Design Science and Social Science Research. ISF 2012.

Gregor, S. and Hevner, A.R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, 37(2), pp.337–355.

Hays, P.A. (2004). Case study research. In Foundations for research methods of inquiry in education and the social sciences. Mahwah, N.J.: L. Erlbaum Associates, pp. 217–234.

Heinrich, L.J., Heinzl, A. and Roithmayr, F. (2004). Wirtschaftsinformatik-Lexikon. München: Oldenbourg.

Hevner, A. and Chatterjee, S. (2010). Design Science Research in Information Systems. In Design Research in Information Systems. Boston, MA: Springer US, pp. 9–22.

Hevner, A.R. et al. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), pp.75–105.

International Federation of Accountants. (2012). ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment.

International Organization for Standardization. (2012). ISO 21500 2011 Guidance on project management.

Krcmar, H. (2010). Informationsmanagement. Berlin; Heidelberg: Springer.

Leist, S. and Rosemann, M. (2011). Research process management. In Proceedings of the 22nd Australasian Conference on Information Systems (ACIS) 2011-Identifying the Information Systems Discipline. pp. 1–11.

March, S.T. and Smith, G.F. (1995). Design and natural science research on information technology. Decision support systems, 15(4), pp.251–266.

Mingers, J. (2001). Combining IS research methods: towards a pluralist methodology. Information systems research, 12(3), pp.240–259.

Mueller-Wickop, N. and Schultz, M. (2013). ERP Event Log Preprocessing: Timestamps vs. Accounting Logic. In Design Science at the Intersection of Physical and Virtual Design. 8th International Conference on Design Science Research in Information Systems and Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 105–119.

Müller-Wickop, N., Schultz, M. and Peris, M. (2013). Towards Key Concepts for Process Audits – A Multi-Method Research Approach. In Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics. ICESAL. Utrecht.

Nunamaker, J.F., Chen, M. and Purdin, T.D.M. (1991). Systems Development in Information Systems Research. Journal of Management Information Systems, 7(3), pp.89–106.

Österle, H. et al. (2010). Memorandum on design-oriented information systems research. European Journal of Information Systems, 20(1), pp.7–10.

Palvia, P. et al. (2003). Management Information System Research: What's There in a Methodolgy? Communications of the Association for Information Systems, 11, pp.289–309.

Palvia, P. et al. (2004). Research methodologies in MIS: an update. Communications of the Association for Information Systems (Volume 14, 2004), 526(542), p.542.

Peffers, K. et al. (2007). A Design Science Research Methodology for Information Systems Research. J. Manage. Inf. Syst., 24(3), pp.45–77.

Peffers, K. et al. (2006). The design science research process: a model for producing and presenting information systems research. In Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006). pp. 83–106.

Project Management Institute. (2013). A guide to the project management body of knowledge: (PMBOK Guide). Newtown Square: Project Management Institute.

Riege, C., Saat, J. and Bucher, T. (2009). Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik. Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik, pp.69–86.

Schauer, C. (2011). Die Wirtschaftsinformatik im internationalen Wettbewerb. Wiesbaden: Gabler.

Takeda, H., Veerkamp, P. and Yoshikawa, H. (1990). Modeling design process. AI magazine, 11(4), p.37.

Vaishnavi, V. and Kuechler, W. (2004). Design Science Research in Information Systems. [online]. Available from: http://www.desrist.org/design-research-in-information-systems/.

Venable, J., Pries-Heje, J. and Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. Design Science Research in Information Systems. Advances in Theory and Practice, pp.423–438.

Walls, J., Widmeyer, G. and El Sawy, O. (1992). Building an information system design theory for vigilant EIS. Information Systems Research, 3(1), pp.36–59.

Werner, M. (2013). Colored Petri Nets for Integrating the Data Perspective in Process Audits. In Proceedings of 32nd International Conference on Conceptual Modeling (ER 2013). 32nd International Conference on Conceptual Modeling (ER 2013). Hong Kong, China: Springer-Verlag, pp. 387–394.

Werner, M. and Gehrke, N. (2011). Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen. In G. Plate, ed. Forschung für die Wirtschaft. Aachen: Shaker Verlag.

Werner, M., Gehrke, N. and Nüttgens, M. (2012). Business Process Mining and Reconstruction for Financial Audits. In Proceedings of the 45th Hawaii International Conference on System Sciences. 45th Hawaii International Conference on System Sciences. Maui, pp. 5350–5359.

Werner, M., Gehrke, N. and Nüttgens, M. (2013). Towards Automated Analysis of Business Processes for Financial Audits. In Proceedings of the 11th International Conference on Wirtschaftsinformatik. Internationale Tagung Wirtschaftsinformatik. Leipzig.

Werner, M. and Nüttgens, M. (2014). Improving Structure - Logical Sequencing of Process Models. In Proceedings of the 47th Hawaii International Conference on System Sciences. 47th Hawaii International Conference on System Sciences. Bis Island.

Weske, M. (2012). Business process management concepts, languages, architectures. Berlin; New York: Springer.

Wilde, T. and Hess, T. (2007). Forschungsmethoden der Wirtschaftsinformatik. Wirtschaftsinformatik, 49(4), pp.280–287.

Wilde, T. and Hess, T. (2006). Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung. Arbeitspapiere des Instituts für Wirtschaftsinformatik und Neue Medien, LMU München, 2.

Winter, R. (2008). Design science research in Europe. European Journal of Information Systems, 17(5), pp.470–475.

Wollnik, M. (1998). Ein Referenzmodell des Informationsmanagements. Information Management, 3(3).

Yin, R.K. (2008). Case Study Research: Design and Methods. Fourth Edition. Sage Publications.

10.2 Process Mining

Number	**2**
Title	**Process Mining**
Appendix	10.2
Primary Related Chapters	4.1
Type	Journal Paper
Journal	wisu - das wirtschaftsstudium
Reference	(Gehrke and Werner, 2013)
Acceptance Rate	[31]
VHB JQ 2.1 Ranking	E (2.86)
WKWI Ranking	-
ERA 2010	-
CORE 2013	-
Review Procedure	Blinded
Number of Reviews	-
Authors	1. Nick Gehrke 2. Michael Werner
Dissertation Points	0.67
Authorship	
Overall	95 %
Design	95 %
Realization	95 %
Writing	95 %
Status	Published
Part of other Dissertations	No
Link	http://www.wisu.de/

[31] Invited for publication

Process Mining

PROF. DR. NICK GEHRKE / DIPL.-WIRT.-INF. MICHAEL WERNER

Elmshorn

Abstract: The increasing integration of information systems for the operation of business processes provides the basis for innovative data analysis approaches. Information systems support or even automate the execution of business transactions in modern companies. Business intelligence aims to support and improve decision making processes by providing methods and tools for analyzing data. Process mining builds the bridge between data mining as a business intelligence approach and business process management. Its primary objective is the discovery of process models based on available event log data. The discovered process models can be used for a variety of analysis purposes.

1 Introduction

New opportunities for data analysis

Companies use information systems to enhance the processing of their business transactions. **Enterprise resource planning (ERP)** and **workflow management systems (WFM)** are the predominant information system types that are used to support and automate the execution of business processes. Business processes like procurement, operations, logistics, sales and human resources can hardly be imagined without the integration of information systems that support and monitor relevant activities in modern companies. The increasing integration of information systems does not only provide the means to increase effectiveness and efficiency. It also opens up new possibilities of data access and analysis. When information systems are used for supporting and automating the processing of business transactions they generate data. This data can be used for improving business decisions.

Business intelligence approaches

The application of techniques and tools for generating information from digital data is called **business intelligence (BI)**. Prominent BI approaches are **online analytical processing (OLAP)** and **data mining** (Kemper et al. 2010 pp. 1-5). OLAP tools allow analyzing multidimensional data using operators like roll-up and drill-down, slice and dice or split and merge (Kemper et al. 2010 pp. 99-106). Data mining is primarily used for discovering patterns in large data sets (Kemper et al. 2010 p. 113).

Blessing and curse

The availability of data as a new source of information is not only a blessing, it can also become a curse. The phenomena of **information overflow** (Krcmar 2010 pp. 54-57), **data explosion** (Van der Aalst 2011 pp. 1-3) and **big data** (Chen et al. 2012) illustrate several problems that arise from the availability of enormous amounts of data. Humans are only able to handle a certain amount of information in a given time frame. When more and more data is available, how can it actually be used in a meaningful manner without overstraining the human recipient?

Aim of process mining

Data mining is the analysis of data for finding relationships and patterns. The patterns are an abstraction of the analyzed data. Abstraction reduces complexity and makes information available for the recipient. The aim of process mining is the extraction of information about business processes (Van der Aalst 2011 p. 1). Process mining encompasses "techniques, tools and methods to discover, monitor and improve real processes (…) by extracting knowledge from event logs (…)" (Van der Aalst et al. 2012 p. 15).

The data that is generated during the execution of business processes in information systems is used for reconstructing process models. These models are useful for analyzing and optimizing processes. Process mining is an innovative approach and a bridge between data mining and business process management.

Process mining research

Process mining evolved in the context of **analyzing software engineering processes** by Cook/Wolf in the late 1990s (Cook/Wolf 1998). Agrawal/Gunopulos (Agrawal et al. 1998) and Herbst/Karagiannis (Herbst/Karagiannis 1998) introduced process mining to the context of workflow management. Major contributions to the field have been added during the last decade by van der Aalst and others by developing mature mining algorithms and addressing a variety of topic related challenges (Van der Aalst 2011). This has led to a well-developed set of methods and tools that are available for scientists and practitioners.

Question 1: Why is process mining important? Why can it be seen as a bridge between data mining and business process management?

2 Process Mining Basics

2.1 Process Models and Event Logs

Graphical representations of business processes

The aim of process mining is the construction of process models based on available logging data. In the context of information system science a model is an immaterial representation of its real world counterpart used for a specific purpose (Becker/Probandt et al. 2012 pp. 1-3). Models can be used to reduce complexity by representing characteristics of interest and by omitting other characteristics. A process model is a graphical representation of a business process that describes the dependencies between activities that need to be executed collectively for realizing a specific business objective. It consists of a set of activity models and constraints between them (Weske 2012 p. 7).

Different modeling languages

Process models can be represented in different process modeling languages for example using the **Business Process Model and Notation (BPMN)**, **Event Driven Process Chains (EPC)** or **Petri Nets**. Petri Nets represent the dominant modeling language in the field of process mining (Tiwari et al. 2008). While the formal expressiveness of the Petri Net language is strong it is less suitable for addressees that are not familiar with its syntax and semantic. BPMN provides more intuitive semantics that are easier to understand for recipients that do not possess a theoretical background in informatics. We therefore rely on BPMN models for illustration in this article.

Figure 1 shows a business process model of a si**mple purchasing process** (we only use a subset of basic BPMN elements and especially do not include participants, data or artifacts for simplification and easier understandability). It starts with the ordering of goods. At some point of time the ordered goods get delivered. After the goods have been received an invoice is issued by the supplier that is finally paid by the company that ordered the goods.

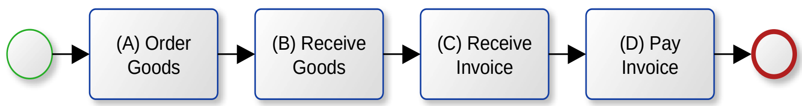


Fig. 1: Ideal Purchasing Process

The illustrated process model was created manually. So we do not know if the model actually reflects reality. There might be occurrences for example that invoices are paid before the goods and invoices have been delivered. For ordered services there might even be no such step as a recorded delivery. The question arises: How can we get reliable information about the real execution of the business process?

Case ID	Event ID	Timestamp	Activity
1	1000	01.01.2013	Order Goods
	1001	10.01.2013	Receive Goods
	1002	13.01.2013	Receive Invoice
	1003	20.01.2013	Pay Invoice
2	1004	02.01.2013	Order Goods
	1005	01.01.2013	Receive Goods
	…	…	…

Fig. 2: Event Log Structure

The approach used in process mining for answering this question bases on the exploitation of data stored in information systems that is created during the processing of business transactions. An information system stores data in log files or database tables when processing transactions. In the case of issuing an order data about the type and quantity of ordered goods, preferred suppliers, time of ordering etc. gets recorded. The stored data can be extracted from the information system and be made available in so called event logs. They constitute the data basis for process mining algorithms.

Cases, events and attributes

An **event log** is basically a table. It contains all recorded events that relate to executed business activities. Each event is mapped to a case. A **process model** is an abstraction of the real world execution of a business process. A single execution of a business process is called **process instance**. They are reflected in the event log as a set of events that are mapped to the same case. The sequence of recorded events in a case is called **trace**. The model that describes the execution of a single process instance is called **process instance model**. A process model abstracts from the single behavior of process instances and provides a model that reflects the behavior of all instances that belong to the same process. Cases and events are characterized by classifiers and attributes. **Classifiers** ensure the distinctness of cases and events by mapping unique names to each case and event. Attributes store additional information that can be used for analysis purposes. An example of an event log is given in Figure 2.

Question 2: What is the difference between event log, case, event, and trace? How do they relate to each other?

2.2 Mining Procedure

Source information systems

Figure 3 provides an overview of the different process mining activities. Before being able to apply any process mining technique it is necessary to have **access to the data**. It needs to be extracted from the relevant information systems. This step is far from trivial. Depending on the type of source system the relevant data can be distributed over different database tables. Data entries might need to be composed in a meaningful manner for the extraction. Another obstacle is the **amount of data**. Depending on the objective of the process mining up to millions of data entries might need to be extracted which requires efficient extraction methods. A further important aspect is **confidentiality**. Extracted data might include personalized information and depending on legal requirements anonymization or pseudonymization might be necessary.

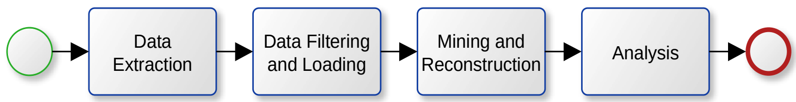


Fig. 3: Mining Procedure

Data filtering

Before the extracted event log can be used it needs to be filtered and loaded into the process mining software. There are different reasons why filtering is necessary. Information systems are **not free of errors**. Data may be recorded that does not reflect real activities. Errors can result from malfunctioning programs but also from user disruption or hardware failures that leads to erroneous records in the event log. Another error source occurs without incorrect processing. A specific process is normally analyzed for a certain time frame. When the data is extracted from the source system process instances can get truncated that were executed over the boundaries of the selected time frame. They need to be deleted from the event log or extracted completely. Otherwise they lead to erroneous results in the reconstructed process models. Event logs commonly do not exclusively contain data for a single process. Filtering is necessary to curtail the event log in a way that it only contains events that belong to the scrutinized process. Such a filtering needs to be conducted carefully because it can lead to truncated process instances as well. A common criterion is the selection of activities that are known to belong to the same process. Data filtering and loading is commonly supported by software tools and performed in a single step. But it can also be done separately.

Process models

Once the data is loaded into the process mining software the actual mining and reconstruction of the process model can take place. The mining includes the discovery of relationships in the event log whereas the reconstruction produces a process model as a graphical representation. The mining and reconstruction are commonly provided by the same software tool in a single step.

Analysis purposes

When the process models are mined and reconstructed they can be used for the intended purpose. We summarize this step with the **term analysis**. A fundamental goal of process mining is the discovery of formerly unknown processes. In this case the reconstruction is the aim itself but not limited to it. The analysis might aim at additional objectives like identifying opportunities for process optimization, organizational aspects or conformance and compliance analysis.

2.3 Mining Algorithms

The main component in process mining is the mining algorithm. It determines how the process models are created. A broad variety of mining algorithms does exist. The following three categories will be discussed in more detail:

— Deterministic mining algorithms

— Heuristic mining algorithms

— Genetic mining algorithms

Deterministic mining algorithms produce defined and reproducible results

Determinism means that an algorithm only produces defined and reproducible results. It always delivers the same result for the same input. A representative of this category is the **α-Algorithm** (Van der Aalst et al. 2002). It was one of the first algorithms that are able to deal with concurrency. It takes an event log as input and calculates the ordering relation of the events contained in the log.

Heuristic mining algorithms take frequencies into account

Heuristic mining also uses deterministic algorithms but they incorporate frequencies of events and traces for reconstructing a process model. A common problem in process mining is the fact that real processes are highly complex and their discovery leads to complex models. This complexity can be reduced by disregarding infrequent paths in the models.

Genetic mining algorithms mimic natural evolution

Genetic mining algorithms use an evolutionary approach that mimics the process of natural evolution. They are not deterministic. Genetic mining algorithms follow four steps: **initialization**, **selection**, **reproduction** and **termination**. The idea behind these algorithms is to generate a random population of process models and to find a satisfactory solution by iteratively selecting individuals and reproducing them by crossover and mutation over different generations. The initial population of process models is generated randomly and might have little in common with the event log. But due to the high number of models in the population, selection and reproduction better fitting models are created in each generation.

Case ID	Event ID	Timestamp	Activity	Resource
1	1000	01.01.2013	(A) Order Goods	Peter
	1001	10.01.2013	(B) Receive Goods	Michael
	1002	13.01.2013	(C) Receive Invoice	Frank
	1003	20.01.2013	(D) Pay Invoice	Tanja
2	1004	02.01.2013	(A) Order Goods	Peter
	1005	03.02.2013	(B) Receive Goods	Michael
	1006	05.02.2013	(C) Receive Invoice	Frank
	1007	06.02.2013	(D) Pay Invoice	Tanja
3	1008	01.01.2013	(A) Order Goods	Louise
	1009	04.01.2013	(C) Receive Invoice	Frank
	1010	05.01.2013	(B) Receive Goods	Michael
	1011	10.01.2013	(D) Pay Invoice	Tanja
4	1016	15.01.2013	(A) Order Goods	Peter
	1017	20.01.2013	(C) Receive Invoice	Claire
	1018	25.01.2013	(D) Pay Invoice	Frank
5	1023	01.01.2013	(A) Order Goods	Michael
	1024	10.01.2013	(B) Receive Goods	Michael
	1025	13.01.2013	(C) Receive Invoice	Michael
	1026	20.01.2013	(D) Pay Invoice	Michael

Fig. 4: Sample Event Log

The outcomes of mining differ depending on the used algorithm. We use the event log displayed in Figure 4 to illustrate the models created by different mining algorithms and for getting an impression how the mining works.

The event log contains a very limited number of events and cases. But it is nevertheless suitable to demonstrate key aspects that are relevant for process mining and the selection of mining algorithms.

Figure 5 shows a mined process model that was reconstructed by applying the α-Algorithm to the sample event log. It was translated into a BPMN model for better comparability. Obviously this model is not the same as the model in Figure 1. The reason for this is that the mined event log includes cases that deviate from the ideal linear process execution that was assumed for modeling in Figure 1. In case 3 the invoice is received before the goods. Due to the fact that both possibilities are included in the event log (goods received before the invoice in case 1, 2, 5 and invoice received before the ordered goods in case 3) the mining algorithm assumes that these activities can be carried out concurrently.

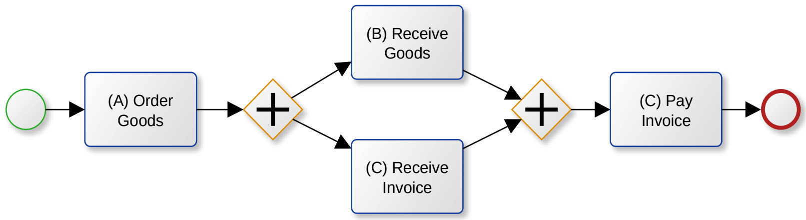


Fig. 5: Mined Process Model Using the α-Algorithm (the original model was generated using the ProM software)

When we look at the model in Figure 5 a little bit closer we can see another important fact. Actually case 4 is not reflected in the process model. No execution sequence in the model is able to reproduce the trace of case 4. It only allows two possible execution sequences: ABCD and ACBD put not ACD. The end-gateway after the "Order Goods" activity requires that both following branches are executed. Therefore it is not possible to have an execution sequence without the execution of activity B. The model has a poor fitness because it is not able to reflect all relevant traces in the event log. The model shown in Figure 6 is able to replay all traces. Due to the exclusive-gateways all three traces ABCD, ACBD and ACD are possible. But now the problem occurs that there are much more execution sequences possible than reflected in the event log. In fact the process model allows for an infinite set of sequences. Now loops are possible either starting from B to C or from C to B leading to possible sequences with infinite iterations of B and C or C and B. The sequence ABCBCD would for example be possible although it is not included as a trace in the event log. If a process model is too general it is called under-fitting. It has a poor precision. A major challenge in process mining is finding an **adequate solution between fitness, precision, simplicity, and generalizability**.

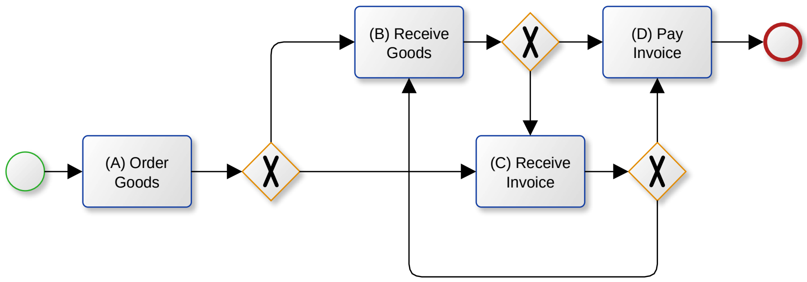


Fig. 6: Under-fitting Process Model

Fuzzy miner algorithm

Various advanced mining algorithms do exist that can be used for different purposes. Figure 7 illustrates the mined model using the heuristic fuzzy miner algorithm (Günther / Van der Aalst 2007). The model does not follow the BPMN notation, instead it uses a dependency graph representation. It does not contain any gateway operators but shows the dependencies between different activities. The dependency graph illustrates for example that A was followed three times by B and two times by C.

Question 3: Why do different process mining algorithms produce different process models?

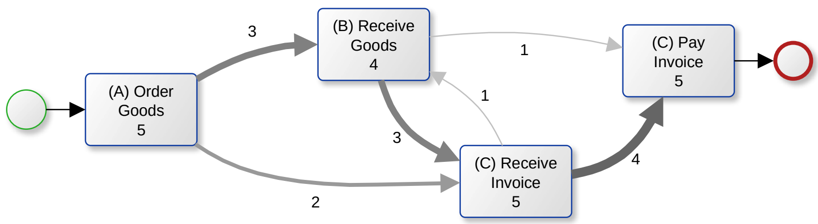


Fig. 7: Mined Process Model Using the Fuzzy Miner Algorithm (the original model was generated using the Disco process mining software)

It is important to identify which requirements need to be considered for achieving the intended objectives for each individual process mining project. The appropriateness of an algorithm should be evaluated depending on the area of application.

3 Application Areas

3.1 Process Discovery and Enhancement

New opportunities for analysis and optimization

A major area of application for process mining is the **discovery** of formerly unknown process models for the purpose of **analysis or optimization** (Van der Aalst et al. 2012 p. 13). Business process reengineering and the implementation of ERP systems in organizations gained strong attention starting in the 1990s. Practitioners have since primarily focused on designing and implementing processes and getting them to work. With maturing integration of information systems into the execution of business processes and the evolution of new technical possibilities the focus shifts to analysis and optimization.

Data stored in information systems is more reliable than manually collected data

Actual executions of business processes can now be described and be made explicit. The discovered processes can be analyzed for performance indicators like average processing time or costs for improving or reengineering the process. The major advantage of process mining is the fact that it uses reliable data. The data that is generated in the source systems is generally hard to manipulate by the average system user. For traditional process modeling necessary information is primarily gathered by interviewing, workshops or similar manual techniques that require the interaction of persons. This leaves room for interpretation and the tendency that ideal models are created based on often overly optimistic assumptions.

Detection, prediction, recommendation

Analysis and optimization is not limited to post-runtime inspections. Instead it can be used for **operational support** by detecting traces being executed that do not follow the intended process model. It can also be used for predicting the behavior of traces under execution. An example for runtime analysis is the prediction of the expected completion time by comparing the instance under execution with similar already processed instances. Another feature can be the provision of recommendations to the user for selecting the next activities in the process. Process mining can also be used to derive information for the design of business processes before they are implemented.

3.2 Conformance Checking

A specific type of analysis in process mining is conformance checking (Adriansyah et al. 2011). The assumption for being able to conduct conformance checking is the existence of a process model that represents the desired process. For this purpose it does not matter how the model was generated either by traditional modeling or by process mining.

Event logs can be replayed to identify conform or deviant behavior

A given event log is then compared with the ideal model for identifying conform or deviant behavior. The process instances presented in the log as cases are replayed as **simulations** in the model. Cases that can be replayed are labeled conform and cases that cannot be replayed deviant.

If we check the conformance of the event log in Figure 4 against the process model in Figure 1 we observe that cases 1, 2 and 5 are conform with the model whereas cases 3 and 4 deviate. The simple statement that cases conform or deviate is generally not sufficient. The hypothetical case with the trace ABCDD meaning that the invoice for an ordered and received good was paid twice is probably a more significant deviation from the ideal process then the incorrect sequence of activities of B and C in the trace ACBD observable in case 3. In general **local diagnostics** can be calculated that highlight the nodes in the model where deviations took place and **global conformance measures** that quantify the overall conformance of the model and the event log.

When conducting conformance checking it should be kept in mind that not every deviation needs to be negative and should therefore be eliminated. Major deviations from the ideal model might also mean that the model itself does not reflect real world circumstances and requirements.

3.3 Compliance Checking

Internal or external rules

Compliance refers to the adherence of internal or external rules. External rules primarily include laws and regulations but can also reflect industry standards or other external requirements. Internal rules include management directives, policies and standards. Compliance checking deals with investigating if relevant rules are followed. It is especially important in the context of internal or external audits.

New possibilities

Process mining offers new and rigorous possibilities for compliance checking (Ramezani et al. 2012). A major advantage is the already mentioned **reliability of used information**. In the context of compliance this has an even higher impact because individuals will normally not admit incompliant behavior in traditional information gathering techniques like interviews due to likely negative consequences for themselves.

Violation discovery

For illustrating the difference of conformance and compliance checking we refer to a well know and common compliance rule that is called the **4-Eyes-Principle**. It means that at least two persons should be involved in the execution of a business process in order to prevent errors or fraud. Errors are more likely to be discovered if a second person is involved and fraud is less likely when conspiracy is needed among individuals.

Let us have a look at case 5 in Figure 4. We have observed in section 3.1 that this case conforms with the ideal process model in Figure 1. The trace does not show any deviations from the ideal execution sequence in the model. However it is not compliant to the 4-Eyes-Principle rule because all activities were executed by the same user. Compliance checking is a relatively novel field of research in the context of process mining. It can be distinguished between forward and backward compliance checking (Ramezani et al. 2012). Forward compliance checking can further be divided into design-time and runtime compliance checking (Becker/Delfmann et al. 2012):

— **Pre-runtime compliance** checking is conducted when processes are designed or redesigned and implemented. The designed model is checked if relevant rules are violated.
— **Runtime compliance** checks if violations occur when a business transaction is processed.
— **Post-runtime** compliance is applied over a certain period of time, when the transactions have already taken place.

Approaches for pre-runtime compliance checking are available and first approaches already do exist for post-runtime compliance checking. But solutions for runtime compliance checking still need to be developed before they will be available for practitioners.

3.4 Organizational Mining

So far we have focused on the control-flow of process models by inspecting the sequence of activities that are possible in a process model. As illustrated in the example for compliance checking other attributes in the event log provides rich opportunities for investigation.

Organizational settings and interactions between resources

Organizational mining aims to analyze information that is relevant from an organizational perspective. This includes the discovery of social networks, organizational structures and resource behavior (Song and Van der Aalst 2008). Metrics like importance, distance, or centrality of individual resources (this can be individuals as shown in the sample log in Figure 4 but also information systems, machines or other resources) can be computed.

Question 4: Explain the difference between conformance and compliance checking.

4 Tool Support

Process mining tools are necessary for the application in practice. Figure 8 lists various available process mining tools.

Product Name	Type	Link
ARIS Process Performance Manager	C	www.softwareag.com
celonis business intelligence	C	www.celonis.de
Disco	C	www.fluxicom.com
Genet/Petrify	O	www.lsi.upc.edu
Interstage Business Process Manager	C	www.fujitsu.com
QPR ProcessAnalysizer	C	www.gqr.com
ProM	O	www.processmining.org
ProcessGold	C	www.processgold.de
Rbminer/Dbminer	O	www.lsi.upc.edu
ReflectOne	C	www.pallas-athena.com
ServiceMosaic	O	soc.cse.unsw.edu.au

Fig. 8: Process Mining Tools — C=commercial, O=open source (adapted from Van der Aalst, p. 271)

Open source software tool

A major academic and open-source tool is **ProM**. It provides plug-ins for many different mining algorithms, as well as analysis, conversion and export modules. **Disco** is a commercial application that benefits from intuitive and easy usability. It also provides integrated functionality for filtering and loading of event logs. It is therefore especially suited for novel process mining users. A non-commercial license is available for academic institutions.

None of the tools support the extraction of event data from the relevant source systems. This means that the data has to be extracted with specialized **data extraction software** or by using export functionalities of the source systems.

5 Challenges and Contemporary Research Questions

5.1 Noise and Incompleteness

Noise refers to rare and infrequent behavior

We mentioned that **incorrect records** in the event log can for example result from software malfunctioning, user disruptions, hardware failures or truncation of process instances during the data extraction. Erroneous records in the event log should be distinguished from a phenomenon called noise. Noise refers to correctly recorded but rare and infrequent behavior (Van der Aalst 2011 p. 148). Noise leads to **increased complexity** in the process model. Process mining approaches should therefore be able to handle or **filter** out noise. But this requirement is debatable because the meaning and role of noise varies depending on the objective of the conducted process mining project. While it might be necessary to abstract from infrequent behavior for **reducing complexity** in the reconstructed model, infrequent behavior and the detection of outliers are key aspects for conformance or compliance checking. And how can noise actually be distinguished from erroneous records? It is therefore necessary to consider the handling of noise for every project individually.

Incompleteness means that not all possible behavior is recorded

Behavior that is not recorded in the event log cannot be considered for mining a process model. On the other hand process models typically allow for much more behavior than recorded in the event log. We have seen that the process model shown in Figure 6 allows for an infinite number of sequences whereas the event log only contains five cases with three different traces. While noise refers to the problem of potentially too much data recorded in the event log incompleteness refers to the problem of having **too little data**. It is unrealistic to assume that an event log includes all possible process executions. When conducting a process mining project it should be ensured that **sufficient event log data** is available to discover the main control-flow structure.

5.2 Competing Model Quality Criteria

Four main quality criteria can be used to specify different quality aspects of reconstructed process models: fitness, simplicity, precision and generalization.

- **Fitness** addresses the ability of a model to replay all behavior recorded in the event log.
- **Simplicity** means that the simplest model that can explain the observed behavior should be preferred.

— **Precision** requires that the model does not allow additional behavior that is very different from the behavior recorded in the event log.

— **Generalization** means that a process model is not exclusively restricted to display the eventually limited record of observed behavior in the event log but that it provides an abstraction and generalizes from individual process instances.

These quality criteria compete with each other as shown in Figure 9. This means that it is normally not possible to perfectly meet all criteria simultaneously. The model in Figure 5 for example has a high precision because it does not allow for any behavior that is not included in the log but a low fitness because it cannot replay all the cases. The model in Figure 6 has a perfect fitness because it is able to replay all cases in the event log. But it has a poor precision because it allows for an infinite number of execution sequences not present in the event log. An adequate balance between the quality criteria should be achieved for every process mining project depending on the intended outcome and further use of the reconstructed process models.

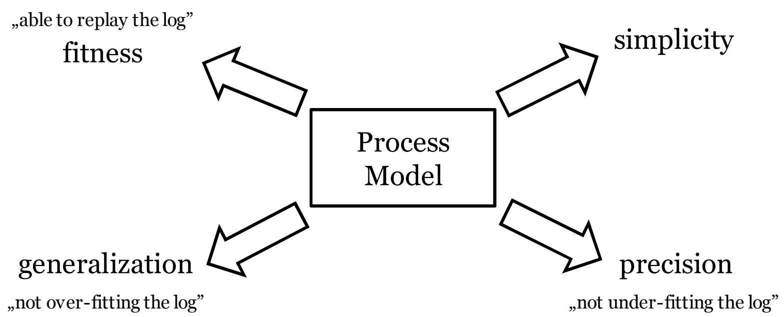


Fig. 9: Quality Dimensions (adapted from Van der Aalst 2011, p. 151)

5.3 Event Log Quality and Labeling

The quality of event logs is crucial for the quality of the mined and reconstructed process models. Business process and workflow management systems provide the highest quality of event logs (Van der Aalst et al. 2012). They primarily focus on supporting and automating the execution of business processes and therefore most likely also store high quality event data that can easily be used for process mining. Data from ERP systems in general do not provide the same quality of event logs. The logging of event data is more a byproduct than intended software functionality.

The quality of an event log depends on the source systems ability to record process relevant data

A fundamental assumption for contemporary process mining approaches is that events in an event log are already mapped to cases. But it depends on the quality of the available event log if this is indeed the case. **ERP systems** as one major source of transactional data in organizations commonly do not store explicit data that maps events to cases. An interesting probabilistic approach for **labeling event data** is presented by (Ferreira/Gillblad 2009). A promising approach could also be the consideration of the **application domain** context. Gehrke / Müller-Wickop present a mining algorithm that is able to operate with unlabeled events by exploiting characteristic structures in the available data thereby mapping events to cases in the process of mining (Gehrke/Müller-Wickop 2010).

5.4 Complex Process Models

Complexity reduction is an important research area

The presented examples in Figure 3 and 5 display very simple process models. Real world processes are commonly much more complex. Their graphical representation can lead to highly complex and **incomprehensible models** as shown in Figure 10. Two typical categories of complex process models are called lasagna and spaghetti processes (Van der Aalst 2011 pp. 277-320) because of their intertwined appearance. The reduction of complexity is a **major chall**enge and subject to recent research (Reichert 2012).

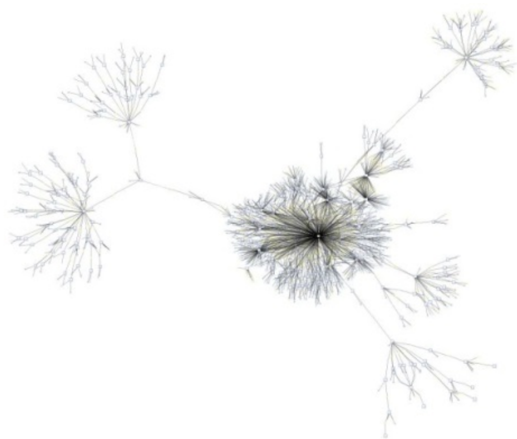


Fig. 10: Example of a Complex Process Model (Werner et al. 2012 p. 5)

5.5 Concept Drift

Processes change over time

When processes are mined and reconstructed it is usually assumed that they are stable over the time of observation. But this might not be the case. A process might work differently over a certain period of time. This is called concept drift.

Assuming stable business processes is therefore a simplistic view and contemporary research introduces approaches how to deal with concept drift in the context of process mining (Bose et al. 2011).

Question 5: Why is it necessary to ensure a balance of different quality criteria for process mining?

References:

Van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes, Berlin, Heidelberg 2011.

Van der Aalst, W.M.P./Andriansyah, A./Alves de Medeiros, A. K./Arcieri, F./Baier, T./Blickle, T./Bose,

J.C./Van den Brand, P./Brandtjen, R./Buijs, J.: Process mining manifesto. In: BPM 2011 Workshops Proceedings, pp. 169-194.

Van der Aalst, W.M.P./Weijters, A. J. M. M./Maruster, L.: Workflow Mining: Which Processes can be

Rediscovered. Presented at the Proc. Int'l Conf. Eng. and Deployment of Cooperative Information Systems (EDCIS), 2002 (Vol. 2480), pp. 45-63.

Adriansyah, A./Van Dongen, B.F./Van der Aalst, W.M.P.: Conformance checking using cost-based fitness analysis. In: Enterprise Distributed Object Computing Conference (EDOC), 2011, 15th IEEE International, pp. 55-64.

Agrawal, R./Gunopulos, D./Leymann, F.: Mining Process Models from Workflow Logs. In: Proc. Sixth Int'l Conf. Extending Database Technology, 1998, pp. 469-483.

Becker, J./Delfmann, P./Eggert, M./Schwittay, S.: Generalizability and Applicability of Model-Based Business Process Compliance-Checking Approaches — A State-of-the-Art Analysis and Research Roadmap. In: BuR — Business Research (5:2), 2012, pp. 221-247.

Becker, J./Probandt, W./Vering, O.: Grundsätze ordnungsmäßiger Modellierung Konzeption und Praxisbeispiel für ein effizientes Prozessmanagement. Berlin, Heidelberg 2012.

Bose, R./Van der Aalst, W./Zliobaite, I./Pechenizkiy, M.: Handling concept drift in process mining. In: Advanced Information Systems Engineering, 2011, pp. 391-405.

Chen, H./Chiang, R.H.L./Storey, V.C.: Business Intelligence and Analytics: From Big Data to Big Impact. In: MIS Quarterly (36:4), 2012, pp. 1165-1188.

Cook, J. E./Wolf, A. L.: Discovering models of software processes from event-based data. In: ACM Trans. Softw. Eng. Methodol. (7:3), 1998, pp. 215-249.

Ferreira, D./Gillblad, D.: Discovering process models from unlabelled event logs. In: Business Process Management, 2009, pp. 143-158.

fluxicon: Process Mining and Process Analysis. www.fluxicon.com.

Gehrke, N./Müller-Wickop, N.: Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems. In: Proceedings of the 16th Americas Conference on Information Systems, 2010. Presented at the Americas Conference on Information Systems, Lima, Peru.

Günther, C./Van der Aalst, W.: Fuzzy mining-adaptive process simplification based on multi-perspective metrics. Business Process Management, 2007, pp. 328-343.

Herbst, J./Karagiannis, D.: Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models. In: Proceedings 9th International Workshop on

Database and Expert Systems Applications 1998, pp. 745-752.

Kemper, H.-G./Mehanna, W./Baars, H.: Business intelligence — Grundlagen und praktische Anwendungen? Eine Einführung in die IT-basierte Managementunterstützung. Wiesbaden 2010.

Krcmar, H.: Informationsmanagement. Berlin 2010.

Process Mining Group: Process Mining. www.processmining.org.

Ramezani, E./Fahland, D./Van der Aalst, W. M. P.: Where Did I Misbehave? Diagnostic Information in

Compliance Checking. In: Business Process Management Lecture Notes in Computer Science, edited by A. Barros/A. Gal/E. Kindler, 2012 (Vol. 7481) Springer, pp. 262-278.

Reichert, M.: Visualizing Large Business Process Models: Challenges, Techniques, Applications. In: 1st Int'l Workshop on Theory and Applications of Process Visualization, Presented at the BPM 2012, Tallin.

Song, M./Van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. Decision Support Systems (46:1), 2008, pp. 300-317.

Tiwari, A./Turner, C.J./Majeed, B.: A review of business process mining: state-of-the-art and future trends. Business Process Management Journal (14:1), 2008, pp. 5-22.

Werner, M./Schultz, M./Müller-Wickop, N./Gehrke, N./Nüttgens, M.: Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits. In: Proceedings of 33rd International Conference on Information Systems Orlando 2012.

Weske, M.: Business process management concepts, languages, architectures. Berlin, New York 2012.

Question and Answers:

Question 1: Why is process mining important? Why can it be seen as a bridge between data mining and business process management?

Business intelligence as an important approach to use data stored in information systems for improving decision making processes and to overcome challenges like data explosion and information overflow. Data mining is the analysis of data for finding relationships and patterns. Process mining uses data mining techniques in the context of business process management and enables the application of innovative approaches for improving the management of business processes.

Question 2: What is the difference between event log, case, event, and trace? How do they relate to each other?

An event log is a collection of events. A case is a record of events that relate to a single executed process instance. An event is a recorded execution of an activity. A trace is a recorded sequence of events that belong to the same case. Events are mapped to cases. Each case has a trace. Different cases can and commonly do embody the same trace.

Question 3: Why do different process mining algorithms produce different process models?

The results of a process mining algorithm depend on its design. Process mining algorithms use different approaches for mining.

They can be deterministic or non-deterministic. They can also use a heuristic or genetic approach. The results differ according to the used approach. Another source of variety is the modeling language. Mining algorithms that use different modeling languages for representation will create different process models.

Question 4: Explain the difference between conformance and compliance checking.

The aim of conformance checking is the identification of process instances that do not conform to a given process model. Compliance checking is the analysis if compliance rules are adhered to. The adherence of compliance rules is commonly safeguarded by controls. Process models and instances can be checked if controls were effective or not.

Question 5: Why is it necessary to ensure a balance of different quality criteria for process mining?

It is generally not possible to achieve a perfect match of opposing quality criteria in real world settings. The objectives and circumstances differ for each process mining project. They need to be identified at the beginning of each project. The relevance of each quality criterion should be evaluated in regard to these objectives and circumstances and they should be assessed accordingly. A mining approach should then be chosen that leads to the desired balance of relevant quality criteria.

10.3 Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen

Number	**3**
Title	**Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen**
Appendix	10.3
Primary Related Chapters	4.2, 5.1
Type	Book Chapter
Book	Forschung für die Wirtschaft
Reference	(Werner and Gehrke, 2011)
Acceptance Rate	32)
VHB JQ 2.1 Ranking	-
WKWI Ranking	-
ERA 2010	-
CORE 2013	-
Review Procedure	Blinded
Number of Reviews	-
Authors	1. Michael Werner 2. Nick Gehrke
Dissertation Points	0.67
Authorship	
Overall	95%
Design	95%
Realization	95%
Writing	95%
Status	Published
Part of other Dissertations	No
Link	http://www.shaker.de/de/content/catalogue/index.asp?lang=de&ID=8&ISBN=978-3-8440-0684-1

[32] Invited for Publication

Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen

MICHAEL WERNER, NICK GEHRKE

NORDAKADEMIE – Hochschule der Wirtschaft, Elmshorn

Abstract: Die Jahresabschlussprüfung stellt einen komplexen und hoch spezialisierten Prozess dar. Die Durchführung der Jahresabschlussprüfung wird vom Gesetzgeber vorgesehen, um die Adressaten der Jahresabschlussprüfung vor Fehlinformationen zu schützen. Im Zuge der Digitalisierung der Wirtschaft und die damit kontinuierlich steigende Durchdringung der Geschäftstätigkeiten durch Informationssysteme stellen die Jahresabschlussprüfer vor neue Herausforderungen. Die zunehmende Automation der Geschäftsprozessverarbeitung durch den Einsatz moderner Informationssysteme lässt traditionelle Prüfungshandlungen insbesondere zur Prüfung von Geschäftsprozessen und internen Kontrollsystemen ineffektiv oder zumindest ineffizient werden. Auf Unternehmensseite erfolgt die Verarbeitung von Geschäftsvorfällen zunehmend systembasiert und automatisiert. Die Prüfungsprozeduren beim Jahresabschluss sind hingegen geprägt durch manuelle Prüfungshandlungen. Damit ergibt sich eine Diskrepanz zwischen einer automatisierten Transaktionsverarbeitung auf Seiten der Unternehmen und manuellen Prüfungsprozeduren auf Seiten der Jahresabschlussprüfer. Eine Möglichkeit, dieser Diskrepanz zu begegnen, besteht darin, Methoden der Prozessextraktion und Prozessrekonstruktion mit Methoden zur automatisierten Prüfung von im Informationssystem eingebetteten Kontrollen zu kombinieren, um so adäquate systembasierte und automatisierte Prüfungsprozeduren zu entwickeln. In diesem Artikel untersuchen wir die Anwendungsmöglichkeiten sowie auch Grenzen dieses Ansatzes.

1 Einleitung

Die Prüfung der externen Finanzberichterstattung von Unternehmen in Form einer Jahresabschlussprüfung stellt in unserem Wirtschaftssystem eine wichtige Kontrolle dar, deren Ziel es ist, die Adressaten des Jahresabschlusses vor Fehlinformationen zu schützen.[33] Die Jahresabschlussprüfung nimmt als Kontrollfunktion eine derart hervorgehobene Stellung ein, dass vom Gesetzgeber die Zuständigkeit zur Durchführung der Jahresabschlussprüfung einer speziellen Berufsgruppe zugewiesen ist, die eine entsprechende Qualifikation aufweist.[34]

Jahresabschlussprüfer, traditionell auf Themenschwerpunkte der Rechnungslegung ausgerichtet, sehen sich mit der fortschreitenden Integration von Geschäftsprozessen und Informationssystemen neuen Herausforderungen gegenübergestellt. Gegenwärtige Prüfungsansätze berücksichtigen bis zu einem gewissen Grad den Gedanken der Prozessorientie-

[33] Die Pflicht zur Prüfung des Jahresabschlusses ergibt sich aus §316 HGB.

[34] Nach §319 HGB erfolgt die Prüfung des Jahresabschlusses durch Wirtschaftsprüfer, Wirtschaftsprüfungsgesellschaften, vereidigte Buchprüfer oder Buchprüfungsgesellschaften.

rung und die Verflechtung von Geschäftsprozessen und Informationssystemen. Sowohl nationale wie auch internationale Prüfungsstandards (vgl. IDW PS 261 [1] und ISA 315 [2]) verlangen die Anwendung risikoorientierter Prüfungsansätze. Bei diesen Ansätzen werden zunächst die wesentlichen Risiken identifiziert, die zu fehlerhaften Darstellungen in der Finanzberichterstattung führen können. Ausgehend von dieser Risikoeinschätzung wird evaluiert, welche Kontrollen innerhalb eines Unternehmens vorhanden sind, um die vorhandenen Risiken zu minimieren. Bei dieser Betrachtung sind auch Kontrollen des internen Kontrollsystems zu würdigen. Darüber hinaus ist die Berücksichtigung von Informationssystemen und deren Prüfung mittlerweile ebenfalls obligatorisch (vgl. IDW PS 330 [3] in Verbindung mit IDW RS FAIT 1[4] bzw. ISA 315.81 [2]).

Eine grundlegende Problematik wird bei bisher angewendeten, risikoorientierten Prüfungsvorgehen jedoch nicht adäquat berücksichtigt. Mit zunehmender Integration von Geschäftsprozessen in Informationssysteme auf Seiten der Unternehmen nimmt die Automatisierung der Verarbeitung von Geschäftsvorfällen zu. Enterprise Resource Planing (ERP) Systeme stellen die am weitesten verbreiteten Informationssysteme zur Unterstützung und Automatisierung zur Transaktionsverarbeitung dar. Sie dienen jedoch nicht nur der Unterstützung und Automatisierung der Geschäftsprozessabwicklung, sondern bilden auch die Basis für die interne und externe Finanzberichterstattung. Dies bedeutet, dass die Informationen, die während der Verarbeitung der Geschäftsvorfälle erzeugt und gespeichert werden, zugleich die Datenbasis bilden für die Berichterstattung, die letztendlich eine Aggregation der zu Grunde liegenden Transaktionsdaten darstellt.

In dem Maße, in dem die Integration der Geschäftsprozesse in ERP Systeme zunimmt, gewinnen diese an Bedeutung für die Finanzberichterstattung und somit auch für die Jahresabschlussprüfung, die letztendlich eine Aussage treffen soll, ob die dargestellten Informationen frei von wesentlichen Fehlern sind.

Der systembasierten und automatisierten Verarbeitung auf Unternehmensseite stehen manuelle Prüfungsprozeduren auf Seiten der Jahresabschlussprüfer gegenüber. Die manuelle Durchführung von Prüfungshandlungen nimmt dabei einen großen Teil der Ressourcen der Abschlussprüfer in Anspruch. Durch geeignete systembasierte und automatisierte Prüfungsprozeduren wäre es möglich, die notwendigen Prüfungshandlungen zur Prüfung integrierter Geschäftsprozesse effektiver und effizienter zu gestalten, was letztendlich Ressourcen freisetzen würde für Prüfungshandlungen zu ungewöhnlichen, von Standardprozessen abweichenden, und komplexen Geschäftsvorfällen, die inhärent ein höheres Risiko für Fehler und Manipulation aufweisen.

In WERNER et al. [5] wird beschrieben, wie Methoden der Prozessextraktion (Process Mining) und der Prozessrekonstruktion in Verbindung mit Methoden zur automatisierten Überprüfung von im Informationssystem eingebetteten Kontrollen (Application Controls bzw. Anwendungskontrollen) kombiniert werden können, um systembasierte und automatisierte Prüfungsprozeduren zu entwickeln und anzuwenden. Dieser Ansatz wird in diesem Artikel aufgegriffen mit dem Ziel, darzustellen, welche Möglichkeiten sich insbesondere aus der Anwendung der Prozessextraktion und Prozessrekonstruktion ergeben und wo sich Grenzen bzw. Restriktionen bei deren Anwendung ergeben.

Im Abschnitt zwei dieses Artikels wird ein Überblick über den derzeitigen Stand der wissenschaftlichen Literatur zum behandelten Themenfeld erörtert. Abschnitt drei schließt mit einer kurzen Erläuterung zum Prozess der Abschlussprüfung an, um darzustellen, in welchen Teilprozessen mittels systembasierter Prüfungsprozeduren Effizienz- und Effektivitätsgewinne erzielt werden können. In Abschnitt vier wird erläutert, wie systembasierte und automatisierte Prüfungsprozeduren eingesetzt werden können. Diese Betrachtung ist notwendig, um erschließen zu können, wie und auf welchen Ebenen sich neue Erkenntnisse durch Prozessrekonstruktion erzielen lassen. In Verbindung mit den vorangegangenen Abschnitten wird in Abschnitt fünf dargestellt, welche Grenzen sich beim Einsatz der entwickelten Prozeduren ergeben und welchen Restriktionen deren Einsatz unterliegen. Abschnitt sechs schließt mit einer Zusammenfassung und einem Ausblick auf weitere Forschungsinhalte. Die in diesem Artikel erläuterten Erkenntnisse basieren auf dem vom Bundesministerium für Bildung und Forschung finanzierten Forschungsprojekt „Virtual Accounting Worlds“ [6].

2 Stand der Wissenschaft

Wesentliche Grundlage für die in diesem Artikel dargestellten Methoden sind Arbeiten zur Prozessextraktion oder Process Mining, dessen Ursprung in den 1990er Jahren liegt. COOK und WOLF [7], [8], [9] untersuchten Process Mining im Zusammenhang mit Software Engineering. Sie beschreiben unterschiedliche Methoden zum Process Mining, jedoch ohne Vorgehen zu präsentieren, die explizit eine Erstellung von Prozessmodellen erlauben. AGRAVAL et al. [10] wenden Process Mining im Zusammenhang mit Workflow Systemen an. HERBST und KARAGIANNIS adressieren Process Mining ebenfalls im Zusammenhang mit Workflow Management unter Verwendung eines induktiven Ansatzes [11], [12], [13], [14], [15], [16]. Weitere Betrachtungen zur Entwicklung von Mining Software wurden durchgeführt von MAXEINER et al. [17] und SCHIMM [18], [19], [20], [21].

Wesentliche Beiträge zum Process Mining wurden von VAN DER AALST et al. [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34] veröffentlicht. Sie handeln vorwiegend von rekonstruierten Prozessmodellen aus Event Logs, die mittels Petri-Netzen dargestellt werden. Diese Arbeiten sind besonders relevant, da die entwickelten Methoden zum Process Mining auf Basis von Event Logs adaptiert werden können für Extraktions- und Rekonstruktionsmethoden zu Prozessen in Finanzbuchhaltungssystemen. Die Arbeiten von VAN DER AALST et al. decken darüber hinaus relevante Themenbereiche zur Workflow Performance, Nebenläufigkeit, Rauschen und Abweichungsanalysen im Zusammenhang mit Process Mining ab. Obwohl diese Arbeiten hilfreiche Grundlagen bieten, sind einige Einschränkungen zu berücksichtigen. Während bei den von VAN DER AALST et al. erarbeiteten Methoden die Rekonstruktion von Prozessmodellen in Form von Petri-Netzen, die die Gesamtheit aller möglichen Prozesspfade eines Event Logs repräsentieren, im Mittelpunkt steht, ist die Zielsetzung bei der Rekonstruktion von Finanzprozessen aus ERP Systemen unterschiedlich. Letztere beabsichtigt, bestimmte Geschäftsvorfälle für spezielle Fragestellungen zu extrahieren und zu einem Prozessmodell zu aggregieren. Bei der Aggregation werden insbesondere Zusammenhänge aus der Rechnungslegung herangezogen, indem die Verknüpfung der zugrundeliegenden Belege über eine offene-Posten-Systematik genutzt wird. Des Weiteren stellen ERP Systeme wesentlich detailliertere Informationen zur Verfügung als herkömmliche Event Logs, die als Basis für spezielle Rekonstruktionen verwendet werden können.

Generell lässt sich feststellen, dass die wissenschaftliche Literatur eine stark technische Prägung aufweist. Die Verbindung zwischen Prozessflüssen, Prozessextraktion, Prozessrekonstruktion und Prozessaggregation im Anwendungsfeld des Rechnungswesens ist in der wissenschaftlichen Literatur wenig vertreten. Der Grund hierfür liegt wahrscheinlich in der thematischen Distanz der klassischen Themenbereiche Rechnungswesen und Compliance auf der einen sowie informationstechnisches Process Mining auf der anderen Seite.

In jüngerer Zeit befassen sich Arbeiten von GEHRKE et al. [35], [36], [37], [38], MÜLLER-WICKOP et al. [39] sowie WERNER et al. [5] mit der Automatisierung von Prozessprüfungen im Rahmen der Jahresabschlussprüfung. Hier werden Methoden entwickelt, die eine Anwendung von Process Mining in ERP Systemen für Finanzprozesse ermöglichen. In WERNER et al. [5] werden die angewendeten Methoden in einen Gesamtzusammenhang gesetzt. Die vorliegende Arbeit greift insbesondere die dort vorgestellten Erkenntnisse auf und erweitert diese um Betrachtungen der Möglichkeiten und Grenzen die sich durch den Einsatz von Prozessextraktion und -rekonstruktion im Rahmen der Jahresabschlussprüfung und auch darüber hinaus ergeben.

Auch JANS et al. [40], [41] und ALLES et al. [42] behandeln den Zusammenhang zwischen Process Mining und dessen Einsatz bei Accounting Information Systems und ERP Systemen. JANS et al. betrachten Process Mining u.a. als Fortentwicklung von Data Mining Techniken und Massendatenauswertungen für spezielle Fragestellungen hinsichtlich der Aufdeckung krimineller Aktivitäten und der Prüfung von Geschäftsprozessen. Die Ansätze von GEHRKE et al. [35], [36], [37], [38], MÜLLER-WICKOP et al. [39] und WERNER et al. [5] nehmen eine unterschiedliche Betrachtung vor, indem die fachliche Perspektive im Sinne der Betrachtung der Zusammenhänge aus Rechnungslegungssicht als Basis für die zu entwickelnden Extraktions-, Aggregations- und Auswertungsmethoden herangezogen wird. Dieser Ansatz unterscheidet sich dahingehend von der technischen Erweiterung von Data Mining Methoden durch Methoden des Process Mining. Ziel des von GEHRKE et al., MÜLLER-WICKOP et al. und WERNER et al. behandelten Forschungsansatzes ist es, automatisierte und systembasierte Methoden für die Prüfung von integrierten Geschäftsprozessen zu entwickeln, die in einem risikoorientierten Prüfungsansatz eine neue Bedeutung einnehmen und bestehende Methoden nicht nur erweitern. Trotz dieses Unterschieds werden insbesondere die von JANS et al. [40] beschriebenen Einsatzfelder von Business Process Mining in diesem Artikel aufgegriffen und auf Anwendbarkeit der speziell von GEHRKE et al. entwickelten Methoden untersucht.

3 Grenzen manueller Prüfungsprozeduren bei hochintegrierten Geschäftsprozessen

Bevor wir bewerten können, warum manuelle Prüfungsprozeduren bei Jahresabschlussprüfungen für Unternehmen, deren Geschäftsprozesse stark in ERP Systeme integriert sind, ineffizient oder im Extremfall auch ineffektiv werden, ist ein Verständnis des Prozesses zur Durchführung des Jahresabschlusses notwendig sowie ein Überblick der zum Einsatz kommenden Prüfungstechniken. Grundlegende Informationen im Hinblick auf die in diesem Artikel dargestellte Problematik sind bereits in WERNER et al. [5] aufgeführt, so dass an dieser Stelle nur wesentliche Aspekte aufgegriffen werden.

3.1 Die Jahresabschlussprüfung

Unternehmen sind gesetzlich dazu verpflichtet, einen Jahresabschluss zu erstellen, der von Jahresabschlussprüfern geprüft und testiert wird (§316 und §319 HGB). Die externen Adressaten des Jahresabschlusses umfassen insbesondere die Unternehmensleitung, Controller, Aufsichtsrat oder Beirat, Finanzverwaltung, Kreditgeber, Gesellschafter und Anteilseigner, Lieferanten, Arbeitnehmer, Kunden, externe Aufsichtsbehörden, Finanzanalysten, Konkurrenz und die generelle Öffentlichkeit [43]. Der Jahresabschlussprüfer hat festzustellen, ob der Jahresabschluss ein getreues Bild der Vermögens-, Finanz- und Ertragslage des Unternehmens wiedergibt und frei ist von wesentlichen Fehlern. Die Anforderungen für die Prüfung des Jahresabschlusses sind in Prüfungsstandards definiert, die von nationalen (IDW) und internationalen Institutionen (IAASB) erstellt werden, nach denen sich die Abschlussprüfer zu richten haben. Gleiches gilt für Standards zur Rechnungslegung. Es bleibt zu erwähnen, dass sowohl Rechnungslegungsstandards wie auch Prüfungsstandards von Land zu Land variieren. Es ist jedoch zu beobachten, dass international eine Angleichung der Standards erfolgt [44]. Eine Diskussion über die Konvergenz der Rechnungslegungs- bzw. Prüfungsstandards ist nicht Gegenstand dieser Arbeit. Allerdings gelten die hier erarbeiteten Ergebnisse nicht nur für den deutschen Raum sondern sind auch übertragbar auf andere Länder, die westlich geprägten Wirtschaftssystemen folgen.

Die Jahresabschlussprüfung stellt einen Prüfungsprozess dar, der aus verschiedenen Teilprozessen besteht. Der erste Teilprozess umfasst im Allgemeinen die Informationssammlung über das zu prüfende Unternehmen und die Identifikation und Bewertung potentieller Fehlerrisiken für die Abschlusserstellung anhand der zur Verfügung stehenden Informationen. Auf die Analyse möglicher Fehlerrisiken folgt eine Informationserhebung im Unternehmen vorhandener Geschäftsprozesse mit dem Ziel, interne Kontrollen in den Prozessen zu identifizieren, die geeignet sind, relevante Risiken zu minimieren. Wenn diese Kontrollen als geeignet beurteilt werden, schließt sich die Funktionsprüfung der Kontrollen an, in der überprüft wird, ob die Kontrollen im jeweils zu betrachtenden Zeitraum effektiv waren, und die Kontrollziele erreicht haben. Nachdem die Kontrollen getestet sind, wird evaluiert, welche weiteren aussagebezogenen Prüfungshandlungen notwendig sind, um eine ausreichende Prüfungssicherheit zu gewährleisten, so dass davon ausgegangen werden kann, dass der geprüfte Abschluss mit hinreichender Sicherheit frei ist von wesentlichen Fehlern. Der Prüfungsprozess endet mit der Erstellung des Prüfungsberichts und der Testierung. Der Prozess der Jahresabschlussprüfung ist in Abbildung 3.1.1 dargestellt.

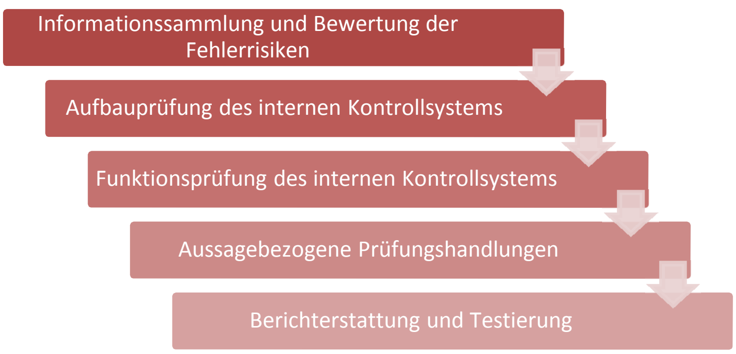


Abbildung 3.1.1 Prüfungsprozess

Innerhalb der einzelnen Teilprozesse kommen verschiedene Prüfungsprozeduren zum Einsatz. Für die Aufbau- und Funktionsprüfung finden Informationserhebungen und Kontrolltests statt. Für die anschließenden aussagebezogenen Prüfungshandlungen werden analytische und substantielle Prüfungshandlungen durchgeführt. Für die Durchführung der Prüfungshandlungen werden unterschiedliche Techniken verwendet wie Interviews, Beobachtung, Begutachtung von Unterlagen (Inspektion) und Wiederholung von Tätigkeiten (Reperformance). Je höher der Verlässlichkeitsgrad der Information für eine bestimmte Prüfungshandlung angesetzt wird, desto investigativer muss die gewählte Prüfungstechnik sein und desto höher ist der Prüfungsaufwand. Abbildung 3.1.2 veranschaulicht diesen Zusammenhang.

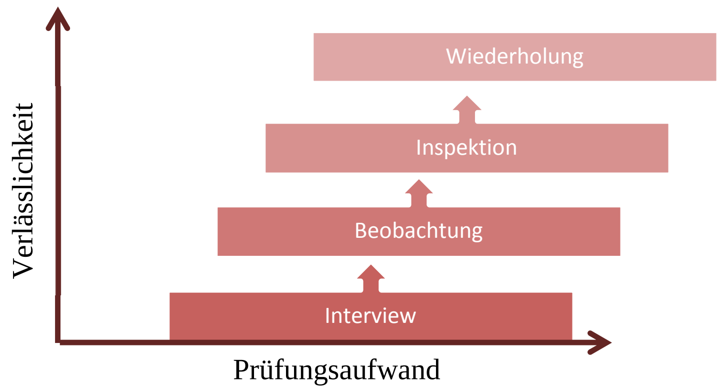


Abbildung 3.1.2 Prüfungstechniken

Zur Informationserhebung und Aufbauprüfung werden vorwiegend Interviews geführt, bei denen Ansprechpartner mit ausreichender Sachkenntnis im Unternehmen befragt werden. Bei der Beobachtung werden Mitarbeiter bei der Durchführung von Tätigkeiten aktiv beobachtet, um daraus die gewünschten Informationen zu ziehen. Bei der Inspektion werden vorhandene Ursprungsbelege gesichtet und evaluiert. Bei der Wiederholung werden Tätigkeiten, die in der Regel durch Mitarbeiter des zu prüfenden Unternehmens als Bestandteil normaler Tätigkeiten zur Geschäftsprozessabwicklung oder zu Kontrollzwecken durchgeführt werden, vom Prüfer unabhängig wiederholt und damit nachvollzogen.

3.2 Manuelle Prüfungsprozeduren bei hochintegrierten Geschäftsprozessen

Geschäftsprozesse spielen bei der Jahresabschlussprüfung sowohl in der Aufbauprüfung als auch in der Funktionsprüfung eine zentrale Rolle. Während der Aufbauprüfung werden vor allem Interviews durchgeführt, um ein Verständnis der als wesentlich identifizierten Prozesse zu erhalten. Die Durchführung der Interviews ist eine rein manuelle Tätigkeit und sehr zeitintensiv. Während der Funktionsprüfung werden vor allem Beobachtungen am System, Inspektionen von zur Verfügung gestellten Unterlagen zu Buchungsbelegen und Kontrollbelegen durchgeführt, sowie Kontrollhandlungen nachvollzogen.

Zu Beginn der Prüfung identifiziert der Prüfer geeignete Interviewpartner, um Informationen über den Prozessablauf und implementierte Kontrollen zu erhalten. Nach Durchführung der Interviews evaluiert er, ob Kontrollen vorhanden und angemessen sind, um die Risiken, die sich aus dem Geschäftsprozess für die Finanzberichterstattung ergeben, zu minimieren. In einem zweiten Schritt sucht der Prüfer eine geeignete Grundgesamtheit, um stichprobenartig eine Funktionsprüfung der Kontrollen durchzuführen. Für die Funktionsprüfung werden bereitgestellte Unterlagen inspiziert oder direkt Prüfungen im System durchgeführt, um die Funktionsfähigkeit eingebetteter Kontrollen (Application Controls bzw. Anwendungskontrollen) zu verifizieren.

Bei manuellen Prozessen mit wenigen Geschäftsvorfällen und geringer Integration der Informationssysteme kann die manuelle Durchführung der notwendigen Prüfungshandlungen eine angemessene Vorgehensweise darstellen. Bei Unternehmen, bei denen die Geschäftsprozesse stark in die Informationssysteme integriert sind, ergeben sich eine Reihe von Faktoren, die die Angemessenheit manueller Prüfungshandlungen in Frage stellen.

Mit zunehmender Integration und Automatisierung der Transaktionsverarbeitung wird der Prozessablauf für die einzelnen beteiligten Personen intransparenter. Dies führt dazu, dass die Ansprechpartner, die im Regelfall der Fachabteilung entstammen, häufig kein vollständiges Verständnis der Verarbeitungsroutinen im System haben. Hinzu kommt, dass die Anwendung von Stichprobenverfahren an Grenzen stößt, wenn die Grundgesamtheit der Transaktionen sehr umfangreich wird. Bei Millionen von Transaktionen pro Geschäftsjahr ist es fraglich, ob die Inspektion von Stichproben noch effektiv ist. Zumindest kann davon ausgegangen werden, dass mit einer notwendigen Ausweitung der Stichproben aufgrund der Zunahme der zu betrachtenden Geschäftsvorfälle die Effizienz der manuellen Prüfungstätigkeiten abnimmt.

4 Anwendungsmöglichkeiten der Prozessrekonstruktion

In diesem Abschnitt werden die Anwendungsmöglichkeiten untersucht, die sich durch den Einsatz von Methoden zur Prozessextraktion, -rekonstruktion und -aggregation ergeben und wie diese mit Methoden zur automatischen Prüfung von Anwendungskontrollen kombiniert werden können. Um die Möglichkeiten der Anwendung zu untersuchen, ist es zunächst notwendig, die grundlegenden Methoden sowie deren Zusammenwirken zu verstehen.

4.1 Methodenkombination

Methoden zur Prozessextraktion für Finanzprozesse erlauben, in ERP Systemen gespeicherte Daten zu extrahieren. Die extrahierten Daten können über entsprechende Algorithmen zu Prozessinstanzen aggregiert und visualisiert werden (vgl. [36] und [37]). Grundlage für die Extraktion und Rekonstruktion ist die Verknüpfung der rechnungslegungsrelevanten Daten über offene-Posten-geführte Buchungen. Der Ausschnitt einer rekonstruierten und visualisierten Prozessinstanz ist in Abbildung 4.1.1 dargestellt.

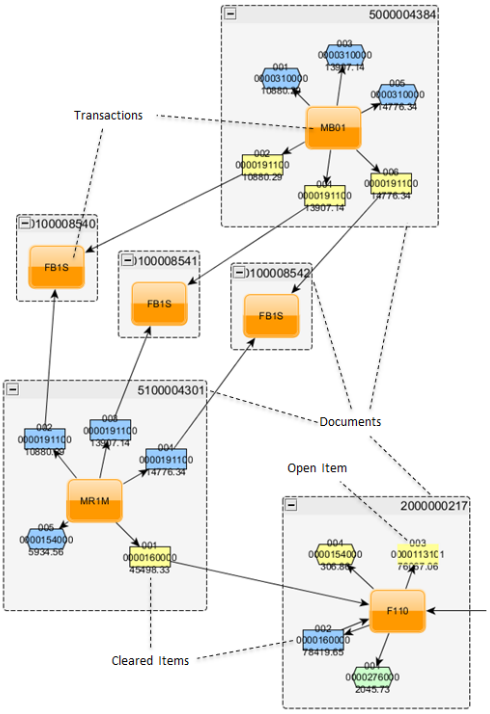


Abbildung 4.1.1 Prozessinstanz aus [5]

Die dargestellte Prozessrekonstruktion kann für beliebig viele und beliebig komplexe Prozessinstanzen durchgeführt werden. Als Ergebnis der Rekonstruktion liegen je nach Auswertungsziel eine bis alle in dem jeweiligen Betrachtungszeitraum im ERP System durchgeführten Prozessinstanzen vor. Die extrahierten und rekonstruierten Instanzen müssen zu Prozessmodellen aggregiert werden. Auf diese Weise werden eine höhere Abstraktionsebene und die Möglichkeit erreicht, die aggregierten Prozessmodelle mit den Ergebnissen automatisierter Prüfung von Anwendungskontrollen zu vereinigen.

Steuerungs- und Kontrollmechanismen in ERP Systemen erlauben, die Durchführung von Transaktionen im System zu beeinflussen und zu regulieren. Sie werden meist bei der Systemeinführung entsprechend der Anforderungen des Nutzers konfiguriert. Anwendungskontrollen steuern und überwachen die Durchführung aller Transaktionen, für die sie aktiviert sind. Während bei manuell durchgeführten Prozessen manuelle Kontrollen dazu dienen, die vollständige und richtige Abarbeitung von Geschäftsvorfällen sicherzustellen, wird diese Funktion bei integrierten Geschäftsprozessen von Anwendungskontrollen wahrgenommen. Die Herausforderung in der Praxis besteht darin, die in den ERP Systemen unterschiedlicher Softwareanbieter heterogen implementierten Kontrollen für die Zwecke der automatisierten Prüfung auszuwerten. In [35] wird hierzu ein Lösungsansatz vorgestellt.

Mittels automatisierter Prozessrekonstruktion und der anschließenden Aggregation zu Prozessmodellen in Kombination mit der Anwendung von Methoden zu automatisierten Prüfung von Anwendungskontrollen ist es möglich, Methoden zur automatisierten Prüfung von Geschäftsprozessen zu entwickeln. Dieser Zusammenhang ist in Abbildung 4.1.2 graphisch dargestellt.

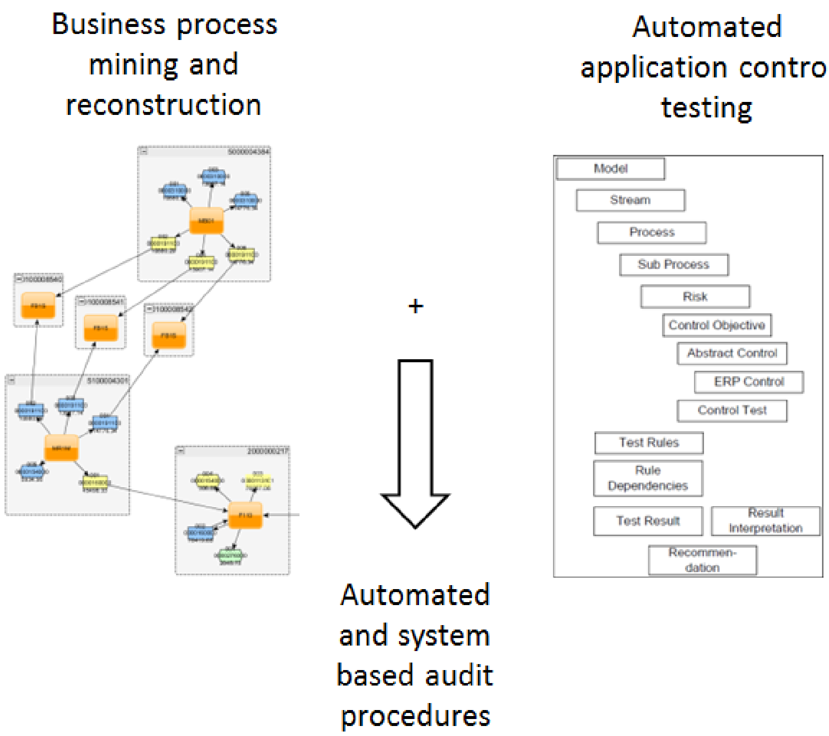


Abbildung 4.1.2 Kombination Prozessrekonstruktion und automatisierte Prüfung von Anwendungskontrollen

4.2 Möglichkeiten der Prozessrekonstruktion für die Jahresabschlussprüfung

Die im vorherigen Abschnitt präsentierten Methoden können nun für Zwecke der Automatisierung der Prüfung von Geschäftsprozessen in Jahresabschlussprüfungen integriert werden.

Zur Verdeutlichung der Integration der Methoden verwenden wir eine Analogie, die in [5] vorgestellt wurde. Wir vergleichen hierbei den Jahresabschluss mit einem See, der mit Wasser gefüllt ist. Das Wasser entspricht den rechnungslegungsrelevanten Daten in ERP Systemen. Der See wird von Flüssen gespeist. Die Flüsse entsprechen den Geschäftsprozessen, die das Wasser aufnehmen, bzw. die Transaktionen erzeugen, und diese in den See bzw. die Finanzberichterstattung tragen. Die Jahresabschlussprüfung vergleichen wir mit der Prüfung der Wasserqualität des Sees. Bei der Anwendung gegenwärtiger Prüfungsverfahren würde der Prüfer aus den Flüssen stichprobenartig Wasserproben entnehmen und deren Qualität prüfen und diese Prüfung ggf. durch weitere Stichproben direkt im See ergänzen. Das Problem bei diesem Ansatz liegt darin, dass einerseits die Flüsse nur stichprobenartig zu bestimmten Zeitpunkten getestet werden und viel gravierender, dass der Prüfer gar nicht weiß, welche Flüsse, Parallelläufe und Verzweigungen tatsächlich existieren. Hierfür kann er lediglich auf seine Erfahrungen aus ähnlichen Prüfungen aufbauen, die zutreffend sein können oder aber auch nicht.

Bei der Anwendung automatisierter und systembasierter Prüfungsmethoden werden diese Informationen transparent. Der Prüfer erhält zunächst eine Karte über alle Zuflüsse, die den See speisen (Prozessrekonstruktion). Des Weiteren erhält er Einsicht, durch welche

Kontrollstationen die Flüsse reguliert werden und ob diese funktionsfähig sind (automatisierte Prüfung von Anwendungskontrollen). Auf Basis dieser Kenntnisse kann der Prüfer seine Prüfungshandlungen auf unkontrollierte oder ungewöhnliche Flussläufe konzentrieren.

Eine konzeptionelle Darstellung unter der Verwendung der beschriebenen Analogie ist in Abbildung 4.2 dargestellt. In Tabelle 4.2 sind Beispiele für die visualisierten Prozessflüsse und Anwendungskontrollen enthalten.

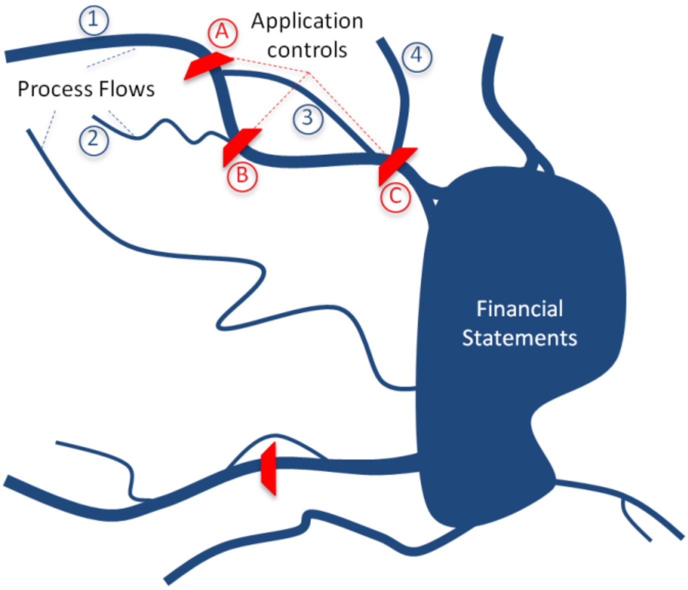


Abbildung 4.2 Prozesskarte aus [5]

Process flows	
#	Description
1	purchase order
2	goods receipt for purchase order
3	services receipt
4	incoming invoice

Application controls	
#	Description
A	system based approval for purchase order
B	Two-way-match
C	Three-way-match

Tabelle 4.2 Beispielprozesse und Application Controls aus [5]

Die gewählte Analogie verdeutlicht das Potential des Einsatzes von Methoden der Prozessrekonstruktion in der Jahresabschlussprüfung. Bei zunehmender Automatisierung der Transaktionsverarbeitung und Integration von Geschäftsprozessen in ERP Systeme stellt die Anwendung automatisierter und systembasierter Kontrollen eine effektive Möglichkeit dar, der Flut automatisierter Transaktionsabwicklungen mit automatischen Kontrollen zu begegnen. Es ist davon auszugehen, dass deren Anwendung eine neue Qualität der Ergebnisse der Prozessprüfung erreicht. Nicht mehr Erfahrung und manuelle Auswertung einzelner Kontrollen bestimmen die Qualität einzelner Prozessprüfungen, sondern die Korrektheit der Rekonstruktionsalgorithmen, die wissenschaftlich bewiesen werden können, sowie die Verfahren zur automatischen Prüfung von Anwendungskontrollen. Einschränkungen zu letzteren werden in Abschnitt fünf erläutert. Allerdings lässt sich festhalten, dass wesentliche Effektivitäts- und Effizienzgewinne insbesondere bei Prüfungen von hochintegrierten Geschäftsprozessen und hoher Transaktionsrate zu erwarten sind.

Einer der wesentlichen Vorteile des präsentierten Ansatzes ist die Systemunabhängigkeit der Methoden. Dies ist darin begründet, dass die Rekonstruktion der Prozesse nicht unmittelbar auf Basis von Verknüpfungen der Daten auf Datenbankebene basiert sondern nur mittelbar. Die Verknüpfung wird hergestellt über die offene-Posten-Buchhaltung rechnungslegungsrelevanter Buchungen. Die offene-Posten-Buchführung ist ein Erfordernis der Rechnungslegung und muss somit von allen Informationssystemen angeboten werden, die für Rechnungslegungszwecke eingesetzt werden. Eine Verknüpfung der Buchungen auf Datenbankebene muss aus diesem Grund gewährleistet sein. Auch wenn die Umsetzung der Verknüpfung technisch unterschiedlich ausfallen mag, kann davon ausgegangen werden, dass sie stets für die Rekonstruktion genutzt werden kann, denn das jeweilige Informationssystem muss auf solche Verknüpfungen zurückgreifen können, um eine offene-Posten-Buchhaltung zu gewährleisten.

4.3 Unterstützung von substantiellen Prüfungshandlungen

Neben der Automatisierung der Prüfung von Geschäftsprozessen können Methoden zur Prozessrekonstruktion eingesetzt werden, um dem Prüfer Hilfestellung zu substantiellen Prüfungshandlungen zu geben. Über die Rekonstruktion und Visualisierung von Prozessflüssen wird offensichtlich, ob Verzweigungen von Prozessflüssen existieren, die auf Ausnahmen von Standardprozessen schließen lassen. Mittels der Auswertung nicht nur der Anzahl von Transaktionen in einem Prozessfluss sondern auch über die Ermittlung der über die Transaktionen abgewickelten Beträge können einzelne Transaktionen mit hohen Beträgen und somit hoher Relevanz für den Jahresabschluss für dezidierte substantielle Prüfungshandlungen identifiziert werden.

Die Generierung von gezielten Stichproben z.B. auf Basis der Betragsgröße abgewickelter Geschäftsvorfälle kann unterstützt werden. Diese Unterstützung ist insbesondere relevant für Prozessflüsse, die nicht durch Anwendungskontrollen reguliert werden.

4.4 Analyse der Metadaten

JANS et al. [40] eröffnen einen interessanten Blickwinkel auf die Auswertung von Metadaten, die bei der Speicherung von Daten zu im System verarbeiteten Transaktionen erfolgen. Metadaten sind hierbei als Daten zu verstehen, die nicht unmittelbar Daten zum bearbeiteten Geschäftsvorfall betreffen, sondern die durch das System automatisch erzeugt und gespeichert werden. Dies beinhaltet z.B. Daten zum Benutzer, der die Transaktion durchgeführt hat, das Datum der Durchführung der Buchung im System (im Gegensatz zum vom Benutzer eingegebenen Buchungsdatum) und Änderungshistorien zu Datenobjekten. Bei manueller Bearbeitung eines Geschäftsvorfalls wären diese Informationen in der Regel nicht verfügbar.

Die Auswertung der Metadaten erlaubt, Prüfungen zu unterstützen, die bei manueller Prüfung nicht möglich wären. Zum Beispiel betrifft dies Abweichungen zum zeitlichen Verlauf von Prozessschritten und Manipulationen an Daten wie Preis oder Menge bestellter Waren.

Ein großer Vorteil bei der Auswertung dieser Daten liegt darin, dass die gespeicherten Daten vor Manipulation geschützt sind, sofern Zugriffsrechte für Administratorberechtigungen auf den relevanten Zugriffsebenen entsprechend eingeschränkt sind.

4.5 Funktionstrennung

Unter Funktionstrennung wird die Trennung kritischer Berechtigungen in ERP Systemen verstanden. Durch die Trennung von Funktionen soll erreicht werden, dass mehrere kritische Transaktionen nicht durch ein und denselben Benutzer, z. B. für wirtschaftskriminelle Tätigkeiten, unbemerkt verwendet werden können. Über die Funktionstrennung soll erreicht werden, dass ein Vier-Augen-Prinzip zu kritischen Geschäftsvorfällen eingehalten wird.

Für die Überprüfung von Funktionstrennung und zur Entdeckung von Funktionstrennungskonflikten existieren anwendungsspezifische Softwarelösungen. Mittlerweile werden entsprechende Funktionalitäten auch durch entsprechende Module verbreiteter Softwareanbieter angeboten. Diese Anwendungen erlauben jedoch lediglich zu analysieren, ob Funktionstrennungskonflikte vorliegen, ob also Berechtigungen für die Durchführung kritischer Transaktionen bei einzelnen Benutzern vorliegen. Sie erlauben aber in der Regel keine Auswertung, ob diese Berechtigungen tatsächlich von dem jeweiligen Benutzer auch eingesetzt wurden. Diese Prüfung muss manuell durchgeführt werden. Durch die Anwendung von Methoden zu Prozessrekonstruktion und Auswertung bestimmter Daten wie Benutzername und Transaktionscode wäre eine automatisierte Auswertung zu Funktionstrennungskonflikten bei tatsächlich eingetretenen Geschäftsvorfällen möglich.

4.6 Leistungsanalyse

Neben dem Einsatz für Prüfungszwecke könnte eine Anwendung zur Messung der Leistung von in ERP Systemen integrierten Geschäftsprozessen ermöglichen. HUFGARD [45] bietet eine umfangreiche Analyse über KPIs für SAP Systeme. Diese kann als Basis für die Entwicklung prozessbezogener und ERP-System-unabhängiger KPIs und Metriken für eine automatisierte Auswertung verwendet werden. Die notwendigen Daten könnten über die Erweiterung der Extraktionsalgorithmen ausgewertet werden. Darüber hinaus ergibt sich die Auswertungsmöglichkeit spezifischer KPIs wie z.B. die durchschnittliche Durchlaufzeit eines nebenläufigen Prozessflusses.

4.7 Abweichungsanalyse und Prozessoptimierung

Die Erstellung von Prozesskarten erlaubt, Abweichungen von Soll-Prozessen zu identifizieren. Sofern Soll-Prozesse definiert sind, können anhand visueller Auswertungen Abweichungen erkannt werden. Die Analyse von Abweichungen erlaubt, deren Ursachen zu evaluieren und ggf. Änderungen zur Optimierung am Prozessablauf vorzunehmen.

5 Restriktionen

In den vorherigen Abschnitten wurden Potential und Anwendungsmöglichkeiten der Methoden zur Rekonstruktion von Prozessen diskutiert. Dieser Abschnitt thematisiert ausgewählte Restriktionen zu deren Einsatz und zeigt wesentliche Anwendungsgrenzen auf.

5.1 Extraktionskomponente

Die Rekonstruktion der Prozessinstanzen erfolgt auf Basis der Verknüpfung der Buchungen, die sich aus der offene-Posten-Buchführung ergibt. Diese Verknüpfung ist fachlicher Natur und nicht systemspezifisch. Die Umsetzung der Verknüpfung erfolgt jedoch auf technischer

Ebene. Um die Rekonstruktion von Prozessinstanzen mit einem Softwareartefakt durchführen zu können, ist die Extraktion der relevanten Daten samt derer Verknüpfungen aus den jeweiligen zugrundeliegenden ERP Systemen notwendig.

Welche Daten für die Extraktion relevant sind, ist für jedes ERP System individuell zu spezifizieren. Aufgrund der Heterogenität der ERP Systeme und deren technischer Umsetzung ergibt sich die Restriktion, dass es keine generische Softwarekomponente für die Datenextraktion geben kann, zumindest solange keine Standards für eine relevante Datenbereitstellung etabliert sind. Über die Modularisierung und Entkopplung der Extraktionskomponente und der Rekonstruktionskomponente in einem Softwareartefakt lässt sich die Universalität der Rekonstruktionskomponente bewahren. Bei einer softwaretechnischen Umsetzung wird die Erstellung von ERP spezifischen Extraktionskomponenten allerdings nicht zu vermeiden sein.

Die gleiche Problematik ergibt sich bei der automatisierten Auswertung von Anwendungskontrollen. Auch hier wird die Entwicklung systemspezifischer Extraktionsmodule nicht vermieden werden können, da die Speicherung von Systemeinstellungen zu Anwendungskontrollen systemspezifisch erfolgt.

5.2 Systemversionen

Eine ähnliche Problematik ergibt sich aus der Versionierung der ERP Systeme. Bei jeder neuen Version eines ERP Systems ist prinzipiell zu prüfen, ob sich die zugrundeliegende Struktur der Speicherung der relevanten Daten geändert hat. Sofern dies der Fall ist, müssen die Extraktionskomponenten entsprechend angepasst werden.

5.3 Rechenkapazitäten

In ERP Systemen werden ggf. Informationen für Millionen von Transaktionen gespeichert. Der Einsatz von Rekonstruktionsmethoden zielt darauf ab, gerade bei hochintegrierten Geschäftsprozessen und hoher Transaktionsrate eingesetzt zu werden. Insofern ist mit großen Datenvolumina zu rechnen, die ausgewertet werden müssen. Einerseits erfordert dies den Einsatz leistungsfähiger Extraktionskomponenten. Andererseits werden die extrahierten Daten ausgewertet, aggregiert und als Graphen visualisiert. Hierbei ergibt sich die Anforderung, hinreichend leistungsfähige Algorithmen zu implementieren. Zum derzeitigen Entwicklungsstand ist unklar, ob die verwendeten Algorithmen in der Lage sind, bei begrenzten Rechnerkapazitäten hinreichend performante Berechnungen zu liefern. Des Weiteren stehen Untersuchungen aus, ob beliebig komplexe Prozessinstanzen bei endlicher Rechenkapazität ausgewertet und repräsentiert werden können.

5.4 Datenschutzaspekte

Einen kritischen Aspekt stellt der Datenschutz dar. In den vorausgegangenen Abschnitten wurden umfangreiche Einsatzmöglichkeiten der vorgestellten Methoden zu Auswertungszwecken erläutert. Die Sammlung und Auswertung von Informationen bietet die Möglichkeit zum Missbrauch. So könnten die erhaltenen Informationen z.B. unberechtigterweise für Leistungsmessung und Beurteilung der Mitarbeiter genutzt werden. Der Einsatz der Methoden ist auf Einsatzgebiete beschränkt, die im Einklang mit bestehenden Datenschutzbestimmungen stehen.

Datenschutzaspekte sollten bereits bei der Konzeptionierung der Softwareartefakte in Form von Anonymisierungs- oder Pseudonymisierungsverfahren berücksichtigt werden, um das Potential von Missbrauch bereits in der Entwicklungsphase zu minimieren.

5.5 Sachlogische Verknüpfung

Eine wesentliche Anwendungsgrenze für die in diesem Artikel vorgestellten Methoden zur Prozessrekonstruktion ergibt sich aus dem prinzipiell größten Vorteil der Methoden, der in der Systemunabhängigkeit der Konstruktionsmethoden liegt. Dieser ist begründet in der Nutzung der Verknüpfung von Buchungen durch die offene-Posten-Buchhaltung. Die Verwendung dieser Verknüpfung bedingt allerdings auch, dass lediglich Geschäftsprozesse und Teilprozesse ausgewertet werden können, die einer offenen-Posten-Buchhaltung folgen.

Für den Einsatz der Methoden für die Jahresabschlussprüfung stellt diese keine gravierende Restriktion dar, da ein Großteil der rechnungslegungsrelevanten Geschäftsprozesse auf offene-Posten-geführte Konten buchen. Hinsichtlich des Einsatzes der Methoden für Analyse- oder Optimierungszwecke stellt dies jedoch eine Einsatzgrenze dar, da hier ggf. Kernwertschöpfungsprozesse von Bedeutung sind, bei deren Durchführung keine Buchungen auf offene-Posten-geführten Konten stattfinden. Insofern stellt sich die Herausforderung, Verknüpfungsmerkmale für Transaktionen bei Geschäftsprozessen zu identifizieren, für die keine Buchung auf offene-Posten-geführte Konten stattfindet.

6 Zusammenfassung und Ausblick

In diesem Artikel wurden die Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen untersucht und diskutiert. Das wesentliche Anwendungsszenario ergibt sich für den Einsatz von Methoden zu Prozessrekonstruktion im Rahmen der Prüfung von Geschäftsprozessen bei Jahresabschlussprüfungen. In Kombination mit Methoden zur automatisierten Prüfung von Anwendungskontrollen bieten diese die Möglichkeit, die Diskrepanz zwischen einer automatisierten Transaktionsverarbeitung auf Seiten der Unternehmen mit systembasierten und automatisierten Prüfungsmethoden auf Seiten der Jahresabschlussprüfer zu begegnen. Insbesondere bei hochintegrierten Geschäftsprozessen und hoher Transaktionsrate ist mit einer wesentlichen Effektivitäts- und Effizienzsteigerung zu rechnen. Der Einsatz der Methoden erlaubt die Freisetzung und Nutzung von Ressourcen für die Prüfung ungewöhnlicher Geschäftsprozesse, die von Standardprozessen abweichen und ein inhärent höheres Risiko aufweisen. Neben dem Einsatz in der Jahresabschlussprüfung können Methoden zur Prozessextraktion Anwendung finden für Auswertungen auf Basis erzeugter Metadaten, die bei manueller Geschäftsprozessabwicklung außerhalb eines ERP Systems gar nicht zu Verfügung stehen. Es ergeben sich Einsatzfelder für die Auswertung zu Funktionstrennungsverletzungen, Leistungsanalysen der Prozesse sowie Abweichungs- und Optimierungsanalysen.

Neben vielfältigen Einsatzmöglichkeiten bestehen Restriktionen zum Einsatz der Methoden durch die notwendige Systemspezifität der Extraktionskomponenten, die sich aus der Heterogenität der ERP Systeme und unterschiedlicher Systemversionen ergeben. Ohne entsprechende Testläufe kann derzeit schwer beurteilt werden, ob die entwickelten Algorithmen leistungsfähig genug sind, um große Datenmengen auswerten und visualisieren zu können. Weitere Einschränkungen zum Einsatz können sich aus Datenschutzaspekten ergeben. Außerdem beschränken sich die Einsatzmöglichkeiten der Methoden bei jetzigem Forschungsstand auf Geschäftsprozesse, bei deren Bearbeitung offene-Posten-geführte Posten Konten involviert sind.

Zusammenfassend ist festzustellen, dass die Methoden zur Prozessrekonstruktion ein großes Potential für die Erreichung wesentlicher Effektivitäts- und Effizienzgewinne bei der Prüfung von in ERP Systeme integrierten Geschäftsprozessen aufweisen und darüber hinaus vielfältige weitere Einsatzmöglichkeiten bieten. Grenzen und Restriktionen für deren Einsatz existieren und müssen durch die Beachtung entsprechender Anforderungen bei der Umsetzung der Softwareartefakte berücksichtigt bzw. durch weitere Forschungstätigkeiten bewältigt werden.[35]

Literaturverzeichnis

[1] IDW PS 261 Feststellung und Beurteilung von Fehlerrisiken und Reaktionen des Abschlussprüfers auf die beurteilten Fehlerrisiken (Quelle: WPg 22/2006, S. 1433 ff., FN-IDW 11/2006, S. 710 ff., WPg Supplement 4/2009, S. 1 ff., FN-IDW 11/2009, S. 533 ff.) vom 09.09.2009

[2] ISA 315 Identifying and Assessing the Risks of Material Misstatement through Understanding in (Quelle: International Federation of Accountants (IFAC-I) (2008) Handbook of international auditing, assurance, and ethics pronouncements 2008 Edition Part I http://web.ifac.org/media/publications/2/2008-handbook-of-internatio/2008-handbook-of-internatio.pdf, abgerufen am 25.02.2010

[3] IDW PS 330 Abschlussprüfung bei Einsatz von Informationstechnologie (Quelle: WPg 21/2002, S. 1167 ff., FN-IDW 11/2002, S. 604 ff.) vom 24.09.2002

[4] IDW RS FAIT 1 Grundsätze ordnungsmäßiger Buchführung bei Einsatz von Informationstechnologie (Quelle: WPg 21/2002, S. 1157 ff., FN-IDW 11/2002, S. 649 ff.) vom 24.09.2002

[5] M. Werner, M. Gehrke, N. Nüttgens: “Business Process Mining and Reconstruction for Financial Audits”, in: Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS-45), Hawaii (2012 angenommen)

[6] Virtual Accounting Worlds (Quelle: www.virtualaccountingworlds.com, abgerufen am 07.12.2011

[7] J.E. Cook, and A.L. Wolf, “Discovering Models of Software Processes from Event-Based Data,” ACM Trans. Software Eng. and Methodology, vol. 7, no. 3, pp. 215-249, 1998.

[8] J.E. Cook and A.L. Wolf, “Event-Based Detection of Concurrency,” Proc. Sixth Int’l Symp. the Foundations of Software Eng. (FSE-6), pp. 35-45, 1998.

[35] Die in dieser Veröffentlichung präsentierten Ergebnisse wurden im Forschungsprojekt Virtual Accounting Worlds erarbeitet. Das Projekt wird vom Bundesministerium für Bildung und Forschung gefördert (Fördernummer 01IS10041). Die Autoren sind verantwortlich für den Inhalt der Veröffentlichung.

[9] J.E. Cook, and A.L. Wolf, “Software Process Validation: Quantitatively Measuring the Correspondence of a Process to a Model,” ACM Trans. Software Eng. and Methodology, vol. 8, no. 2, pp. 147-176, 1999.

[10] R. Agrawal, D. Gunopulos, and F. Leymann, “Mining Process Models from Workflow Logs,” Proc. Sixth Int’l Conf. Extending Database Technology, pp. 469-483, 1998.

[11] J. Herbst, “A Machine Learning Approach to Workflow Management,” Proc. 11th European Conf. Machine Learning, pp. 183-194, 2000.

[12] J. Herbst, “Dealing with Concurrency in Workflow Induction,” Proc. European Concurrent Eng. Conf., U. Baake, R. Zobel, and M. Al- Akaidi, eds., 2000.

[13] J. Herbst, “Ein induktiver Ansatz zur Akquisition und Adaption von Workflow-Modellen,” PhD thesis, Universität Ulm, Nov. 2001.

[14] J. Herbst, and D. Karagiannis, “An Inductive Approach to the Acquisition and Adaptation of Workflow Models,” Proc. Workshop Intelligent Workflow and Process Management: The New Frontier for AI in Business, M. Ibrahim and B. Drabble, eds., pp. 52-57, Aug. 1999.

[15] J. Herbst, and D. Karagiannis, “Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models,” Proc. Ninth Int’l Workshop Database and Expert Systems Applications, pp. 745-752, 1998.

[16] J. Herbst, and D. Karagiannis, “Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models,” Int’l J. Intelligent Systems in Accounting, Finance, and Management, vol. 9, pp. 67-92, 2000.

[17] M.K. Maxeiner, K. Küspert, and F. Leymann, “Data Mining von Workflow-Protokollen zur teilautomatisierten Konstruktion von Prozessmodellen” Proc. Datenbanksysteme in Büro, Technik und Wissenschaft, pp. 75-84, 2001.

[18] G. Schimm, “Generic Linear Business Process Modeling,” Proc. ER 2000 Workshop Conceptual Approaches for E-Business and The World Wide Web and Conceptual Modeling, S.W. Liddle, H.C. Mayr, and B. Thalheim, eds., pp. 31-39, 2000.

[19] G. Schimm, “Process Mining Elektronischer Geschäftsprozesse,” Proc. Elektronische Geschäftsprozesse, 2001.

[20] G. Schimm, “Process M6ning linearer Prozessmodelle - Ein Ansatz zur Automatisierten Akquisition von Prozesswissen,” Proc. 1. Konferenz Professionelles Wissensmanagement, 2001.

[21] G. Schimm, “Process Miner - A Tool for Mining Process Schemes from Event-Based Data,” Proc. Eighth European Conf. Artificial Intelligence (JELIA), S. Flesca and G. Ianni, eds., pp. 525-528, 2002.

[22] W. M. P. van der Aalst, „Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing“, Requirements Engineering Journal, Bd. 10, Nr. 3, S. pp. 198-211, 2005.

[23] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1. Aufl. Springer Berlin Heidelberg, 2011.

[24] W. M. P. van der Aalst, A. J. M. M. Weijters, und L. Maruster, „Workflow Mining: Which Processes can be Rediscovered“, presented at the Proc. Int’l Conf. Eng. and Deployment of Cooperative Information Systems (EDCIS 2002), 2002, Bd. 2480, S. 45-63.

[25] W.M.P. van der Aalst, “The Application of Petri Nets to Workflow Management,” The J. Circuits, Systems and Computers, vol. 8, no. I, pp. 21-66, 1998.

[26] W.M.P. van der Aalst, “Verification of Workflow Nets”, Application and Theory of Petri Nets, P. Azema and G. Balbo, eds., pp. 407-426, Berlin: Springer-Verlag, 1997.

[27] W.M.P. van der Aalst, and B.F. van Dongen, “Discovering Workflow Performance Models from Timed Logs,” Proc. Int’l Conf. Eng. and Deployment of Cooperative Information Systems (EDCIS 2002), Y. Han, S. Tai, and D. Wikarski, eds., vol. 2480, pp. 45-63, 2002.

[28] W. M. P. van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A.J.M.M. Weijters, “Workflow Mining: a Survey of Issues and Approaches”, Beta: Research School for Operations Management and Logistics, Working Paper 74, 2002.

[29] W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster, “Workflow Mining: Which Processes can be Rediscovered?” BETA Working Paper Series, WP 74, Eindhoven Univ. of Technology, Eindhoven, 2002.

[30] L. Maruster, W.M.P. van der Aalst, A.J.M.M. Weijters, A. van den Bosch, and W. Daelemans, “Automated Discovery of Workflow Models from Hospital Data,” Proc. 13th Belgium-Netherlands Conf. Artificial Intelligence (BNAIC 2001), B. Kro¨se, M. de Rijke, G. Schreiber, and M. van Someren, eds., pp. 183-190, 2001.

[31] L. Maruster, A.J.M.M. Weijters, W.M.P. van der Aalst, and A. van den Bosch, “Process Mining: Discovering Direct Successors in Process Logs,” Proc. Fifth Int’l Conf. Discovery Science (Discovery Science 2002), pp. 364-373, 2002.

[32] A.J.M.M. Weijters, and W.M.P. van der Aalst, “Process Mining: Discovering Workflow Models from Event-Based Data,” Proc. 13th Belgium-Netherlands Conf. Artificial Intelligence (BNAIC 2001), B. Kro¨ se, M. de Rijke, G. Schreiber, and M. van Someren, eds., pp. 283-290, 2001.

[33] A.J.M.M. Weijters and W.M.P. van der Aalst, “Workflow Mining: Discovering Workflow Models from Event-Based Data,” Proc. ECAI Workshop Knowledge Discovery and Spatial Data, C. Dousson, F. Höppner, and R. Quiniou, eds., pp. 78-84, 2002.

[34] A.J.M.M. Weijters, and W.M.P. van der Aalst, “Rediscovering Workflow Models from Event-Based Data,” Proc. 11th Dutch-Belgian Conf. Machine Learning (Benelearn 2001), V. Hoste and G. de Pauw, eds., pp. 93-100, 2001.

[35] N. Gehrke, “The ERP AuditLab - A prototypical Framework for Evaluating Enterprise Resource Planning System Assurance”, in: Proceedings of the 43th Hawaii International Conference on System Sciences (HICSS-43), Hawaii (2010).

[36] N. Gehrke, and N. Müller-Wickop, "Basic Principles of Financial Process Mining", Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010.

[37] N. Gehrke, and N. Müller-Wickop, "Rekonstruktion von Geschäftsprozessen im Finanzwesen mit Financial Process Mining", in: Lecture Notes in Informatics, Proceedings der Jahrestagung Informatik 2010, IT-supported Service Innovation and Service Improvement, Leipzig, 2010.

[38] Gehrke, N., Wolf, P.: Towards Audit 2.0 - A Web 2.0 Community Platform for Auditors, in: Proceedings of the 43th Hawaii International Conference on System Sciences (HICSS-43), Hawaii (2010).

[39] Müller-Wickop, N., Schultz, M., Gehrke, N., Nüttgens, M.: Towards Automated Financial Process Auditing: Aggregation and Visualization of Process Models, Proceedings of the Enterprise Modelling and Information Systems Architectures (EMISA 2011), Germany, Hamburg, 2011

[40] M. Jans, M. Alles, und M. Vasarhelyi, „Process mining of event logs in auditing: opportunities and challenges", Working paper. Hasselt University. Belgium, 2010.

[41] M. Jans, N. Lybaert, K. Vanhoof, und J. M. Van Der Werf, „Business process mining for internal fraud risk reduction: Results of a case study", 2008.

[42] M. Alles, M. Jans, und M. Vasarhelyi, „Process Mining: A New Research Methodology for AIS", in CAAA Annual Conference 2011, 2011.

[43] K. Küting, and M. Reuter, "Bilanzierung im Spannungsfeld unterschiedlicher Adressaten", DSWR, Vol. 9, 2004.

[44] P. Leibfried, and P. Meixner, "Konvergenz der Rechnungslegung: Bestandsaufnahme und Versuch einer Prognose", Der Schweizer Treuhänder, Vol. 4, 2006.

[45] A. Hufgard, ROI von SAP-Lösungen verbessern, 1. Aufl. SAP PRESS, 2010.

10.4 Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung

Number	**4**
Title	**Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung**
Appendix	10.4
Primary Related Chapters	4.2, 5.1
Type	Book Chapter
Book	Forschung für die Wirtschaft
Reference	(Werner, 2012)
Acceptance Rate	[36]
VHB JQ 2.1 Ranking	-
WKWI Ranking	-
ERA 2010	-
CORE 2013	-
Review Procedure	Blinded
Number of Reviews	-
Authors	Michael Werner
Dissertation Points	1.00
Authorship	
Overall	100%
Design	100%
Realization	100%
Writing	100%
Status	Published
Part of other Dissertations	No
Link	https://cuvillier.de/de/shop/publications/6274-forschung-fur-die-wirtschaft-2012

[36] Invited for Publication

Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung

MICHAEL WERNER

NORDAKADEMIE – Hochschule der Wirtschaft, Elmshorn

Abstract: Die Automatisierung der Durchführung von Geschäftsprozessen durch den Einsatz von Informationssystemen erlaubt es Unternehmen, ihre Prozesse effizienter und effektiver zu gestalten. Mit der Zunahme der Automatisierung steigt die Verfügbarkeit von Daten, die im Zuge der Bearbeitung durch Informationssysteme gespeichert werden. Business Intelligence Tools bieten Funktionalitäten, um aus den immer stärker anwachsenden Datenbeständen Informationen effizient zu extrahieren und in geeigneter Weise zur Verfügung zu stellen. Während der Einsatz von Business Intelligence Software in den vergangenen Jahren mit der gleichzeitigen Verfügbarkeit digital auswertbarer Daten in vielen Industriezweigen zugenommen hat, ist dies für die Prüfungsbranche nicht oder nur in einem geringen Ausmaß zu beobachten gewesen. Wirtschaftsprüfer werden mit der Prüfung und Testierung der Jahresabschlüsse von Unternehmen beauftragt. Ein wesentlicher Bestandteil der Jahresabschlussprüfung ist die Prüfung der Geschäftsprozesse, die für die Finanzberichterstattung relevant sind. Für diese Prüfung werden durch den Wirtschaftsprüfer vorwiegend manuelle Tätigkeiten durchgeführt. Process Mining Methoden bieten Möglichkeiten, aus vorhandenen Event Logs Prozessmodelle zu rekonstruieren. Der Einsatz dieser Methoden eignet sich auch für die Prüfung von Geschäftsprozessen, indem Prozessmodelle automatisch rekonstruiert und analysiert werden können. In diesem Beitrag wird veranschaulicht, wie spezielle Process Mining Methoden auf die Anwendungsdomäne der Jahresabschlussprüfung angewendet werden können.

1. Einleitung

Der Zweck von Business Intelligence[37] besteht darin, umfangreiche Daten zu analysieren und somit Informationen zugänglich zu machen, die andernfalls nicht zur Verfügung stehen würden. Die Bedeutung von Business Intelligence hat in den vergangenen Jahren mit dem Anwachsen auswertbarer, digital vorliegender Datenbestände stark zugenommen und Eingang in Softwarewerkzeuge bedeutender Anbieter gefunden.[38] Je stärker die Unterstützung und Automatisierung der Verarbeitung von Geschäftsvorfällen in Informationssystemen voranschreitet, desto wichtiger ist es, die verfügbaren Datenstände für Analysen auswerten zu können. Primäres Anwendungsfeld von Business Intelligence Techniken ist die Unterstützung von Entscheidungsprozessen [5], [6]. Sie eignen sich allerdings auch zur Analyse von Geschäftsprozessen für die Zwecke der Jahresabschlussprüfung.

[37] Business Intelligence umfasst „Anwendungen und Techniken, die sich darauf konzentrieren, Daten aus verschiedenen Quellen zu sammeln, zu speichern, zu analysieren und den Zugriff auf sie zu ermöglichen, um den Benutzern zu helfen, bessere Entscheidungen zu treffen.“[1] S. 503.

[38] Gängige Softwarelösungen werden zum Beispiel von IBM [2], Oracle [3] und SAP [4] angeboten.

Die Jahresabschlussprüfung ist ein Kontrollinstrument, das dazu dient, die Adressaten des Jahresabschlusses vor Falschinformationen zu schützen. Sie stellt ein wichtiges Regulativ im Rahmen der Wirtschaftsordnung westlicher Volkswirtschaften dar. Aufgrund ihrer Bedeutung werden an die Durchführung der Prüfung und die Qualifikation der Prüfer hohe Anforderungen gestellt, die durch den Gesetzgeber kodifiziert sind.[39] Des Weiteren werden die Anforderungen an die Durchführung der Prüfung in Prüfungsstandards konkretisiert.[40] Eine bedeutende Tätigkeit während der Abschlussprüfung ist die Prüfung von Geschäftsprozessen. Den Buchungseinträgen auf den Konten der Bilanz und Gewinn- und Verlustrechnung liegen Geschäftsvorfälle zu Grunde, die im Unternehmen abgewickelt wurden. Die Prüfung der Geschäftsprozesse soll sicherstellen, dass nur vollständige und richtige Einträge auf den entsprechenden Buchhaltungskonten gebucht werden, die auch auf tatsächlich stattgefundenen Geschäftsvorfällen basieren. Die Prüfung von Geschäftsprozessen wird vorgenommen aufgrund der Annahme, dass wohlkontrollierte Geschäftsprozesse zur vollständigen und richtigen buchhalterischen Erfassung der bearbeiteten Geschäftsvorfälle führen. Von besonderer Bedeutung für Prüfung von Geschäftsprozessen sind die International Standards on Auditing 315 [9] und 330 [10] sowie auf nationaler Ebene der IDW Prüfungsstandard 261 [11]. In diesen wird festgelegt, wie die Prüfung von Geschäftsprozessen, internen Kontrollen und relevanten Informationssystemen zu berücksichtigen ist.

Bei der Durchführung der Prüfung von Geschäftsprozessen stehen die Wirtschaftsprüfer vor der Herausforderung, die Zuverlässigkeit der Kontrollmechanismen zu den betrachteten Geschäftsprozessen zu würdigen. Dazu ist es notwendig, ein der Realität entsprechendes Verständnis der zugrundeliegenden Geschäftsprozesse sowie der involvierten internen Kontrollen zu erhalten und letztere auf deren Angemessenheit und operative Funktionsfähigkeit zu prüfen. Dies erfolgt zumeist anhand manueller Prüfungstätigkeiten. Diese umfassen die Durchführung von Interviews oder die Inspektion vorhandener Dokumente auf Stichprobenbasis. Solche Prüfungshandlungen sind im Umfeld stark integrierter und automatisierter Verarbeitung durch Informationssysteme kritisch zu hinterfragen. Werner und Gehrke [12] zeigen, dass mit steigender Automatisierung der Geschäftsprozessabwicklung und der Anzahl der zu verarbeiteten Geschäftsvorfälle manuelle Prüfungstätigkeiten ineffizient und ggf. ineffektiv werden.

Eine Lösung für dieses Problem stellt der Einsatz von Business Intelligence Methoden für die Zwecke der Jahresabschlussprüfung dar. Ganzheitliche konzeptionelle Überlegungen für den Einsatz von Business Intelligence Methoden zur Unterstützung und Teil-automatisierung der Geschäftsprozessprüfung finden sich in [13] und [14]. In diesem Beitrag wird auf einen konkreten Teilaspekt des konzeptionellen Entwurfs von Werner et al. [13] eingegangen, indem illustriert wird, wie Process Mining Methoden als Analysemethoden für die Aufbauprüfung von Geschäftsprozessen eingesetzt werden können.

[39] §319 HGB regelt die Zuständigkeit der Prüfung des Jahresabschlusses und legt fest, dass diese ausschließlich durch Wirtschaftsprüfer, Wirtschaftsprüfungsgesellschaften, vereidigte Buchprüfer oder Buchprüfungsgesellschaften durchgeführt werden darf.

[40] Auf nationaler Ebene übernimmt das Institut der Wirtschaftsprüfer in Deutschland e.V. (IDW) [7] als Vereinigung von Wirtschaftsprüfern und Wirtschaftsprüfungsgesellschaften die fachliche Entwicklung der Regeln zu deren Berufsausübung und veröffentlicht in diesem Zusammenhang anzuwendende Prüfungsstandards. Auf internationaler Ebene übernimmt diese Aufgabe das International Auditing and Assurance Standards Board (IAASB)[8] mit der Veröffentlichung der International Standards on Auditing (ISA).

Ziel der Aufbauprüfung als Bestandteil der Geschäftsprozessprüfung ist die Informationsgewinnung über die für die Finanzberichterstattung relevanten Prozesse und Kontrollen. Process Mining Methoden können eingesetzt werden, um durchgeführte Geschäfts-vorfälle als Modelle von Prozessinstanzen zu visualisieren und somit Informationen über die Ausgestaltung der beteiligten Prozesse zu erhalten. Bei der Anwendung dieser Methoden ist die Berücksichtigung von Anforderungen von großer Bedeutung, die sich aus der Anwendungsdomäne ergeben.

Das Vorgehen sowie die Ergebnisse, die sich durch die Anwendung ausgewählter Methoden erzielen lassen, werden in diesem Beitrag erläutert. Die hier vorgestellten Methoden können dabei zum derzeitigen Forschungstand nur als Meilenstein betrachtet werden auf dem Weg zur Entwicklung ganzheitlicher automatisierter Analyse- und Prüfungsmethoden für die Geschäftsprozessprüfung. Insofern liegt ein weiterer Schwerpunkt auf der Betrachtung, welche Hinweise sich für weitere Entwicklungsmöglichkeiten aus den dargestellten Ergebnissen ableiten lassen.

In den folgenden Abschnitten zwei und drei werden der derzeitige Stand der Wissenschaft zu den relevanten Themengebieten sowie der gewählte Forschungsansatz und Methoden vorgestellt, die zur Erlangung der dargestellten Erkenntnisse herangezogen wurden. Um die Bedeutung des Einsatzes der Process Mining Techniken für die Zwecke der Aufbauprüfung von Geschäftsprozessen zu veranschaulichen, bietet Abschnitt vier einen Überblick über die Bedeutung der Analyse von Geschäftsprozessen für die Jahres-abschlussprüfung. Für die Modellierung von Modellen ist es notwendig, eine angemessene Modellierungssprache und Repräsentationsform zu wählen. Die hier dargestellten Modelle werden als Petri-Netze modelliert. Hintergrund für diese Wahl sowie formale Aspekte werden in Abschnitt fünf erörtert. In Abschnitt sechs werden schließlich bisher erzielte Auswertungsergebnisse dargestellt. Der Beitrag endet mit einer Zusammenfassung und Diskussion der dargestellten Ergebnisse.

2. Stand der Wissenschaft

Ausgangspunkt der in diesem Beitrag dargelegten wissenschaftlichen Untersuchung sind die Forschungsarbeiten aus dem Gebiet Process Mining. Erste Ansätze hierzu entstanden zum Ende der 1990er Jahre durch Cook und Wolf [15–17] ursprünglich im Bereich des Software Engineering. Seit Beginn der 2000er Jahre hat sich die Forschung auf diesem Gebiet deutlich intensiviert, was zu einer Vielzahl von Veröffentlichungen geführt hat. Einen recht aktuellen Überblick hierzu bietet Tiwari [18]. Ein bedeutender Vertreter dieser Disziplin ist van der Aalst. In [19] bietet er einen umfassenden Überblick über Methoden und wesentliche Grundkenntnisse zum Process Mining.

In den letzten Jahren ist zu beobachten, dass sich wissenschaftliche Arbeiten vermehrt mit einer Ausweitung des Anwendungskontextes befassen. Dies gilt auch für das An-wendungsgebiet von Process Mining Techniken für Prüfungszwecke. Hierbei ist zu unterscheiden zwischen Arbeiten, die sich mit Conformance und solchen, die sich mit Compliance beschäftigen. Erstgenannte Untersuchungen haben zum Ziel, Abweichungen von rekonstruierten Geschäftsvorfällen anhand eines Soll-Modells zu identifizieren. Hierzu werden verfügbare Event Logs gegen Sollmodelle verglichen und Abweichungen identifiziert. Dieser Ansatz lässt sich unter anderem finden in [20]. Er ist interessant, für die in diesem Beitrag dargestellten Vorgehensweisen aber weitestgehend irrelevant. Die Anwendung von Conformance Checking Methoden setzt das Vorhandensein eines Sollmodells voraus. Beim Einsatz der in diesem Beitrag dargestellten Methoden geht es aber gerade darum, Prozessmodelle anhand verfügbarer Daten erst zu entdecken und nicht rekonstruierte Modelle gegen ein bereits existierendes Sollmodell zu vergleichen.

Anders verhält es sich mit Forschungsarbeiten, die sich mit Compliance-Aspekten beschäftigen. Unter Compliance wird allgemein die Einhaltung von internen oder externen Vorgaben verstanden. Bei der Jahresabschlussprüfung geht es darum, festzustellen, ob externen Vorgaben in Form von Gesetzen und Normen im Zuge der Rechnungslegung entsprochen wurde. Insofern stellt die Prüfung der Einhaltung dieser Vorgaben einen Teilbereich der Gesamtthematik dar, die unter dem Begriff Compliance zusammengefasst wird. Interessante Ansätze für die Überprüfung der Einhaltung von Compliance-Regeln werden in [21–24] gezeigt. Ihnen ist allerdings gemein, dass sie nicht speziell auf das Anwendungsfeld der Jahresabschlussprüfung ausgerichtet sind. In diesem Anwendungsfeld ist insbesondere die Berücksichtigung der Werteflüsse innerhalb der Prozesse notwendig. Diese werden bei den zuvor genannten Arbeiten nicht explizit berücksichtigt.

Speziell für den Anwendungsbereich entwickelte Methoden präsentieren hingegen Gehrke und Müller-Wickop [25]. Sie stellen einen Mining Algorithmus vor, der in der Lage ist, Modelle von Prozessinstanzen zu rekonstruieren und dabei für die Prüfung relevante Informationen zu modellieren. Auf diese Methoden wird im Folgenden zurückgegriffen.

3. Methodik

Die Forschungstätigkeiten, die dieser Arbeit zugrunde liegen, folgen einem Design Science Ansatz [26–28]. Bei dieser gestaltungsorientierten Herangehensweise liegt der Fokus auf der Entwicklung von Artefakten. Der Erkenntnisgewinn ergibt sich als Ergebnis des Gestaltungsprozesses und der Evaluation der geschaffenen Artefakte. Primäre Artefakte bilden in dieser Arbeit die Methoden, die zur Extraktion relevanter Daten sowie zur Rekonstruktion von Modellen und deren Visualisierung entwickelt wurden.

Im Gegensatz zu behavioristischen Ansätzen wird gestaltungsorientierten Vorgehens-weisen häufig eine fehlende Rigorosität vorgeworfen. Riege et al. [29] weisen in diesem Zusammenhang auf die Notwendigkeit der Evaluation der Erkenntnisse sowohl gegen die identifizierte Forschungslücke wie auch gegen die Realität hin, um eine ausreichende Aussagekraft der Erkenntnisse nachweisen zu können. Um diesen Überlegungen Rechnung zu tragen, wurden für die vorliegende Forschungsarbeit verschiedene Evaluationsmethoden verwendet.

Abbildung 1 stellt die verschiedenen Phasen des Forschungsprozesses in Anlehnung an [28] dar. Für die Analysephase wurden Experteninterviews durchgeführt, um generelle Anforderungen für den Einsatz von Process Mining Methoden zu identifizieren. Primäre Quelle für die Identifikation von Anforderungen bildeten die Recherche relevanter Literatur und insbesondere die Sichtung von Standards und Normen zur Rechnungslegung und Jahresabschlussprüfung.

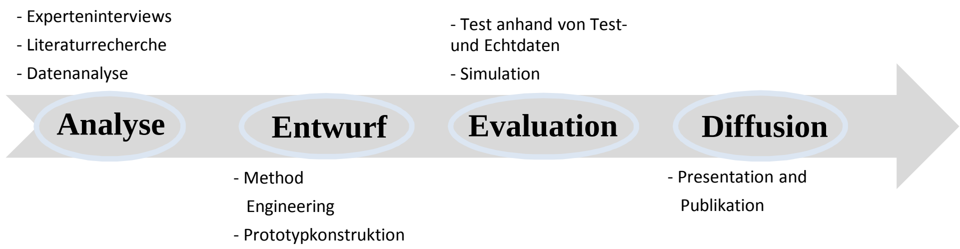


Abbildung 1 Phasen des Forschungsprozesses

Für die technische Ebene der Entwicklung wurden vorhandene Datenbestände ausgewertet. Die Entwicklung der Methoden erfolgte in Form eines Method Engineering [30], indem bereits bestehende Methoden kombiniert wurden, wenn hiermit das gewünschte Ergebnis erzielt werden konnte. Wenn notwendig wurden zudem neuartige Methoden entwickelt, die prototypisch in einem Softwareartefakt implementiert wurden. Zur Evaluation wurden diese anhand von Test- und Echtdaten getestet. Die Prozessmodelle, die mit der Software erstellt werden können, wurden wiederum durch die Simulation der Erreichbarkeit relevanter Schaltzustände auf ihre Korrektheit hin überprüft.

Bei dieser Vorgehensweise wurde der Forschungsprozess nicht einmalig durchlaufen sondern in mehrfachen Iterationen, in denen immer wieder Analyse, Entwurf, Evaluation und Diffusion, letzteres in Form von Publikationen und Vorträgen, aufeinander aufbauten.

4. Geschäftsprozessprüfung in der Jahresabschlussprüfung

Um die Einsatzmöglichkeiten von Process Mining Methoden für Zwecke der Jahresabschlussprüfung bewerten zu können, ist es notwendig, die Bedeutung der Geschäftsprozessprüfung innerhalb der Jahresabschlussprüfung zu beleuchten.

Anhand der Abbildung 2 lässt sich veranschaulichen, warum das Prozessverständnis für die Prüfung wichtig ist. Es stellt die typischen Phasen einer Jahresabschlussprüfung dar, die nach der Auftragsannahme im Allgemeinen mit der Erhebung relevanter Informationen und einer Beurteilung wesentlicher Risiken beginnt. Ziel der Prüfung ist es insgesamt, wesentliche Risiken für Fehler in der Finanzberichterstattung zu identifizieren und diese durch gezielte Prüfungshandlungen zu adressieren. Ein wesentliches Risiko besteht in der Regel darin, dass getätigte Buchungen nicht vollständig und richtig sind. Um dieses Risiko zu adressieren, ist es sinnvoll, die Prozesse zu prüfen, die zu den jeweiligen Buchungen geführt haben. Hierbei wird ausgenutzt, dass Kontrollaktivitäten[41] in den Prozessabläufen die korrekte Verarbeitung und somit auch Buchung sicherstellen. Wenn diese Kontrollen einwandfrei funktionieren, kann davon ausgegangen werden, dass auch die zugehörigen Buchungen vollständig und fehlerfrei sind.

[41] Ein häufig angeführtes Beispiel für eine solche Kontrolle ist der sogenannte Three-Way-Match. Bevor eine Zahlung veranlasst werden kann, wird hierbei überprüft, ob Menge und Wert bei Bestellung, Warenlieferung und eingegangener Rechnung übereinstimmen. Wenn dies nicht der Fall ist, wird die Zahlung verhindert.

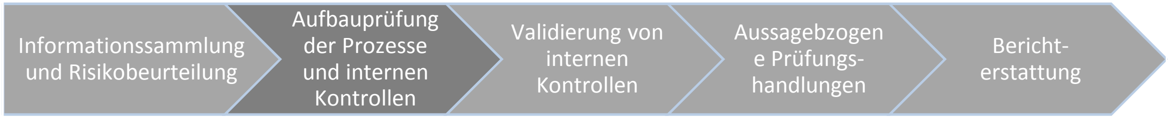


Abbildung 2 Prozessphasen der Jahresabschlussprüfung

Um auf die Prüfung der internen Kontrollen zurückgreifen zu können, muss der Prüfer zunächst wissen, wie die Prozesse aussehen, um beurteilen zu können, welche Kontrollen denn für eine Prüfung in Frage kommen. Dieser Zusammenhang wird aus Abbildung 2 deutlich. In der sogenannten Aufbauprüfung verschafft sich der Prüfer einen Überblick über die relevanten Prozesse und internen Kontrollen. Erst wenn dies erfolgt ist, kann die Überprüfung der identifizierten Kontrollen erfolgen, wenn feststeht, welche Kontrollen überhaupt relevant sind.

In Abbildung 3 ist das einfache Modell eines Bestellprozesses dargestellt. Die Rechtecke symbolisieren Aktivitäten, die in einem Informationssystem ausgeführt werden. Wenn dies erfolgt, erzeugen sie Buchungen auf bestimmten Konten der Finanzbuchhaltung. Diese sind rechts und links der Aktivitäten dargestellt. Gemäß den Regeln der doppelten Buchführung erfolgt für jede Buchung eine Soll- und eine Habenbuchung, so dass der Saldo der Buchung null ist. Wenn in einem Unternehmen Waren bestellt werden, hat dies zunächst keine Buchung auf Konten der Bilanz oder der Gewinn- und Verlustrechnung (GuV) zur Folge. Dies geschieht erst mit dem Eingang der bestellten Ware. Dieser Vorgang führt zu einer Erhöhung der Rohstoffbestände und zu einer Gegenbuchung auf dem Wareneingangs-/Rechnungseingangskonto (WeRe-Konto). Erst wenn die Rechnung des Zulieferers dem Unternehmen zugeht, entsteht eine Verbindlichkeit, die auf dem Kreditorenkonto für den entsprechenden Lieferanten gebucht wird.

Die Gegenbuchung findet wiederum auf dem WeRe-Konto statt und gleicht die zugehörige Haben-Buchung aus dem Wareneingang aus. Als nächste Aktivität findet die Zahlung statt, mit der die Verbindlichkeit an den Kreditor beglichen und in diesem Fall das Bankkonto belastet wird.

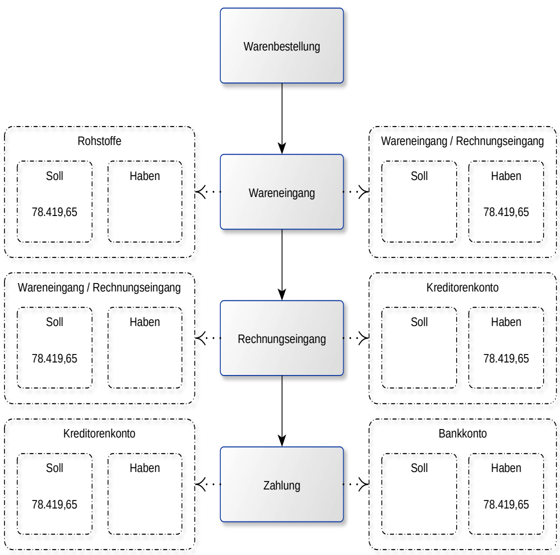


Abbildung 3 Einfaches Prozessmodell

Bei dem dargestellten Beispiel handelt es sich um einen sehr einfachen Standardprozess. Es illustriert aber, worauf es bei der Prüfung von Geschäftsprozessen in der Jahresabschlussprüfung ankommt. Für den Prüfer ist es wichtig zu verstehen, wie die Prozesse des Unternehmens mit den Buchungen auf den Finanzbuchhaltungskonten zusammen-hängen. In diesem Beispiel ist der Zusammenhang klar ersichtlich. Es ist zudem nach-vollziehbar, welche Werte durch den hier dargestellten Geschäftsvorfall auf die jeweiligen Konten geflossen sind.

Es stellt sich nun die Frage, wie der Prüfer Informationen über die Prozessabläufe und die Zusammenhänge mit den Buchhaltungskonten für die Aufbauprüfung erhält und welche Vorteile mit dem Einsatz von Process Mining Methoden in dieser Prüfungsphase erzielt werden können. Die herkömmliche Vorgehensweise besteht in der Durchführung von Interviews mit Ansprechpartnern beim zu prüfenden Mandanten. Auf Basis der so erhaltenen Informationen werden meist einfache Prozessdarstellungen in Form von Flussdiagrammen wie in Abbildung 3 erstellt, die gegebenenfalls durch textuelle Beschreibungen erweitert werden. Bestenfalls werden diese durch Prozessschreibungen, die beim Mandanten verfügbar sind, ergänzt.

Werner und Gehrke [12], [13] erläutern, dass dieses Vorgehen ineffizient und eventuell ineffektiv wird, wenn die Integration von Geschäftsprozessen und Informationssystemen zunimmt. Mit der Zunahme der Integration erhöht sich in der Regel die Komplexität der Prozesse. Des Weiteren können Aktivitäten vollständig automatisiert ablaufen, die für die operativ beteiligten Personen nicht mehr sichtbar sind, so dass sie auch keine verlässliche Auskunft über den tatsächlichen Prozessablauf geben können. Bestenfalls kann mit dieser manuellen Erhebungsweise ein Sollmodell erstellt werden. Ob dieses mit der Realität übereinstimmt, lässt sich nicht beurteilen. Erfahrungen aus der Praxis zeigen, dass dies in der Regel nicht der Fall ist.

Vor diesem Hintergrund können Process Mining Techniken eingesetzt werden, um au-tomatisiert Kenntnisse über den Zusammenhang von Prozessen und Buchungen auf Finanzbuchhaltungskonten zu erhalten. Der Einsatz dieser Techniken erlaubt es, anhand der tatsächlich stattgefundenen Geschäftsvorfälle und deren zugehörigen Daten Prozessmodelle zu erstellen. Damit besteht keine Abhängigkeit mehr zwischen der Verlässlichkeit der erhalten Informationen und der Darstellung der zu prüfenden Informationen. Des Weiteren erlaubt die automatisierte Erstellung eine wesentlich detaillierte Modellierung und bietet die Möglichkeit für umfassende Auswertungen.

5. Modellierung und Petri-Netz Darstellung

Die im vorherigen Abschnitt genannten Einsatzmöglichkeiten sollen nun an einem Bei-spiel illustriert werden. Für die Erstellung von Modellen spielt die Zweckmäßigkeit der gewählten Modellierungssprache eine wesentliche Rolle. Für die in dieser Arbeit dargestellten Prozessmodelle wurde die Modellierung in Form von gefärbten Petri-Netzen gewählt. Petri-

Netze stellen eine formal robuste und mathematisch ausdrucksstarke Modellierungssprache dar [31], [32]. Zudem kann durch die Modellierung von Petri-Netzen auf bereits umfangreiche Arbeiten zurückgegriffen werden.[42]

Abbildung 4 zeigt das Modell einer Prozessinstanz. Es basiert auf Daten, die aus einem SAP Testsystem extrahiert und für die Rekonstruktion verwendet wurden. Für die Extraktion und Rekonstruktion ist die Verwendung eines geeigneten Mining-Algorithmus notwendig. Es gibt unterschiedliche Verfahren wie etwa den Alpha-Algorithmus, heuristische Verfahren und genetische Algorithmen. Grundsätzlich sind diese Algorithmen geeignet, um Prozessmodelle in verschiedenen Anwendungsdomänen zu rekonstruieren.

Bei der Rekonstruktion von Prozessmodellen für den Zweck der Jahresabschlussprüfung kommt es jedoch darauf an, die Wertflüsse dazustellen und das Buchungsverhalten nachvollziehen zu können. Dies muss durch den einzusetzenden Mining-Algorithmus berücksichtigt werden. Zudem operieren die zuvor genannten Algorithmen auf Event Logs, die unterschiedliche Geschäftsvorfälle (Cases) enthalten. Die rekonstruierten Prozessmodelle bilden somit eine Abstraktion der im Event Log abgebildeten Prozessinstanzen. Hierbei kann es zu Unter- wie auch Überdeckungen in der Prozessdarstellung kommen. D.h. es werden unter Umständen Prozessabläufe modelliert, die gar nicht stattgefunden haben, oder es werden ggf. existente Prozessabläufe nicht abgebildet. Aus Sicht der Jahresabschlussprüfung sind lediglich solche Informationen wünschenswert, die den tatsächlichen Verlauf exakt widergeben. Der von Gehrke und Müller-Wickop [25], [33] entwickelte Algorithmus erfüllt die zuvor genannten Anforderungen und wurde aus diesem Grund für die vorliegende Arbeit verwendet und erweitert.

Der ursprüngliche Algorithmus stellt primär die Funktionalität dar, die Einträge im Event Log einzelnen Prozessinstanzen (oder auf Log Ebene Cases) zuzuordnen und die Prozessinstanzen in einer mehr oder weniger informellen Darstellungsweise zu visualisieren. Für ersteres wird die Struktur von offene-Posten-geführten Buchungen verwendet [13]. Der Mining-Algorithmus wurde dahingehend erweitert, dass er nun in der Lage ist, die rekonstruierten Prozessmodelle als Petri-Netze und somit in einer formalen Darstellungsform abzubilden.

[42] Tiwari et al. [18] zeigen, dass die Mehrheit der von ihnen untersuchten Artikel zum Thema Process Mining Petri-Netze als Modellierungssprache einsetzten.

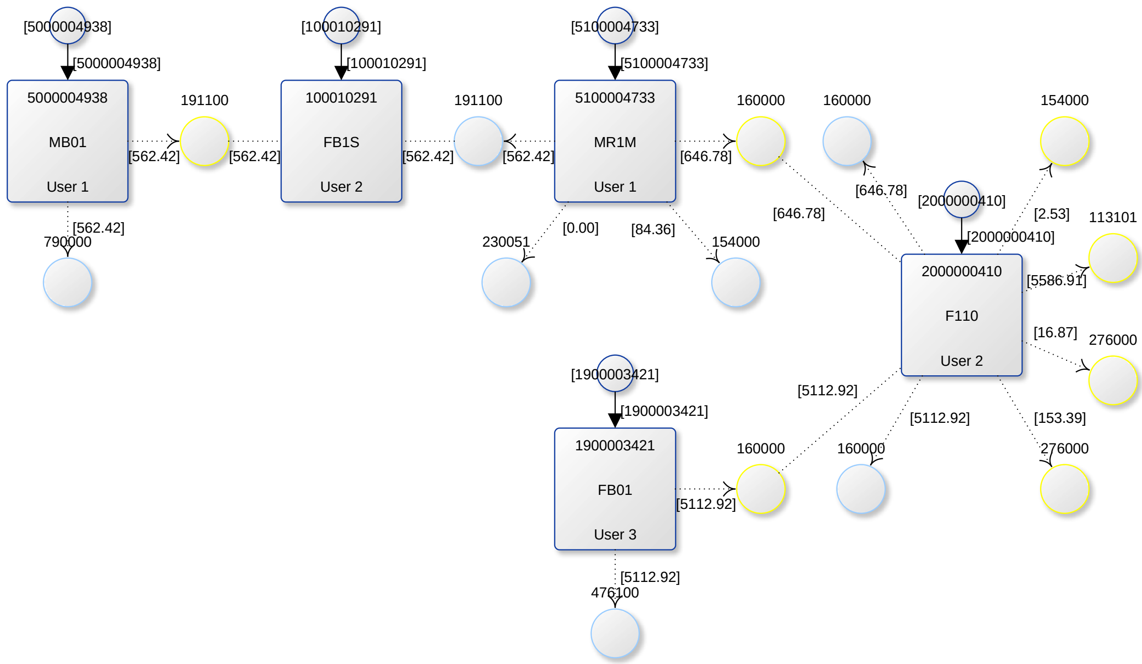


Abbildung 4 Beispiel einer rekonstruierten Prozessinstanz

Ein Petri-Netz besteht grundsätzlich aus Transitionen und Stellen, die über Kanten mit-einander verbunden sind. Dies wird aus Abbildung 4 deutlich. Die Transitionen (Recht-ecke) repräsentieren die Transaktionen, die im zugrundeliegenden ERP System ausgeführt wurden. Wenn Transaktionen in einem ERP System ausgeführt werden, erzeugen diese Buchungseinträge. Diese Buchungseinträge werden im Petri-Netz als gefärbte Plätze (Kreise) dargestellt. Die Färbung der Plätze ergibt sich aus der Kontonummer des Kontos, auf dem die Buchung stattgefunden hat sowie der Information, ob es sich um eine Soll- oder Habenbuchung gehandelt hat. Eine gestrichelte Kante mit Pfeil zeigt an, dass eine Transaktion die entsprechende Buchung erzeugt hat. Gestrichelte Kanten ohne Pfeil zeigen, dass eine Buchung durch eine weitere Transaktion ausgeglichen wurde. Bei diesen Kanten handelt es sich um Testkanten. Bei einer normalen Kante von einer Transition zu einer Stelle wird beim Schalten der Transition eine Markierung auf der verbundenen Stelle erzeugt. Die Beschriftung der jeweiligen Kante legt dabei fest, welchen Wert (bzw. Farbe) diese Markierung erhält. Dabei muss die Farbe der Markierung mit der Farbe der Transition Übereinstimmen, damit die Transition schalten kann. Bei einer Testkante wird hingegen überprüft, ob eine Markierung mit einer der Beschriftung der Testkante entsprechendem Wert vorliegt. Diese wird beim Schalten der zugehörigen Transition allerdings nicht konsumiert. Dieses Verhalten spiegelt den buchhalterischen Vorgang der Auszifferung von offene-Posten-geführten Buchungen wider. Hierbei wird der Posten zwar ausgeglichen, die ursprüngliche Buchung bleibt aber erhalten und wird nicht etwa wieder gelöscht. Des Weiteren sind in dem Modell zusätzliche Plätze enthalten. Diese sind mit Startmarkierungen belegt, die der Buchungsnummer der zugehörigen Transaktion entsprechen, und sorgen dafür, dass das rekonstruierte Modell ein schaltfähiges Petri-Netz darstellt.

Das hier illustrierte Beispiel stellt somit das Modell einer Prozessinstanz in Form eines schaltfähigen Netzes dar, das die Durchführung der tatsächlich stattgefundenen Transaktionen und Buchungen imitiert. MB01 ist eine SAP Transaktion, mit der Wareneingänge erfasst werden. Mit MR1M werden Eingangsrechnungen gebucht. Über F110 wird der Zahllauf durchgeführt und mit FB1S werden offene Posten ausgeglichen. Mit der Transaktion FB01 können allgemeine Buchungen vorgenommen werden. Das Modell zeigt somit eine Instanz eines Einkaufsprozesses in dessen Zuge zuerst eine Ware eingegangen ist, dann die Eingangsrechnung verarbeitet und hiermit der offene Posten durch den Wareneingang ausgeglichen wurde. Nach Eingang der Rechnung wurde diese über einen Zahllauf beglichen. Des Weiteren wurde mit dem Zahllauf eine weitere Verbindlichkeit ausgeglichen, deren Gegenbuchung auf ein Aufwandskonto für EDV-Material gebucht wurde. Die Bedeutung der einzelnen Transaktionscodes und die Kontenbezeichnungen sind in Tabelle 1 und 2 aufgeführt.

Tabelle 1 Transaktionscodes

Transaktion	Bedeutung
MB01	Wareneingang zur Bestellung buchen
FB1S	Ausgleichen Sachkonto
MR1M	Eingangsrechnung erfassen
F110	Parameter für maschinelle Zahlung
FB01	Beleg buchen

Tabelle 2 Kontenbezeichnungen

Kontonummer	Kontobezeichnung
790000	Unfertige Erzeugnisse
191100	WE/RE-Verrechnung -Eigenfertigung-
230051	Erfolg Euro Umstellung / Beleg Differenz
154000	Eingangssteuer
160000	Kreditoren-Verbindlichkeiten Inland
113101	Deutsche Bank (Ausgangs-Schecks)
276000	Skonto-Ertrag
476100	EDV-Material

Zu Evaluationszwecken wurden rekonstruierte Modelle mit Hilfe der Software Renew [34] untersucht. Mit dieser Software ist es möglich, die Ausführung von Petri-Netzen zu simulieren. Die Simulation hat gezeigt, dass durch den verwendeten Algorithmus vollständig erreichbare und damit korrekt modellierte Netze erzeugt werden.

6. Bisherige Auswertungsergebnisse

Im vorausgegangenen Abschnitt wurden das Vorgehen zur Erzeugung und die verwendete Repräsentationsform anhand eines einfachen Beispiels für eine Prozessinstanz eines Einkaufsprozesses erläutert. Mittels der vorgestellten Methodik ist es möglich, Prozessinstanzen zu rekonstruieren. Für die Aufbauprüfung im Zuge der Jahresabschlussprüfung können diese Verfahren eingesetzt werden, um den tatsächlichen Ablauf von Geschäftsvorfällen zu visualisieren. Der Prüfer wird damit in die Lage versetzt, auf Basis der in den zugrundeliegenden Systemen gespeicherten Daten Informationen über den Zusammenhang zwischen der Ausführung von Prozessen und den Buchhaltungskonten zu erhalten, ohne auf manuelle Informationsgewinnung zurückgreifen zu müssen. Dies ist ohne den Einsatz der vorgestellten Process Mining Methoden nicht möglich.

Abbildung 5 zeigt ebenfalls eine Prozessinstanz für einen Einkaufsprozess. Hierbei ist zu erkennen, dass das dargestellte Modell bereits erheblich komplexer ist. Im vorherigen Beispiel beinhaltete die Instanz fünf Transaktionen. Die Prozessinstanz in Abbildung 5 hingegen enthält 46 Transaktionen. Dabei ist zu erkennen, dass die Teilprozesse für Wareneingang, Rechnungseingang und Ausgleich der offenen Posten (MB01-FB1S-MR1M) mit den Teilprozessen in Abbildung 4 vergleichbar sind. Beim ersten Beispiel wird durch den Zahllauf ein Einkaufsteilprozess beendet, im zweiten Beispiel acht.

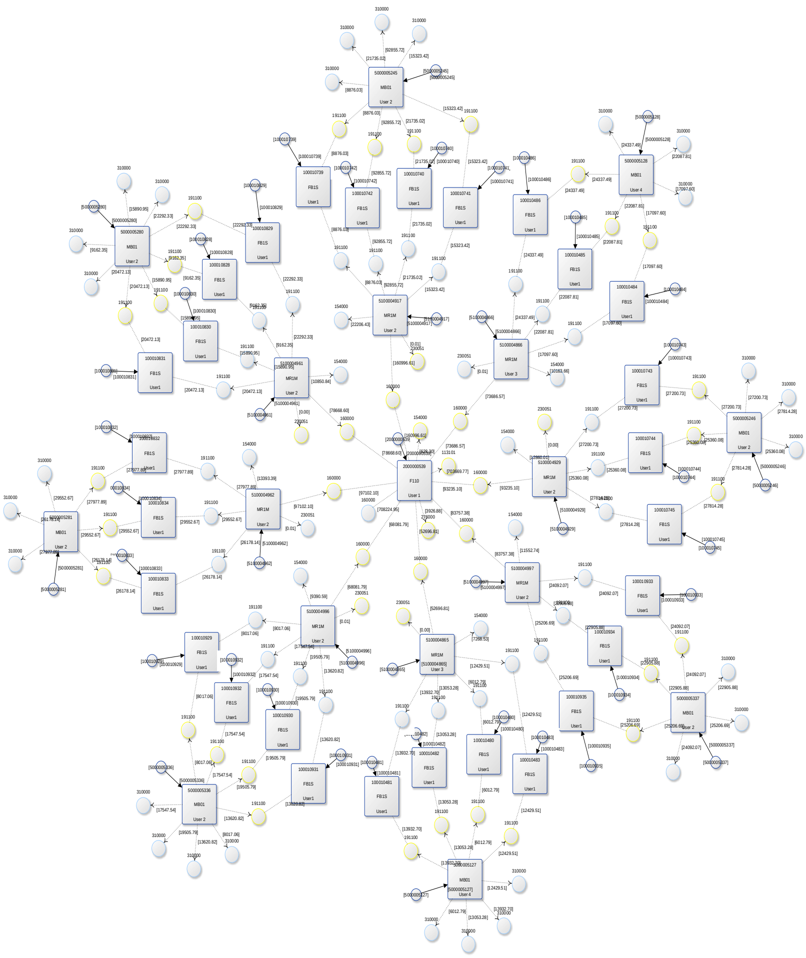


Abbildung 5 Umfangreiche Prozessinstanz eines Einkaufsprozesses

Für die Bewertung der Anwendbarkeit der dargestellten Process Mining Methode wurden neben den dargestellten Prozessinstanzen umfangreiche Event Logs ausgewertet. Hierbei handelte es sich um Testdaten aus dem SAP IDES System [35] sowie um Echtdaten aus produktiven Systemen von Unternehmen. Hiervon ist eins in der Einzelhandelsbranche tätig sowie das zweite in der verarbeitenden Industrie. Tabelle 3 gibt einen Überblick über den Umfang der verwendeten Daten.

Tabelle 3 Übersicht Event Logs

Event Logs	#1 SAP IDES	#2 Einzelhandel	#3 Verarbeitende Industrie
Buchhaltungsbelege	115,060	92,487	1,764,773
Belegsegmente	419,106	222,901	7,395,434
Anzahl rekonstruierter Prozessinstanzen	81,171	40,130	1,035,805

Die rekonstruierte Instanz mit der höchsten Komplexität umfasste ca. 470.000 Transaktionen. Die in Abbildung 6 dargestellte Instanz beinhaltet 2.966 Transaktionen. Obwohl die bisher entwickelten Methoden sicherlich nutzbringend für die Aufbauprüfung von Prozessen eingesetzt werden können, zeigt dieses Beispiel deutlich deren Grenzen. Denn anhand der bloßen Visualisierung einer solch komplexen Instanz können keine für die Prüfung verwertbaren Informationen gewonnen werden. Hierfür bedarf es weiterer Betrachtungen.

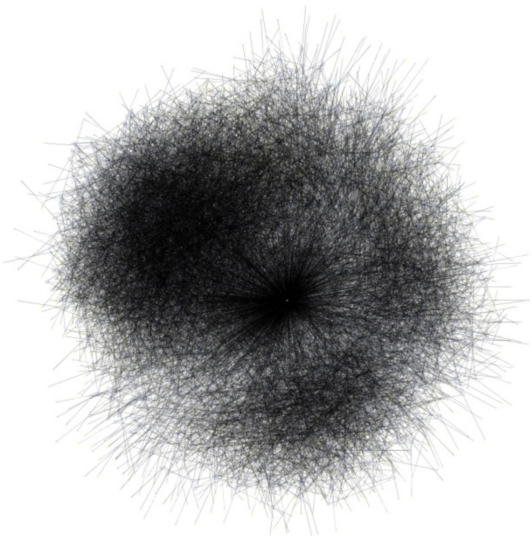


Abbildung 6 Komplexe Prozessinstanz

Erste Ansatzmöglichkeiten ergeben sich aus der Analyse der Verteilungsfunktionen der Anzahl der Instanzen über deren Komplexität[43]. Diese zeigen, dass es nur extrem wenige Instanzen mit einer extrem hohen Anzahl von Transaktionen gibt, wohingegen die über-wiegende Mehrzahl der Instanzen aus sehr wenigen Transaktionen bestehen. Ebenso ist zu beobachten, dass auch in sehr großen Prozessinstanzen nur relative wenige Konten verwendet werden.

Diese Beobachtungen bieten Ansatzpunkte für mögliche Verfahren zur Komplexitätsreduzierung der rekonstruierten Modelle. Dabei ist zu berücksichtigen, dass die Vereinfachung von rekonstruierten Modellen weder im Forschungsfeld der Graphentheorie noch im Gebiet der Forschung zur Modellierung von Geschäftsprozessen ein neuartiges Problem dar-

[43] Gemessen als Anzahl der Knoten und Kanten der jeweiligen Graphen.

stellt. Algorithmen aus der Graphentheorie befassen sich mit der Identifizierung isomorpher Teilgraphen [36], [37]. Im Forschungsbereich zum Thema Process Mining wird für diese Problematik der anschauliche Begriff „Spaghetti-Prozesse“ verwendet [38] und mittels Abstraktionsmethoden adressiert [39]. Wie diese Methoden für Process Mining eingesetzt werden können, bleibt zukünftiger Forschung überlassen.

7. Zusammenfassung und Diskussion

Durch die Automatisierung der Verarbeitung von Geschäftsvorfällen in Unternehmen werden Daten in den beteiligten Informationssystemen gespeichert. Der Einsatz von Business Intelligence Software ermöglicht die Auswertung dieser Daten. Wirtschaftsprüfer stehen im Zuge der Jahresabschlussprüfung vor der Herausforderung, komplexe und zunehmend in Informationssysteme integrierte Geschäftsprozesse prüfen zu müssen. Traditionelle manuelle Prüfungsmethoden stoßen hierbei an ihre Grenzen. Dies führt zu einem Ungleichgewicht zwischen automatisierter Transaktionsverarbeitung auf der Seite der zu prüfenden Unternehmen und manuellen Prüfungshandlungen auf der Seite der Jahresabschlussprüfer.

Zielgerichtete Prozess Mining Methoden bieten eine vielversprechende Möglichkeit, dieses Ungleichgewicht zu reduzieren. In diesem Beitrag wurde gezeigt, wie Process Mining Methoden für die Aufbauprüfung von Prozessen eingesetzt werden können, um anhand rekonstruierter Modelle von Prozessinstanzen Informationen über den Zusammenhang der untersuchten Geschäftsprozesse und der Buchhaltungskonten zu erlangen.

Neben den erzielten Ergebnissen zeigen die Auswertungen von Prozessmodellen bestehende Grenzen zu deren praktischen Einsatzmöglichkeiten. Sobald es sich um komplexe Prozessinstanzen handelt, bietet deren Modellierung und Visualisierung mit den vorgestellten Mitteln wenig Mehrwert. Hier gilt es Methoden zu finden, welche für die Komplexitätsreduktion verwendet werden können. Ein weiterer Aspekt betrifft die Zusammenfassung gleichartiger Prozessinstanzen zu Prozessmodellen. Obwohl bestehende Algorithmen die Erstellung von Prozessmodellen erlauben, ist zu prüfen, inwiefern diese den Anforderungen der Anwendungsdomäne genügen und ob diese im Sinne eines Method Engineering zweckmäßig verwendet werden können. [44]

Literaturverzeichnis

[1] K. C. Laudon, J. P. Laudon, und D. Schoder, Wirtschaftsinformatik : eine Einführung. München [u.a.]: Pearson Studium, 2006.

[2] IBM, „IBM Cognos Business Analytics und Performance Management Software - Deutschland“. [Online]. Available: http://www-01.ibm.com/software/de/data/cognos/. [Accessed: 17-Sep-2012].

[44] Die in dieser Veröfentlichung präsentierten Ergebnisse wurden im Forschungsprojekt Virtual Accounting Worlds erarbeitet. Das Projekt wird vom Bundesministerium für Bildung und Forschung gefördert (Fördernummer 01IS10041). Die Autoren sind verantwortlich für den Inhalt der Veröffentlichung.

[3] Oracle, „Oracle Business Intelligence Enterprise Edition". [Online]. Available: http://www.oracle.com/technetwork/middleware/bi-enterprise-edition/overview/index.html. [Accessed: 17-Sep-2012].

[4] SAP, „SAP Deutschland - Komponenten & Werkzeuge von SAP NetWeaver: SAP NetWeaver Business Intelligence". [Online]. Available: http://www.sap.com/germany/plattform/netweaver/components/businessintelligence/index.epx. [Accessed: 17-Sep-2012].

[5] E. Turban, J. E. Aronson, T.-P. Liang, und R. Sharda, Decision support and business intelligence systems. Upper Saddle River, N.J.; London: Pearson Education International, 2007.

[6] C. Vercellis, Business intelligence. Chichester: Wiley, 2009.

[7] IDW, „IDW Aktuell - Willkommen im Institut der Wirtschaftsprüfer". [Online]. Available: http://www.idw.de/. [Accessed: 17-Sep-2012].

[8] IAASB, „IAASB | International Accounting | Auditing Standards | Quality Assurance - IFAC". [Online]. Available: http://ifac.org/auditing-assurance. [Accessed: 17-Sep-2012].

[9] International Federation of Accountants, „ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment". 2012.

[10] International Federation of Accountants, „ISA 330 The Auditor's Responses to Assessed Risks", in Handbook of international quality control, auditing, review, other assurance, and related services pronouncements., Bd. 1, New York, NY: International Federation of Accountants, 2010.

[11] IDW, IDW PS 261 Feststellung und Beurteilung von Fehlerrisiken und Reaktionen des Abschlussprüfers auf die beurteilten Fehlerrisiken. 2009.

[12] M. Werner und N. Gehrke, „Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen", in Forschung für die Wirtschaft, Shaker Verlag, 2011.

[13] M. Werner, N. Gehrke, und M. Nüttgens, „Business Process Mining and Reconstruction for Financial Audits", in Hawaii International Conference on System Sciences, 2012, S. 5350–5359.

[14] W. van der Aalst, K. van Hee, J. M. van der Werf, A. Kumar, und M. Verdonk, „Conceptual model for online auditing", Decision Support Systems, Bd. 50, Nr. 3, S. 636–647, Feb. 2011.

[15] J. E. Cook und A. L. Wolf, „Software process validation: quantitatively measuring the correspondence of a process to a model", ACM Trans. Softw. Eng. Methodol., Bd. 8, Nr. 2, S. 147–176, 1999.

[16] J. E. Cook und A. L. Wolf, „Discovering models of software processes from event-based data", ACM Transactions on Software Engineering and Methodology, Bd. 7, Nr. 3, S. 215–249, Juli 1998.

[17] J. E. Cook und A. L. Wolf, „Event-based detection of concurrency“, SIGSOFT Softw. Eng. Notes, Bd. 23, Nr. 6, S. 35–45, 1998.

[18] A. Tiwari, C. Turner, und B. Majeed, „A review of business process mining: state-of-the-art and future trends“, Business Process Management Journal, Bd. 14, Nr. 1, S. 5–22, 2008.

[19] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st Edition. Springer Berlin Heidelberg, 2011.

[20] M. de Leoni, F. M. Maggi, und W. M. P. van der Aalst, „Aligning Event Logs and Declarative Process Models for Conformance Checking“, in Business Process Management, 2012, Bd. 7481, S. 82–97.

[21] J. M. E. M. van der Werf, H. M. W. Verbeek, und W. M. P. van der Aalst, „Context-Aware Compliance Checking“, in Business Process Management, 2012, Bd. 7481, S. 98–113.

[22] E. Ramezani, D. Fahland, und W. M. P. van der Aalst, „Where Did I Misbehave? Diagnostic Information in Compliance Checking“, in Business Process Management, 2012, Bd. 7481, S. 262–278.

[23] S. Banescu, M. Petković, und N. Zannone, „Measuring Privacy Compliance Using Fitness Metrics“, in Business Process Management, 2012, Bd. 7481, S. 114–119.

[24] R. Accorsi und A. Lehmann, „Automatic Information Flow Analysis of Business Process Models“, in Business Process Management, 2012, Bd. 7481, S. 172–187.

[25] N. Gehrke und N. Müller-Wickop, „Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems“, in Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010.

[26] A. R. Hevner, S. T. March, J. Park, und S. Ram, „Design science in information systems research“, Mis Quarterly, S. 75–105, 2004.

[27] S. T. March und G. F. Smith, „Design and natural science research on information technology“, Decision support systems, Bd. 15, Nr. 4, S. 251–266, 1995.

[28] H. Österle, J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, und E. J. Sinz, „Memorandum zur gestaltungsorientierten Wirtschaftsinformatik“, Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, Bd. 62, Nr. 9, S. 662–672, 2010.

[29] C. Riege, J. Saat, und T. Bucher, „Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik“, Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik, S. 69–86, 2009.

[30] S. Brinkkemper, „Method engineering: engineering of information systems development methods and tools“, Information and Software Technology, Bd. 38, Nr. 4, S. 275–280, 1996.

[31] W. M. P. van der Aalst und C. Stahl, Modeling business processes : a petri net-oriented approach. Cambridge, Mass.: MIT Press, 2011.

[32] R. Valk, „Lecture Notes: Formale Grundlagen der Informatik II (FGI 2) Modellierung & Analyse paralleler und verteilter Systeme“. Universität Hamburg, 2008.

[33] N. Gehrke und N. Müller-Wickop, „Rekonstruktion von Geschäftsprozessen im Finanzwesen mit Financial Process Mining“, in Lecture Notes in Informatics, Proceedings der Jahrestagung Informatik, Leipzig, 2010.

[34] University of Hamburg, „Renew - The Reference Net Workshop“, 2012. [Online]. Available: http://www.renew.de/. [Accessed: 03-Mai-2012].

[35] SAP, „SAP-UCC“, 2012. [Online]. Available: http://www.sap-ucc.com/. [Accessed: 03-Mai-2012].

[36] J. Huan, W. Wang, und J. Prins, „Efficient mining of frequent subgraphs in the presence of isomorphism“, in Data Mining, 2003. ICDM 2003. Third IEEE International Conference on, 2003, S. 549–552.

[37] X. Yan und J. Han, „gSpan: graph-based substructure pattern mining“, 2002, S. 721–724.

[38] C. Günther und W. van der Aalst, „Fuzzy mining–adaptive process simplification based on multi-perspective metrics“, Business Process Management, S. 328–343, 2007.

[39] M. Reichert, „Visualizing Large Business Process Models: Challenges, Techniques, Applications“, in 1st Int’l Workshop on Theory and Applications of Process Visualization, Tallin, 2012.

10.5 Business Process Mining and Reconstruction for Financial Audits

Number	**5**
Title	**Business Process Mining and Reconstruction for Financial Audits**
Appendix	10.5
Primary Related Chapters	5.1
Type	Conference Paper
Conference	45th Hawaii International Conference on System Sciences (HICSS 2012)
Reference	(Werner et al., 2012a)
Acceptance Rate	54 %
VHB JQ 2.1 Ranking	C (6.44)
WKWI Ranking	B
ERA 2010	A
CORE 2013	A
Review Procedure	Double blinded
Number of Reviews	6
Authors	1. Michael Werner 2. Nick Gehrke 3. Markus Nüttgens
Dissertation Points	0.50
Authorship	
Overall	90%
Design	90%
Realization	90%
Writing	90%
Status	Published
Part of other Dissertations	No
Link	http://ieeex-plore.ieee.org/stamp/stamp.jsp?tp=&ar-number=6149542&tag=1

Business Process Mining and Reconstruction for Financial Audits

MICHAEL WERNER
Nordakademie, University of Applied Sciences, Germany
michael.werner@nordakademie.de

NICK GEHRKE
Nordakademie, University of Applied Sciences, Germany
nick.gehrke@nordakademie.de

MARKUS NÜTTGENS
University of Hamburg, Germany
markus.nuettgens@wiso.uni-hamburg.de

Abstract: In modern companies business processes and information systems are highly integrated and transactions are executed system based and automated. The data generated in the course of processing transactions commonly provides the basis for internal and external financial reporting. The financial statements are subject to audits due to regulatory requirements. Contemporary audit approaches take into account internal control frameworks over relevant business processes and underlying information systems, but they lack adequate audit procedures needed to handle voluminous data flows when business processes are highly integrated and automated. We face a discrepancy between an integrated and automated transaction processing on the one side and manual audit procedures on the other. Financial audits would be more effective and efficient if an audit approach with system based and automated procedures would be applied. This article describes how business process mining and reconstruction of mined processes can be used to overcome this discrepancy.

1. Introduction

The execution of business processes in companies is regularly based on information systems. The integration between business process and information systems ranges from support for manual executions to completely automated processing. Enterprise Resource Planning (ERP) systems represent the dominant type of information systems that are implemented to support and automate transaction processing. Depending on the industry and types of business processes that are integrated into the ERP systems millions or even billions of transactions may be processed within a financial period.

ERP systems do not only support or automate the execution of transactions but they also commonly provide the data basis for the internal and external financial reporting as well as integrated functionality for preparing the financial statements including the balance sheet and profit and loss statements. This means that the financial statements represent an aggregation of the information stored in the ERP system that is made up by the processing of myriad numbers of transactions.

Companies are required to prepare financial statements in order to inform addressees primarily about the financial situation of the company. To protect addressees from misinformation the financial statements are subject to independent audits by financial auditors. The requirements for the audit are specified in local or international laws, regulations and standards.

International Standards on Auditing require the application of a risk based audit approach that takes into account the internal control framework over relevant business processes and underlying information systems (ISA 315). A risk based audit approach requires the identification of relevant risks for material misstatements and the evaluation how internal controls are able to mitigate existing risks. When business processes are integrated with ERP systems application controls represent a significant type of internal controls that have to be considered in the audit. Contemporary audit approaches consider internal controls and application controls as a special type of internal controls that are embedded in the integrated system. Although relevant business processes and internal controls are taken into account contemporary audit procedures are generally not system based. The selection and test of controls is done manually.

This situation leads to a discrepancy. Business processes are highly integrated with ERP systems and transactions are processed automatedly. Auditors identify significant risks and manually evaluate relevant internal controls that are in place to mitigate these risks. On the company side we observe system based and automated processing and on the auditor side a risk based approach with manual evaluation procedures. An approach with automated and system based audit procedures would lead to more efficient and effective audits.

The need for automated audit procedures has been pointed out by major market participants [17], but an approach that includes system based and automated audit procedures has not been developed yet due to the fact that adequate methods and software artifacts have not been available.

Recent research by GEHRKE et al. [14], [15] has revealed how financially relevant information can be extracted via financial business process mining from information systems. The mined information can be used to reconstruct processes based on that information. This paper presents how business process mining and reconstruction can be used to apply a risk based audit approach with system based and automated audit procedures.

The article starts with an overview of related theoretical work in section 2. Section 3 provides a brief summary of contemporary audit approaches and their limitations in system based and highly integrated environments. Section 4 presents the concepts of business process mining and reconstruction. Section 5 discusses the relevance of application controls for highly integrated and automated business processes. In section 6 we discuss how the concepts of business process mining and reconstruction can be combined with automated application control testing for developing system based and automated audit procedures. Section 7 closes with a discussion how stakeholders benefit, which limitations exist and what further developments are needed.

2. Related Work

The idea of process mining evolved in the 1990s. COOK and WOLF [9], [10], [11] investigated process mining in the context of software engineering. They describe different methods for process discovery. However, they do not provide an approach to generate explicit process models. The idea to apply process mining in the context of workflow management was first introduced by AGRAVAL et al. [8]. Further research was undertaken by MAXEINER et al. [28] and SCHIMM [33], [34], [35], [36] who developed mining tools. HERBST and

KARAGIANNIS also address process mining in the context of workflow management using an inductive approach [18], [19], [20], [21], [22], [23]. Substantial research has been published by VAN DER AALST et al. [1], [2], [3], [4], [5], [6], [7], [26], [27], [38], [39], [40] that deals with the mining and rediscovery of process models from event logs. This research is especially relevant because the provided methods and algorithms allow the reconstruction of petri nets which represent the process models mined from event logs. They also cover considerations of workflow performance, concurrency, noise and conformance checking.

Although this research is valuable for the research subject of this article several limitations have to be considered. The process mining research by VAN DER AALST et al. focuses on event based logs and intends to reconstruct graphs that completely represent the processes that produce these event logs. Mining of business processes in ERP systems for the financial audit entails different environmental settings and intends to achieve different aims. First, the stored data in ERP systems includes much more detailed information than common event logs from workflow systems. They store accounting information about journal entries that provide more specific data usable for process mining. Second, for the proposed mining we intend to rediscover single representative process instances that are aggregated into process models, we do not intend to rediscover complete representations of the mined data.

The discipline of process mining is characterized by technical research approaches. The connection between process flows, process mining, process reconstruction and accounting has not been extensively covered in scientific work so far. We assume that the connection between informatics oriented process mining and the business management topics accounting and compliance has not been the focus of interest so far due to the thematically distance of the two disciplines.

An exception is the research work by GEHRKE et al. [13], [14], [15], [16] that has been derived from the research project Virtual Accounting Worlds [37]. The developed methods and concepts bridge the gap between accounting, compliance and process mining. They represent an application of fundamental concepts from VAN DER AALST et al. for mining of business processes that are relevant for financial accounting. Within this paper we relate to these methods and concepts and include them in a wider consideration in order to demonstrate how they can be applied for developing an audit approach that includes system based and automated audit procedures.

Further relevant research covers topics like process data warehousing by EDER et al. [12] and ZUR MÜHLEN et al. [29], [30], [31] which can be used for developing software artifacts needed to apply the discussed approach in practice.

3. Risk based audit approach in the context of integrated and automated business processes

Before we can understand the shortcomings of the application of contemporary audit approaches in business environments with integrated and automated business processes we need to clarify their relevant characteristics.

Companies are required to apply generally accepted accounting principles (GAAP) when preparing their financial statements. This ensures that the published statements display correct and comparable information to the addressees. External addressees are shareholders, creditors, tax and regulatory authorities, employees, clients, financial analysts, competitors and the general public [24].

In order to protect addressees from misinformation the financial statements are audited by financial auditors. The obligation to engage auditors for auditing the financial statements is generally mandated by law. Auditors follow standards on auditing to ensure that adequate audit procedures are applied. They assess if the audited statements give a fair and true view of the financial situation of a company and if the statements are free of material misstatements.

Standards on accounting and standards on auditing are issued by regulatory bodies such as the International Accounting Standards Board (IASB) for the International Financial Reporting Standards (IFRS) or International Auditing and Assurance Standards Board (IAASB) for the International Standards on Auditing (ISA). Laws, regulations and standards differ between countries. But in recent years we observe a convergence between internationally significant accounting frameworks especially between the IFRS and US GAAP [25]. We do not intend to focus on differences of the accounting and audit frameworks in this article. For our purpose it is sufficient to point out that a risk based audit approach is mandated by ISA (e.g. ISA 315) as well as local regulations or standards such as the Sarbanes-Oxley-Act in the USA. For the remainder of this article we primarily refer to IFRS and ISA while pointing out that the same considerations and conclusions provided in this article are applicable to other accounting and audit frameworks.

ISA 315 requires the application of a risk based audit approach: "The objective of the auditor is to identify and assess the risks of material misstatement, whether due to fraud or error (…) through understanding the entity and its environment, including the entity's internal control (…)" (ISA 315.3). The auditor has to identify and to evaluate the risks that might lead to material misstatements. The auditor further needs to identify if internal controls do exist that mitigate existing risks: "The auditor shall obtain an understanding of internal control relevant to the audit (…)" (ISA 315.12). The underlying axiom of the approach is the assumption that well organized and controlled processes lead to correct financial reporting.

In practice the audit takes place by identifying risks that are significant to the audit. A general significant risk is that business transactions are not recorded completely or correctly. Following a risk based approach it is not necessary to consider all business processes within a company but only those where errors in the processing might lead to a material misstatement in the financial statements.

Typical business processes relevant for financial accounting are purchase, sales, payroll, production and logistics processes.

When the scope of the audit is determined and relevant processes identified the auditor has to gain an understanding of the processes and the internal control over these processes. The auditor has to evaluate if the controls are properly designed and operative to achieve the desired control objectives. The procedure to understand business processes, to evaluate and test internal controls is a manual and highly time-consuming activity. It generally includes interviews with knowledgeable contact persons and manual reviews of provided documentation.

We illustrate the procedure for the following example. For producing goods a company creates purchase requisitions and orders individually and with paper based forms. The orders have to be approved by signature by the purchase representative. The responsible warehouse worker checks if the amount and quantity of the received goods equal the amount and quantity of the purchase order when the goods are delivered. When the invoice for the delivered goods is received from the supplier a responsible person in the accounting department checks if the billed amount and quantity equal the amount and quantity of purchase order and the goods received.

For understanding the process and for evaluating the relevant internal controls an auditor first performs interviews with the persons involved in the process. He evaluates if the controls in place are adequate to control the process and to achieve the desired control objectives. Based on the understanding of the process and the controls he performs tests to evaluate if the controls are carried out continuously and effectively throughout the relevant reporting period. The testing of the operating effectiveness requires the review of relevant documents. In the mentioned example the auditor would draw a representative sample of purchase transactions and verify if check marks and signatures are available on the provided documentation.

The example illustrates that the audit procedures for auditing business processes and internal controls are highly manual and time-consuming in nature.

The described procedure is practical for manually executed business processes but it is insufficient when business processes are highly integrated with ERP systems and executed automated. Under such conditions contact persons from the relevant business functions generally lack sufficient knowledge about the integration and type of automation with the underlying systems. A common observation is that provided information does not correctly reflect the process implementation within the ERP systems. Second, with an increasing number of executed transactions manual review of available evidence becomes increasingly inefficient or even ineffective. A manual review of even hundreds of documents does not provide sufficient audit comfort when millions of such transactions are executed within the relevant period.

We illustrate the shortcomings of contemporary audit procedures in an integrated and automated business process environment with a second example.

A company has integrated its production and purchase processes in an ERP system. Production orders automatically initiate purchase requisitions and purchase orders based on item lists maintained in the system. The program routines initiate purchase orders only if required items are not available in the warehouse. Purchase requisitions and orders are approved automatically up to a certain amount. Only purchase orders exceeding that amount are subject to a system based approval by the purchase department. The system further blocks purchase orders randomly for manual but system based approval in order to prevent manipulation. The warehouse clerk can only accept received goods if the quantity and amount match the purchase order (two-way-match). Otherwise an exception handling sub-process is initiated. The accounting department can only process incoming invoices if the billed amount and quantity matches the amount and quantity of the purchase order and the goods received (three-way-match). Otherwise an exception handling sub-process is initiated.

The example demonstrates a highly integrated business process with automated executions and automated and systems based internal controls also referred to as application controls. In the described environment performing interviews with contact persons from the functional departments might not provide sufficient information because they may lack the information how transactions are processed automatically, when no human interaction occurs, and especially which application controls do exist. A common occurrence is that contact persons think application controls are in place and effective which in fact is not the case. A second dilemma becomes obvious when controls actually get tested. The automation of execution means that paper based evidence might not be available. In such a situation it is necessary to manually evaluate and to test relevant application controls. The evaluation and testing of application controls requires a specialized knowledge of the ERP system in use. Furthermore, the review of control settings, commonly based on the customizing settings, requires extensive access rights and is a manual time-consuming work. Third, even if these procedures are applied no information is available if the controls really cover complete transaction flows or if controls are bypassed by concurrent transaction flows differing from the general transaction flows, by manual journal entries or manipulation. Fourths, generally it is hard to test if the application controls were effective over the whole relevant period, for example if specific application controls were disabled for a specific timeframe.

Computer assisted audit techniques (CAAT) for supporting the testing of application controls and business processes integrated into ERP systems do exist [13]. But they only support the manual execution of tests or provide functionality to analyze mass data for journal entry testing. The described fundamental problems are not solved.

An audit approach is needed that counters the system based and automated processing by applying system based and automated audit procedures.

4. Business process mining and process reconstruction

Business process mining and process reconstruction provides the concepts and methods needed to implement system based and automated audit procedures.

VAN DER AALST et al. [1], [3], [4], [5], [6], [26], [27], [38], [39] focus on event logs in order to mine and reconstruct workflows. We can rely on these concepts to mine and reconstruct process models in ERP systems. In contrast to the systems and log files described by VAN DER AALST et al. ERP systems provide much more detailed information about the processed transactions. Every financially relevant business transaction executed in an ERP system is recorded as a journal entry posted to an account in a main or sub ledger. Basically, entries in the accounting of an ERP system are structured in a simple way [32]. Each entry consists of an accounting document and at least two items posted as credits and debits.

Technically documents and items are stored as entries in data tables in the underlying database of the ERP system. The stored data for each transaction contains information that allows identifying relationships between the transactions. The transactions can be traced back to the executed instance of a business process they belong to.

Mining of business processes that are relevant for financial accounting can be applied where open item accounting is enabled, which is the case for most relevant processes. If open item accounting is in use for a particular account, each item contains a flag that indicates if the item has already been cleared or not. If an item has been cleared, it also contains a reference to the entry / document which cleared the item.

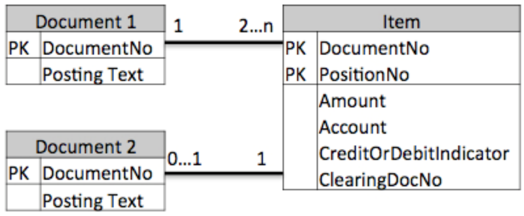


Figure 1 General data structure of an open item accounting entry

Figure 1 shows the general data structure of an accounting entry in a database. One document consists of two or more items posted on different accounts. Each item can be linked to one (other) document (=item cleared) or does not refer to (another) document (=item still open).

We illustrate the data structure with the example of an execution of a purchase process. The receipt of an ordered material (transaction 1) is recorded as a debit posting on raw materials and a credit posting on the goods received / invoices received account. Upon receipt of the incoming invoice (transaction 2) the posting on the goods received / invoices received account is cleared with a corresponding credit posting on the creditor account. The document number of transaction 2 becomes the clearing document number for the posting item from transaction 1. With the payment run (transaction 3) the creditor account is cleared. The document number of transaction 3 becomes the clearing document number for the posting item of transaction 2. The example illustrates how the execution of a business process instance is recorded in the system and how the processing produces a chain of journal entries within the system that is traceable.

If we abstract from this example we can conclude that transactions processed in an ERP system leave a digital trace within the system. GEHRKE AND MÜLLER-WICKOP [14], [15] present an algorithm that is able to mine these traces. The algorithm takes off with a start document and iteratively mines corresponding documents by identifying the relevant clearing document. If no further documents can be found the algorithm terminates. The mined information can be used for graphically reconstructing and representing the mined process instance.

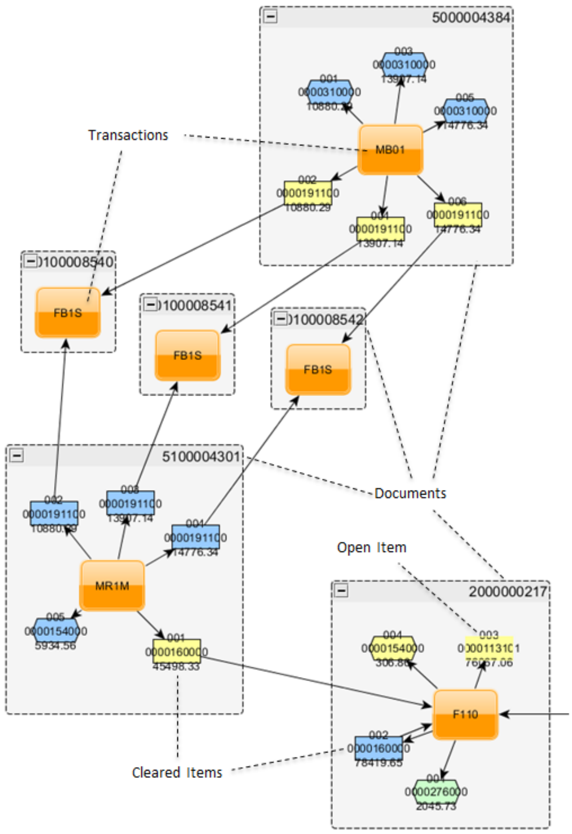


Figure 2 Reconstruction of a mined process instance

Figure 2 shows a section of the reconstruction for a mined purchase process instance from a SAP system.

Transactions are represented as rounded orange rectangles. Simple rectangles represent items of financial entries involved in open item accounting. The rectangles with black borders represent cleared items of financial entries. Items illustrated as rectangles without borders are not cleared and still open. Hexagons represent entries not involved in open item accounting. The color of the rectangles and hexagons indicates if the item is a debit or credit posting on a general ledger or a profit and loss account. An arrow from a business activity to an item means that the business activity has produced the item as a part of the complete entry. An arrow from an item to a business activity means that the item has been cleared by the corresponding accounting document. Detailed information such as item number, account and amount is displayed for each item.

The shown process instance starts with the transaction MB01 (post goods receipt for purchase order). The open items are cleared by the transaction MR1M (enter incoming invoice). When entering an invoice via MR1M it is possible to explicitly reference corresponding open items. In the mined instance MR1M was executed without such explicit references. In this case the items are cleared by the ERP system automatically via the execution of transaction FBS1 (clear G/L account). FBS1 does not represent a separate business transaction and therefore does not follow the structure displayed in Figure 1. The open item posted by MR1M is cleared by transaction F110 (payment run).

5. Application controls

ERP systems provide control mechanisms in order to govern and control the processing within the system. Control mechanisms that are inherently embedded in software are called application controls. Application controls represent a type of internal controls [13]. Examples are automatic reconciliation procedures, prevention of entering duplicate transactions, system forced approvals or system based two- and three-way-matches.

Application controls play a key role for auditing system based and automated processes. They provide a means for overcoming the problem that manual testing of business transactions becomes inefficient for integrated and automated processes. Instead of testing single business transactions it is possible to test the design and effectiveness of application controls that cover whole process flows independent from the number of business transactions that are processed.

By relying on application controls provided by the system for the purpose of the financial audit automated processing of transactions can be countered by automated control mechanisms.

Unfortunately the testing of application controls itself is a manual and time-consuming procedure. Application controls are generally configured and enabled during the implementation of the system by setting relevant customizing settings. These settings need to be evaluated. However, settings for application controls are stored within the ERP systems and methods and software artifacts exist that allow to extract relevant settings and to test them in an automated way [13].

6. System based and automated audit procedures

In the previous sections we illustrated how process instances can be automatically mined and reconstructed. We point out that business process mining and reconstruction does not merely represent another CAAT. Indeed it allows introducing a new audit procedure that is adequate for the audit of companies with highly integrated and automated business processes as shown by the following considerations.

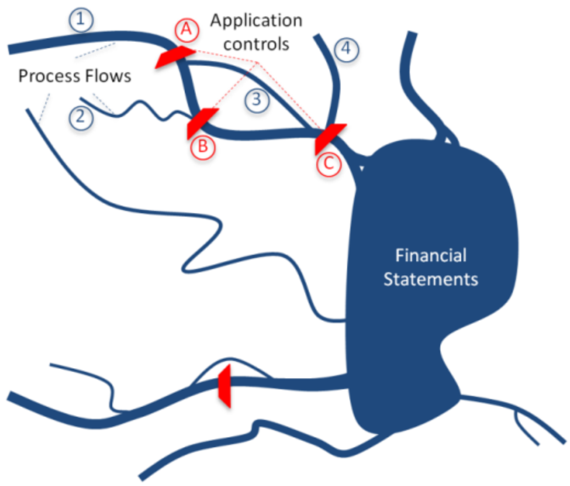
When applying a contemporary audit approach the auditor decides which financial accounts are in scope from a risk perspective and evaluates which business processes have to be considered. This decision is based on professional judgment derived primarily from experience. The scoping may be appropriate or not. It is a manual procedure and highly dependent on the knowledge and experience of the auditor. Information from the underlying information systems is not considered or only to a marginal extent although the information systems indeed provide all necessary information for a precise scoping.

Business process mining and reconstruction provides the possibility to analyze how the transactions flow throughout the system and which processes effect relevant accounts. The methods to reconstruct and visualize a single execution of a business process were presented in section 4. In order to implement automated and system based audit procedures it is necessary to aggregate mined process instances to process models that represent the process flows within the system. The aggregation of mined process instances is possible but adequate algorithms for automated aggregations are currently still under research.

We define a process flow as a collection of executed similar business activities. The process flows can be made explicit and the auditor can virtually see how they interact with the relevant accounts.

For illustrating the possibilities that business process mining offers, we use the following analogy. We compare the financial statements of a company to a lake of water. The auditor has to provide an opinion if the lake only contains water from specific sources with a defined quality. This requirement is the analogy to the real life requirement that financial statements present a fair and true view of the financial situation of the company. Rivers feed our imaginary lake. The rivers represent transaction flows and the water transaction data. Following a contemporary audit approach the auditor would manually take samples from different places in the lake to verify the water quality (=substantive testing). Based on experience and professional judgment he would also choose several rivers for inspection and verify manually if control mechanisms are in place (=manual controls testing) that regulate the flow and quality of the water, but without knowing which rivers and concurrent flows indeed exist and how much water they actually carry into the lake.

By applying business process mining and reconstruction the auditor first develops a map with all relevant rivers that flow into the lake with information which control mechanisms control the flow and quality of the water. The auditor gathers information about how much water each river carries and which rivers or concurrent streams flow uncontrolled (=business process mining and reconstruction). Based on this understanding the auditor can decide precisely which rivers are significant and instead of taking random samples the auditor can decide specifically which control mechanisms should be tested to cover relevant process flows (=automated controls testing). Process flows with no application controls in place can be identified for targeted samples.



Process flows	
#	Description
1	purchase order
2	goods receipt for purchase order
3	services receipt
4	incoming invoice

Application controls	
#	Description
A	system based approval for purchase order
B	Two-way-match
C	Three-way-match

Figure 3 Process flow map

Figure 3 illustrates on an aggregated level a map of process flows. It illustrates how different process flows feed the financial statements. The diagram further shows how application controls interact with the process flows and how they control them.

In the upper left corner of the diagram we illustrate how a typical automated purchase process as already described in section 3 would be represented.

The process flow of purchase orders (process flow 1) is controlled by a system based approval (application control A), the combined process flow of purchase orders and goods receipt (process flow 2) is controlled by an automated two-way-match (application control B). When transaction flows 1 and 2 combine with the process flow of incoming invoices (process flow 4) the combined flow is controlled by an automated three-way-match (application control C). The diagram also shows that the concurrent process flow of receipt services (process flow 3) is not controlled by the two-way-match (application control B). The reason is that commonly no receipt data for the delivery of services is available that could be subject of an automated control activity. A matching for delivered services commonly takes place between the purchase order and the invoice via application control C.

7. Conclusion

In modern companies business processes and information systems are highly integrated and transactions are executed system based and automatedly. In section 3 we have shown that contemporary audit procedures for financial audits are not adequate in environments where business processes are highly integrated and automated. It is ineffective and inefficient to audit automated business processes and internal controls with manual audit procedures.

Business process mining and reconstruction provides methods and procedures that base directly on the information stored in the underlying ERP systems. These methods and procedures combined with methods and procedures usable for automated application control testing can be applied to implement system based and automated audit procedures for financial audits. Business process mining and reconstruction allows visualizing process flows within an ERP system and how these process flows interact with application controls embedded in the system. The automated and system based analysis and its graphical representation enables auditors to handle the complexity of integrated and system based business processes and internal controls.

Auditing firms have recognized the need to introduce automated audit procedures in order to keep up with technological progress [17]. Via the application of system based and automated audit procedures as introduced in this article it is possible to meet this requirement. It is expected that the introduction of system based and automated audit procedures will lead to significant gains in effectiveness and efficiency of financial audits.

The conclusions presented in this paper base on the research work derived from the research project Virtual Accounting Worlds (VAW) [37]. A major market participant of the auditing industry participates in the project as an associated project partner. The prototype for financial business process mining and reconstruction as well as the prototype for automated application control testing has been developed within the VAW project. It is intended to develop a software artifact providing the functionality for system based and automated audit procedures as described in this article in further research.

Internal and external auditors are not the only stakeholder that would benefit from the availability of methods and artifacts for automated auditing. The process mining, reconstruction and visualization provides the basis for analyses, performance and optimization consideration that are of interest for process owners and managers, risk management and business management in general.

We have to point out that in order to implement system based and automated audit procedures further issues need to be researched. The prototypes referred to in this article [13], [14] provide software artifacts that proof the correctness and applicability of the underlying concepts and methods. Nevertheless no information is available how the artifacts will behave and perform in real live environments. The described methods for process mining allow the mining of single process instances. In order to analyze and to visualize reconstructed process flows it is necessary to aggregate mined process instances to process models. Adequate algorithms for the automated aggregation of mined process instances are still under research as well as methods for an automated visualization. Further attention has to be paid to the selection of representative process instances if the complete mining of all instances is not a viable option due to the amount of processed instances which will be a common occurrence in real life settings.

The aspects discussed in this paper focus on methods for automating business process and internal controls testing. The overall audit of financial statements is a complex and difficult task carried out by qualified experts. Process and controls testing is only a part of a financial audit. The possibilities for automating audit procedures are limited to the extent to how the underlying transaction processing is automated. Due to the fact that companies act as market participants in changing and volatile environments there will always be unique business transactions such as mergers or acquisitions that need to be evaluated by manual and substantive audit procedures. The aim of introducing system based and automated audit procedures is to counter automated processing with adequate audit procedures and to set free resources for more sophisticated audits of unique, uncontrolled or exceptional transactions that generally comprise higher risks than standard transactions.

Although further research is needed for developing mature software artifacts that allow the implementation of system based and automated audit procedures for financial audits presented in this article we conclude that the basic methods have already been developed and proofed valid.

Further research has to focus on issues such as process instance selection and automated aggregation, visualization, complexity and viability of developed algorithms. These aspects will be the focus of further research within the VAW project that has already been initiated. Specific information concerning these topics and the developed software artifact will be provided in subsequent publications.

8. References

[1] W.M.P. van der Aalst, “Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing”, Requirements Engineering Journal, volume 10, issue 3, pp. 198-211, 2005.

[2] W.M.P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st Edition, Springer Berlin Heidelberg, 2011.

[3] W.M.P. van der Aalst, “The Application of Petri Nets to Workflow Management,” The J. Circuits, Systems and Computers, vol. 8, no. I, pp. 21-66, 1998.

[4] W.M.P. van der Aalst, “Verification of Workflow Nets”, Application and Theory of Petri Nets, P. Azema and G. Balbo, eds., pp. 407-426, Berlin: Springer-Verlag, 1997.

[5] W.M.P. van der Aalst, and B.F. van Dongen, “Discovering Workflow Performance Models from Timed Logs,” Proc. Int’l Conf. Eng. and Deployment of Cooperative Information Systems (EDCIS 2002), Y. Han, S. Tai, and D. Wikarski, eds., vol. 2480, pp. 45-63, 2002.

[6] W. M. P. van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A.J.M.M. Weijters, “Workflow Mining: a Survey of Issues and Approaches”, Beta: Research School for Operations Management and Logistics, Working Paper 74, 2002.

[7] W.M.P. van der Aalst, A.J.M.M. Weijters, and L. Maruster, “Workflow Mining: Which Processes can be Rediscovered?” BETA Working Paper Series, WP 74, Eindhoven Univ. of Technology, Eindhoven, 2002.

[8] R. Agrawal, D. Gunopulos, and F. Leymann, “Mining Process Models from Workflow Logs,” Proc. Sixth Int’l Conf. Extending Database Technology, pp. 469-483, 1998.

[9] J.E. Cook, and A.L. Wolf, “Discovering Models of Software Processes from Event-Based Data,” ACM Trans. Software Eng. and Methodology, vol. 7, no. 3, pp. 215-249, 1998.

[10] J.E. Cook and A.L. Wolf, “Event-Based Detection of Concurrency,” Proc. Sixth Int’l Symp. the Foundations of Software Eng. (FSE-6), pp. 35-45, 1998.

[11] J.E. Cook, and A.L. Wolf, “Software Process Validation: Quantitatively Measuring the Correspondence of a Process to a Model,” ACM Trans. Software Eng. and Methodology, vol. 8, no. 2, pp. 147-176, 1999.

[12] J. Eder, and G.E. Olivotto, and W. Gruber, “A Data Warehouse for Workflow Logs,” Proc. Int’l Conf. Eng. and Deployment of Cooperative Information Systems (EDCIS 2002), Y. Han, S. Tai, and D. Wikarski, eds., pp. 1-15, 2002.

[13] N. Gehrke, “The ERP AuditLab - A prototypical Framework for Evaluating Enterprise Resource Planning System Assurance”, in: Proceedings of the 43th Hawaii International Conference on System Sciences (HICSS-43), Hawaii (2010).

[14] N. Gehrke, and N. Müller-Wickop, “Basic Principles of Financial Process Mining”, Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010.

[15] N. Gehrke, and N. Müller-Wickop, “Rekonstruktion von Geschäftsprozessen im Finanzwesen mit Financial Process Mining“, in: Lecture Notes in Informatics, Proceedings der Jahrestagung Informatik 2010, IT-supported Service Innovation and Service Improvement, Leipzig, 2010.

[16] Gehrke, N., Wolf, P.: Towards Audit 2.0 - A Web 2.0 Community Platform for Auditors, in: Proceedings of the 43th Hawaii International Conference on System Sciences (HICSS-43), Hawaii (2010).

[17] K. Heese, H. Kreisel, “Püfung von Geschäftsprozessen“, Die Wirtschaftsprüfung, Vol. 18, pp. 907-919, 2010.

[18] J. Herbst, “A Machine Learning Approach to Workflow Management,” Proc. 11th European Conf. Machine Learning, pp. 183-194, 2000.

[19] J. Herbst, “Dealing with Concurrency in Workflow Induction,” Proc. European Concurrent Eng. Conf., U. Baake, R. Zobel, and M. Al- Akaidi, eds., 2000.

[20] J. Herbst, “Ein induktiver Ansatz zur Akquisition und Adaption von Workflow-Modellen,” PhD thesis, Universität Ulm, Nov. 2001.

[21] J. Herbst, and D. Karagiannis, “An Inductive Approach to the Acquisition and Adaptation of Workflow Models,” Proc. Workshop Intelligent Workflow and Process Management: The New Frontier for AI in Business, M. Ibrahim and B. Drabble, eds., pp. 52-57, Aug. 1999.

[22] J. Herbst, and D. Karagiannis, “Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models,” Proc. Ninth Int’l Workshop Database and Expert Systems Applications, pp. 745-752, 1998.

[23] J. Herbst, and D. Karagiannis, “Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models,” Int’l J. Intelligent Systems in Accounting, Finance, and Management, vol. 9, pp. 67-92, 2000.

[24] K. Küting, and M. Reuter, “Bilanzierung im Spannungsfeld unterschiedlicher Adressaten“, DSWR, Vol. 9, 2004.

[25] P. Leibfried, and P. Meixner, “Konvergenz der Rechnungslegung: Bestandsaufnahme und Versuch einer Prognose“, Der Schweizer Treuhänder, Vol. 4, 2006.

[26] L. Maruster, W.M.P. van der Aalst, A.J.M.M. Weijters, A. van den Bosch, and W. Daelemans, “Automated Discovery of Workflow Models from Hospital Data,” Proc. 13th Belgium-Netherlands Conf. Artificial Intelligence (BNAIC 2001), B. Kro¨se, M. de Rijke, G. Schreiber, and M. van Someren, eds., pp. 183-190, 2001.

[27] L. Maruster, A.J.M.M. Weijters, W.M.P. van der Aalst, and A. van den Bosch, “Process Mining: Discovering Direct Successors in Process Logs,” Proc. Fifth Int’l Conf. Discovery Science (Discovery Science 2002), pp. 364-373, 2002.

[28] M.K. Maxeiner, K. Küspert, and F. Leymann, “Data Mining von Workflow-Protokollen zur teilautomatisierten Konstruktion von Prozessmodellen” Proc. Datenbanksysteme in Büro, Technik und Wissenschaft, pp. 75-84, 2001.

[29] M. zur Mühlen, “Process-Driven Management Information Systems—Combining Data Warehouses and Workflow Technology,” Proc. Int’l Conf. Electronic Commerce Research (ICECR-4) B. Gavish, ed., pp. 550-566, 2001.

[30] M. zur Mühlen, “Workflow-Based Process Controlling-Or: What You Can Measure You Can Control,” Workflow Handbook 2001, Workflow Management Coalition, L. Fischer, ed., pp. 61-77, Lighthouse Point, Fla.: Future Strategies, 2001.

[31] M. zur Mühlen, and M. Rosemann, “Workflow-Based Process Monitoring and Controlling - Technical and Organizational Issues,” Proc. 33rd Hawaii Int’l Conf. System Science (HICSS-33), R. Sprague, ed., pp. 1-10, 2000.

[32] M. Romney, and M. Steinbart, Accounting Information Systems, International Version, 11th Edition, Prentice Hall, pp. 56-69, 2008.

[33] G. Schimm, “Generic Linear Business Process Modeling,” Proc. ER 2000 Workshop Conceptual Approaches for E-Business and The World Wide Web and Conceptual Modeling, S.W. Liddle, H.C. Mayr, and B. Thalheim, eds., pp. 31-39, 2000.

[34] G. Schimm, “Process Mining Elektronischer Geschäftsprozesse,” Proc. Elektronische Geschäftsprozesse, 2001.

[35] G. Schimm, “Process M6ning linearer Prozessmodelle - Ein Ansatz zur Automatisierten Akquisition von Prozesswissen,” Proc. 1. Konferenz Professionelles Wissensmanagement, 2001.

[36] G. Schimm, “Process Miner - A Tool for Mining Process Schemes from Event-Based Data,” Proc. Eighth European Conf. Artificial Intelligence (JELIA), S. Flesca and G. Ianni, eds., pp. 525-528, 2002.

[37] Virtual Accounting Worlds, www.virtualaccountingworlds.com, retrieved June 14, 2011

[38] A.J.M.M. Weijters, and W.M.P. van der Aalst, “Process Mining: Discovering Workflow Models from Event-Based Data,” Proc. 13th Belgium-Netherlands Conf. Artificial Intelligence (BNAIC 2001), B. Kro¨ se, M. de Rijke, G. Schreiber, and M. van Someren, eds., pp. 283-290, 2001.

[39] A.J.M.M. Weijters, and W.M.P. van der Aalst, “Rediscovering Workflow Models from Event-Based Data,” Proc. 11th Dutch-Belgian Conf. Machine Learning (Benelearn 2001), V. Hoste and G. de Pauw, eds., pp. 93-100, 2001.

[40] A.J.M.M. Weijters and W.M.P. van der Aalst, “Workflow Mining: Discovering Workflow Models from Event-Based Data,” Proc. ECAI Workshop Knowledge Discovery and Spatial Data, C. Dousson, F. Höppner, and R. Quiniou, eds., pp. 78-84, 2002.

10.6 Colored Petri Nets for Integrating the Data Perspective in Process Audits

Number	**6**
Title	**Colored Petri Nets for Integrating the Data Perspective in Process Audits**
Appendix	10.6
Primary Related Chapters	5.2
Type	Conference Paper
Conference	32nd International Conference on Conceptual Modeling (ER 2013)
Reference	(Werner, 2013)
Acceptance Rate	32 %[45)]
VHB JQ 2.1 Ranking	B[45)] (7.59)
WKWI Ranking	B[45)]
ERA 2010	A[45)]
CORE 2013	A[45)]
Review Procedure	Double Blinded
Number of Reviews	3
Authors	Michael Werner
Dissertation Points	1.00
Authorship	
Overall	100%
Design	100%
Realization	100%
Writing	100%
Status	Published
Part of other Dissertations	No
Link	http://link.springer.com/chapter/10.1007%2F978-3-642-41924-9_31

[45] Submitted as full paper (14 pages), accepted as short paper (8 pages), presented at the ER 2013 and published in the conference proceedings, overall acceptance rate for all papers was 32 %

Colored Petri Nets for Integrating the Data Perspective in Process Audits

MICHAEL WERNER

University of Hamburg, Germany
michael.werner@wiso.uni-hamburg.de

Abstract: The complexity of business processes and the data volume of processed transactions increase with the ongoing integration of information systems. Process mining can be used as an innovative approach to derive information about business processes by analyzing recorded data from the source information systems. Although process mining offers novel opportunities to analyze and inspect business processes it is rarely used for audit purposes. The application of process mining has the potential to significantly improve process audits if requirements from the application domain are considered adequately. A common requirement for process audits is the integration of the data perspective. We introduce a specification of Colored Petri Nets that enables the modeling of the data perspective for a specific application domain. Its application demonstrates how information from the application domain can be used to create process models that integrate the data perspective for the purpose of process audits.

Keywords: Business Process Audits, Petri Nets, Business Process Modeling, Process Mining, Business Intelligence, Business Process Intelligence

1 Introduction

The integration of information systems for supporting and automating the operation of business processes in organizations opens up new ways for data analysis. Business intelligence is an academic field that investigates how data can be used for analysis purposes. It provides a rich set of analysis methods and tools that are well accepted and applied in a variety of application domains, but it is rarely used for auditing purposes.

This article deals with the application of Colored Petri Net models that combine the control flow and data perspective for process mining in the context of process audits. We refer to the example application domain of financial audits for illustration purposes. The benefit of this application domain is the fact that the event data which is necessary for the application of process mining displays structural characteristics that are particularly suitable to be used for the integration of a data perspective. These characteristics relate to the structure of financial accounting and are independent from the used source system. The objective of this article is to illustrate an approach for the integration of the data perspective into process models in the context of process mining and process audits. This approach is not restricted to the illustrated example application domain but can be applied in a variety of application scenarios where information about the involved data is available and valuable for process mining purposes.

2 Related Research

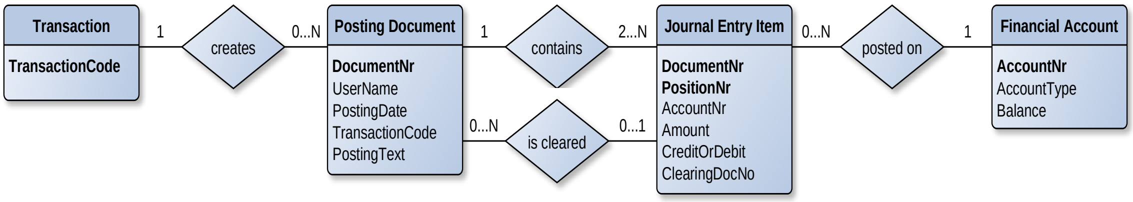
Process Mining is a research area that emerged in the late 1990s. Tiwari et al. provide a good overview of the state-of-the-art in process mining until 2008 [1] and Van der Aalst [2] provides a comprehensive summary of basic and advanced mining concepts that have been researched during the last decade. Jans et al. investigate the application of process mining for auditing purposes [3–5]. They provide an interesting case study [6] for compliance checking by using the Fuzzy Miner [7] implemented in the ProM software framework [8]. The study shows how significant information can be derived from using process mining methods for internal audits. The research results are derived from analyzing the control flow of discovered process models and the interaction of users. We are not aware of any implementations or case studies that consider the data perspective in the context of process mining for process audits. One of the reasons may be the fact that the data perspective in process mining has generally not been investigated extensively yet in the academic community [9, 10]. Exceptions are the research results published by Accorsi and Wonnemann [11] and de Leoni and van der Aalst [10]. The research presented by Accorsi and Wonnemann is motivated by finding control mechanisms to identify information leaks in process models. The authors introduce information flow nets (IFnets) as a meta-model based on Colored Petri Nets that are able to model information flows. Instead of using tokens exclusively for the modeling of the control flow colored tokens are used to represent data items that are manipulated during the process execution. De Leoni and van der Aalst use a different approach. Their intention is to incorporate a data perspective for analyzing why a certain path in a process model is taken for an individual case. The modeled data objects influence the course of routing. They introduce Petri Nets with data (DPN-nets) that base on Petri Nets but that are extended by a set of data variables that are modeled as graphical components in the DPN-nets.

3 Application Domain and Requirements

Financial information is published to inform stakeholders about the financial performance of a company. The published information is prepared based on data that is recorded by information systems in the course of transaction processing. Public accountants audit financial statements for ensuring that the financial information is prepared according to relevant rules and regulations. The understanding of business processes plays a significant role in financial audits [12]. The rationale of considering business processes is the assumption that well controlled business processes lead to complete and correct recording of entries to the financial accounts. Auditors traditionally collect information about business processes manually by performing interviews and inspecting available process documentation. These procedures are extremely time-consuming and error-prone [13]. Process mining allows an effective and efficient reconstruction of reliable process models. Its application would significantly improve the efficiency and effectiveness of financial process audits [14].

Most mining algorithms focus on the reconstruction of control flows in process models that determine the relation and sequence of process activities [9, 10]. Information about control flows is important for auditors to understand the structure of a business process. But this information alone is not sufficient from an audit perspective because an auditor additionally needs to understand how the business processes relate to the entries in the financial accounts [15, 16]. It is therefore necessary to receive information on how the execution of activities in a business process relate to recorded financial entries. This can be achieved by incorporating the data perspective. The relationship between transactions, journal entries and financial accounts is illustrated in Figure 1.

Fig. 1. Accounting Structure Entity-Relationship-Model



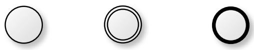
A second important concept in financial process audits is materiality [15, 16]. Auditors just inspect those transactions that could have a material effect on the financial statements. To be able to identify which business processes are material information is needed about the amounts that were posted on the different accounts. It is therefore also necessary to model the value of the posted journal entries in the produced process models.

4 Integrating the Data Perspective

The relevant data objects in financial audits are journal entries. They are created during the execution of a business process but their values do not influence the course of routing. They can be interpreted as passive information objects and we therefore refer to the approach used by Accorsi and Wonnemann [11] for integrating them into the process models. Petri Net places and tokens are normally used in process mining to model the control flow. The general approach for integrating the data perspective is to model data objects as colored tokens that are stored in specific places.

The integration is illustrated in the following specification. A Colored Petri Net can formally be expressed as a tuple CPN = $(T, P, A, \Sigma, V, C, G, E, I)$ [17]. Table 1 presents the formal definition of each tuple element, the used net components, and their meaning when applied in the context of this paper. For ease of reference we refer to this type of nets as Financial Petri Nets (FPN) for the remainder of this paper.

Table 1. Specification of Colored Petri Nets for Process Mining in Financial Audits

T is a finite set of transitions	
	The transitions represent the activities that were executed in the process. They display the name of the activity. Further information, for example the transaction code name, can be added.
P is a finite set of places	
Places in the FPN represent financial accounts and control places.	
Control Places  Source Place, Sequence Place, Sink Place	Control places determine the control flow in a process model. For every process model one source place is modeled that connects to the start transactions. The sink place marks the termination of the process. The control places between the start and end transition determine the execution sequence of the process model. A control place belongs to the set of control places CP.
Account Places  debit side of a balance sheet account  credit side of balance sheet account  credit side of a profit and loss account  debit side of a profit and loss account	The account places represent financial accounts that are affected by the execution of activities in a process. The symbol color indicates the meaning of an account. Account places belong to the set of account places AP.
$A \in P \times T \cup T \times P$ is a set of arcs also called flow relation	
⟶ Control Arc	Control arcs connect control places with transitions. They model the control flow in the model.
- - - -> Posting Arc	Posting arcs illustrate the relationship between activities represented as transitions in the model and financial accounts that are modeled as account places.
<- - - -> Clearing Arc	Clearing arcs are used to model that an activity cleared an entry on the corresponding account. Clearing arcs are double-headed arcs and are used as an syntactical abbreviation for two arcs (p,t) and (t,p).
Σ is a set of non-empty color sets	
colset *VALUES* = double	The color set contains the possible values that are posted or cleared.
colset *ACCOUNTS* = string	The color set contains all account numbers.
colset *ACOUNTTYPE* = boolean	The color set contains {1,0} indicating if the represented account is a balance sheet or a profit and loss account.
colset *CREDorDEB* = boolean	The color contains {1,0} indicating if the account place is a representation of the debit or credit side of an account.
colset *EXECUTIONS* = int	The color contains {1,…,n} indicating how often a path was chosen in the FPN.
colset *ACCOUNTPLACES*	*ACCOUNTPLACES* is the color set as a product of *VALUES* * *ACCOUNTS* * *ACCOUNTTYPE* * *CREDorDEB*
V is a finite set of typed variables such that Type$[v] \in$ Σ for all variables $v \in V$.	
In FPN models arc inscriptions are modeled as constants. Variables are therefore not necessary and $V = \{\}$.	

$C: P \rightarrow \Sigma$ is a color set function that assigns a color set to each place.	
The color set function in FPN assigns different color sets to places depending if they belong to the group of control or account places:	$C(p) \begin{cases} ACCOUNTPLACES \text{ } if \text{ } p \in AP \\ EXECUTIONS \text{ } if \text{ } p \in CP \end{cases}$
$G: T \rightarrow EXPR_V$ is a guard function that assigns a guard to each transition t such that $Type\ [G(t)] = boolean$.	
$EXPR$ is the set of expressions that is provided by the used inscription language. FPN do not explicitly include guards because they do not model dynamic behavior of transitions that depends on specific input but illustrate the processing of already executed processes. The guard function for FPN is therefore defined as $G(t) = true\ for\ all\ t \in T$.	
$E: A \rightarrow EXPR_V$ is an arc expression function that assigns an arc expression to each arc a such that $Type\ [E(a)] = C(p)_{MS}$, where p is the place connected to the arc a.	
The arc expressions in a FPN are constants. The arc expression function assigns to each posting and clearing arc a set of constants that denote the posted or cleared value, the account number, account type and an indicator if it is a credit or debit posting. For each control flow arc the number of execution times is assigned indicating how often this path was chosen in the process model. $E(a) \begin{cases} \{val \in VALUES, acc \in ACCOUNTS, acctyp \in ACCOUNTTYPE, cord \in CREDorDEB\} \\ with\ Type\ [E(a)] = C(p)_{MS} = ACCOUNTPLACES\ if\ p \in AP \\ \{ex \in EXECUTIONS\ with\ Type\ [E(a)] = C(p)_{MS} = EXECUTIONS\ if\ p \in CP \end{cases}$	
$I: P \rightarrow EXPR_\emptyset$ is an initialization function that assigns an initialization expression to each place p such that $Type\ [I(p)] = C(p)_{MS}$	
The initialization function of a FPN assigns initialization expressions to each place as follows: $I(p) \begin{cases} n'ex \in EXECUTIONS\ with\ Type\ [I(p)] = C(p)_{MS} = EXECUTIONS\ if\ p = source \\ \emptyset_{MS}\ otherwise \end{cases}$ Only the source place is initialized in a FPN. The initialization expression for $p = source$ generates n tokens in the initial marking $M_0(p)$, one for each connected start transition. The inscription of each token is a member of the set $EXECUTIONS$.	

Figure 2 illustrates a FPN model for a purchasing process. The model includes:

Transitions: $T = \{MB01, MIRO, F110\}$
Places: $P = \{Source, S1, S2, Sink, 100\_D, 200\_D, 200\_C, 300\_D, 300\_C, 400\_D\}$
$CP \subseteq T = \{Source, S1, S2, Sink\}$
$AP \subseteq T = \{100\_D, 200\_D, 200\_C, 300\_D, 300\_C, 400\_D\}$.
Color sets: $VALUES = \{50{,}000\}$, $ACCOUNTS\{100, 200, 300, 400\}$, $ACOUNTTYPE = \{0,1\}$, $CREDorDEB = \{0,1\}$; $EXECUTIONS = \{1\}$; $ACCOUNTPLACES = VALUES * ACCOUNTS * ACCOUNTTYPE * CREDorDEB$.
Initialization: $I(p) = \{1'1\ for\ p = source\ and\ \emptyset_{MS}\ for\ p \neq source$

The model shows that the processing of received goods created journal entries with the amount of 50,000 on the raw materials and the goods receipt / invoices receipt (GR/IR) account. The receipt of the corresponding invoice for the purchased goods was processed with activity MIRO which led to journal entries on the GR/IR account and a creditor account. It also cleared the open debit item on the GR/IR account that was posted by MB01. The received invoice was finally paid by executing the activity F110 which posted a clearing item on the creditor account and a debit entry on the bank account.

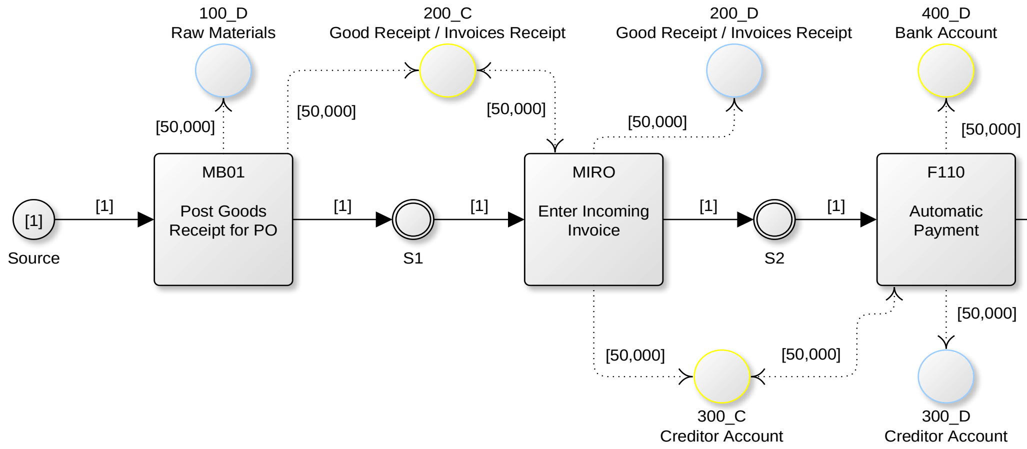


Fig. 2. Simple FPN Example of a Purchase Process

The example in Figure 2 demonstrates how the used Petri Net specification can be used to model the control flow and the data perspective simultaneously in a single model. The transitions in the model create colored tokens when they fire. They store information on the values that are posted on the connected account places. The illustrated model does not only mimic the execution behavior of the involved activities but also the creation of journal entries on the financial accounts. The model shows the execution sequence and the value flows that are produced.

5 Implementation and Experimental Evaluation

We applied the described FPN specification on real world data to evaluate if the theoretical constructs can actually be used in real world settings. The data base for the evaluation included about one million cases of process executions from a company operating in the manufacturing industry. The raw data was extracted from a SAP ERP system and checked for first and second order data defects [18]. We used an adjusted implementation of the Financial Process Mining (FPM) algorithm [19] that is able to produce FPN models. The mining was limited to 100,000 process instances that affected a specific raw materials account. The mining resulted in 113 process variants. They were analyzed in the evaluation phase by observation using the yEd Graph Editor [20] for verifying if the process models presented the desired information adequately. Selected models were further tested using the Renew software [21] for evaluating if correct FPN were created by simulating the execution of the models. The evaluation demonstrated that FPN can be created correctly by using an adapted FPM algorithm. The produced FPN are able to adequately model the control and data flow based on the used real world data.

The same modeling procedure was used in a different scenario to analyze technical customer service processes which is not described in this paper due to place restrictions.

6 Summary and Conclusion

Process Mining is an innovative approach for analyzing business processes but it is rarely used in the context of process audits. The successful application of process mining requires the consideration of domain specific requirements. A common requirement is the incorporation of a data perspective which has not been addressed extensively yet in the academic community. We have introduced a specification of Colored Petri Nets that allows the modeling of the control flow and data perspective simultaneously. We referred to the application domain of financial audits as a representative example to demonstrate how the data perspective can be included by referring to relevant application domain requirements.

The evaluation of mined process models shows the suitability of the presented specification in real world settings. The evaluation included the data from a SAP system of a single company. It can therefore not be concluded that the results also hold true for other companies or ERP systems. But the presented specification bases on the structure of financial accounting and is therefore independent from any proprietary ERP software implementation or industry. Evaluation results from further current research indicate that the presented results are also applicable in other settings.[46]

7 References

1. Tiwari, A., Turner, C.J., Majeed, B.: A review of business process mining: state-of-the-art and future trends. Business Process Management Journal. 14, 5–22 (2008).
2. Van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin Heidelberg (2011).
3. Jans, M., Alles, M., Vasarhelyi, M.: Process mining of event logs in auditing: opportunities and challenges. Working paper. Hasselt University. Belgium (2010).
4. Jans, M.J.: Process Mining in Auditing: From Current Limitations to Future Challenges. In: Daniel, F., Barkaoui, K., and Dustdar, S. (eds.) Business Process Management Workshops. pp. 394–397. Springer, Berlin, Heidelberg (2012).
5. Jans, M., van der Werf, J.M., Lybaert, N., Vanhoof, K.: A business process mining application for internal transaction fraud mitigation. Expert Systems with Applications. 38, 13351–13359 (2011).
6. Jans, M., Alles, M., Vasarhelyi, M.: Process Mining of Event Logs in Internal Auditing: A Case Study. 2nd International Symposium on Accounting Information Systems (2011).
7. Günther, C., van der Aalst, W.: Fuzzy mining–adaptive process simplification based on multi-perspective metrics. Business Process Management. 328–343 (2007).
8. Process Mining Group: ProM, http://www.processmining.org/prom/start.

[46] The research results presented in this paper were developed in the research project EMOTEC sponsored by the German Federal Ministry of Education and Research (grant number 01FL10023). The authors are responsible for the content of this publication.

9. Stocker, T.: Data flow-oriented process mining to support security audits. Service-Oriented Computing-ICSOC 2011 Workshops. pp. 171–176 (2012).

10. De Leoni, M., van der Aalst, W.M.: Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments. (2013).

11. Accorsi, R., Wonnemann, C.: InDico: Information Flow Analysis of Business Processes for Confidentiality Requirements. Security and Trust Management. pp. 194–209. Springer (2011).

12. International Federation of Accountants: ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment, (2012).

13. Werner, M.: Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung. In: Plate, G. (ed.) Forschung für die Wirtschaft. pp. 199–214. Cuvillier Verlag, Göttingen (2012).

14. Werner, M., Gehrke, N., Nüttgens, M.: Business Process Mining and Reconstruction for Financial Audits. Hawaii International Conference on System Sciences. pp. 5350–5359. , Maui (2012).

15. Schultz, M., Müller-Wickop, N., Nüttgens, M.: Key Information Requirements for Process Audits – an Expert Perspective.pdf. Proceedings of the 5th International Workshop on Enterprise Modelling and Information Systems Architectures. , Vienna (2012).

16. Müller-Wickop, N., Schultz, M., Peris, M.: Towards Key Concepts for Process Audits - A Multi-Method Research Approach. Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics. , Utrecht (2013).

17. Jensen, K., Kristensen, L.M.: Coloured petri nets. Springer (2009).

18. Kemper, H.-G., Mehanna, W., Baars, H.: Business intelligence - Grundlagen und praktische Anwendungen : eine Einführung in die IT-basierte Managementunterstützung. Vieweg + Teubner, Wiesbaden (2010).

19. Gehrke, N., Müller-Wickop, N.: Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems. Proceedings of the 16th Americas Conference on Information Systems. , Lima, Peru (2010).

20. yWorks GmbH: yEd - Graph Editor, http://www.yworks.com/de/products_yed_about.html.

21. University of Hamburg: Renew - The Reference Net Workshop, http://www.renew.de/.

10.7 Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits

Number	**7**
Title	**Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits**
Appendix	10.7
Primary Related Chapters	5.3
Type	Conference Paper
Conference	33rd International Conference on Information Systems (ICIS 2012)
Reference	(Werner et al., 2012b)
Acceptance Rate	29 % [47)]
VHB JQ 2.1 Ranking	A[47)] (8.48)
WKWI Ranking	A[47)]
ERA 2010	A[47)]
CORE 2013	A*[47)]
Review Procedure	Double Blinded
Number of Reviews	5
Authors	1. Michael Werner 2. Martin Schultz 3. Niels Müller-Wickop 4. Nick Gehrke 5. Markus Nüttgens
Dissertation Points	0.33
Authorship	
Overall	80%
Design	80%
Realization	80%
Writing	80%
Status	Published
Part of other Dissertations	No
Link	http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1230&context=icis2012

[47] Submitted an accepted as Research-in-Progress paper (12 pages), presented at the ICIS 2014 and published in the conference proceedings

Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits

Research-in-Progress

MICHAEL WERNER

University of Hamburg
Chair for Information Systems
Max-Brauer-Allee 60
D-22765 Hamburg
michael.werner@wiso.uni-hamburg.de

MARTIN SCHULTZ

University of Hamburg
Chair for Information Systems
Max-Brauer-Allee 60
D-22765 Hamburg
martin.schultz@wiso.uni-hamburg.de

NIELS MÜLLER-WICKOP

University of Hamburg
Chair for Information Systems
Max-Brauer-Allee 60
D-22765 Hamburg
niels.mueller-wickop@wiso.uni-hamburg.de

NICK GEHRKE

Nordakademie
Chair for Information Systems
Köllner Chaussee 11
D-25337 Elmshorn
nick.gehrke@nordakdemie.de

MARKUS NÜTTGENS

University of Hamburg
Chair for Information Systems
Max-Brauer-Allee 60
D-22765 Hamburg
markus.nuettgens@wiso.uni-hamburg.de

Abstract: A key objective of implementing business intelligence tools and methods is to analyze voluminous data and to derive information that would otherwise not be available. Although the overall significance of business intelligence has increased with the general growth of processed and available data it is almost absent in the auditing industry. Public accountants face the challenge to provide an opinion on financial statements that are based on the data produced by the automated processing of countless business transactions in ERP systems. Methods for mining and reconstructing financially relevant process instances can be used as a data analysis tool in the specific context of auditing. In this article we introduce and evaluate an algorithm that effectively reduces the complexity of mined process instances. The presented methods provide a part of the foundation for implementing automated analysis and audit procedures that can assist auditors to perform more efficient and effective audits.

Keywords: business process modeling, data mining, process mining, business intelligence, data analysis, enterprise resource planning systems, financial audits

1 Introduction

Enterprise resource planning (ERP) systems are key components for supporting and automating the processing of business transactions in modern companies. While ERP systems are primarily used for supporting and automating business processes they commonly also provide the functionality to prepare the financial statements that companies are required to publish. The provided information plays a critical role in the economic system. It enables stakeholders to acquire information about the financial situation of the entity they are interested in. Due to their informative significance governments and regulatory institutions have issued laws and regulations that intend to safeguard the correctness of published financial statements. They have entrusted public accountants to carry out audits for ensuring that accounting standards are adhered to and that the published information is free of material misstatements. But when auditors perform their audits they encounter a significant problem. While on the one hand business transactions are processed automatically in ERP systems auditors on the other hand apply mainly manual audit procedures to achieve their audit comfort. With increasing integration of ERP systems and rising numbers of processed transactions manual audit procedures become inefficient and ineffective.

This situation is quite astonishing as the precondition for implementing automated audit procedures is given in every case where the mere number of processed transactions makes manual audit procedures inefficient or even ineffective. Whenever this is the case systems must be involved that support or automate the processing. Otherwise the number of transactions could still be effectively audited with manual audit procedures. While processing ERP systems produce data that is stored in the systems' databases. The stored data includes the journal entries and other logging information that can be used to reconstruct relationships between the stored data entries and their corresponding originally executed transactions. The content of the stored financially relevant data in ERP systems is generally suitable for automated process mining and analysis purposes. It is clearly structured and needs to comply with accounting requirements like completeness and accuracy. It is possible to capitalize on the available structures by using purposeful mining and analysis methods.

Gehrke and Müller-Wickop (2010a, 2010b) present an algorithm for mining business processes from journal entries stored in ERP databases. The approach is similar to mining process models from event logs (van der Aalst 2011) but applied to financially relevant business processes. Werner et al. (2012) show how these process mining methods can be combined with automated testing of controls that are embedded in ERP systems. The combination generally allows implementing system based and automated audit procedures that are efficient and effective to audit highly integrated and automated financially relevant business processes. This way the imbalance between automated processing on the companies' side and manual audit procedures on the auditors' side can be overcome.

The application of these methods can be seen as a business intelligence tool that enables especially auditors to receive information from available mass data that they otherwise do not have access to. Their application would enable the auditor to efficiently gather information for understanding the relationship between the business processes and the financial statements for the audited entity. They would also allow the identification of unusual business transactions that normally inherit higher risk than standard transactions (Werner and Gehrke 2011).

In this paper we focus on the results and issues that arise from analyzing mined process instances. We therefore address a specific sub-problem on the path towards developing automated analysis and audit methods for financial audits. Analyzing mined processes from real life data shows that process instances do exist that contain up to tens of thousands executed transactions leading to very complex graphs consisting of hundreds of thousands elements. The complexity of these graphs does not allow sophisticated interpretation for the purpose of auditing. It is necessary to find mechanisms to reduce their complexity.

In this article we illustrate how the complexity of mined process instances can be reduced by using a graph transformation system. We start with an overview of related work and chosen research methodology. We continue with an illustration of how mined process instances can be represented as Petri nets by using an illustrative example of a mined process instance. Based on this representation we introduce an aggregation algorithm that operates as a graph transformation system by aggregating net elements within a process instance. We graphically show the aggregation results for the used example and provide evaluation results that were derived from applying the aggregation algorithm to test and real life data. The paper closes with a summary and conclusion of the presented methods and derived results.

We focus our attention on financial audits in order to stay in reasonable limits but we like to point out that the application of the discussed process mining, reconstruction and graph transformation methods is not restricted to financial audits but is also relevant for performance and optimization considerations (Werner and Gehrke 2011).

2 Related Work and Research Methodology

The concept of process mining forms the foundation of the research presented in this paper. Process mining first evolved in the 1990s for investigating process mining in the context of software engineering (Cook and Wolf 1998a, 1998b, 1999). The idea of applying process mining to workflow logs was first introduced by Agrawal et al. (1998). Maxeiner et al. (2001) and Schimm (2000; 2001a; 2001b; 2002) developed mining tools whereas Herbst and Karagiannis (Herbst 2000a; 2000b; 2003; Herbst and Karagiannis 1998; 1999) addressed process mining in the context of workflow management. Substantial research work exists for mining and rediscovering process models from event logs (van der Aalst 1997, 1998, 2005, 2011; van der Aalst et al. 2002a, 2002b; van der Aalst and Dongen 2002; van der Aalst and Weijters 2002; Maruster et al. 2001a, 2001b; Weijters and van der Aalst 2001). Their work covers a variety of aspects like delta analysis, conformance, concurrency, workflow verification and performance. Gehrke and Müller-Wickop (2010a, 2010b) build a bridge between process mining and financial accounting by applying process mining to reconstruct financially relevant business processes.

A traditional application area of business intelligence is the support for decision making processes (Turban et al. 2007; Vercellis 2009). Anandarajan et al. (2004) illustrate business intelligence techniques from an accounting and finance perspective but we are not aware of publications that address business intelligence tools from an explicit auditing perspective. For developing an aggregation method we refer to the theory of graph grammars and transformation (Heckel 2006; Rozenberg 1997).

The research presented in this paper follows a design science approach (March and Smith 1995; Österle et al. 2010; Hevner et al. 2004). For the purpose of developing adequate complexity reduction methods we implemented a software artifact. The artifact allows reconstructing and observing process instances. Based on the observation derived from extensive real life data we engineered formal methods (Brinkkemper 1996) and finally evaluated them against test and real life data.

3 Automated Reconstruction and Petri Net Representation

The mined processes presented by Gehrke and Müller-Wickop (2010a, 2010b) use a more or less informal presentation that focusses on illustrating essential process components. Müller-Wickop et al. (2011) introduce a Business Process Model and Notation (BPMN) based representation which primarily aims to integrate the process and financial perspective by including graph elements that capture financially relevant aspects of modeled process instances. In order to be able to develop comprehensive methods for complexity reduction we need a formally robust presentation. We chose a Petri net representation due to the fact that Petri nets constitute a formally sound and mathematically powerful modeling language (Valk 2008). Modeling mined process instances as colored Petri nets makes it possible to apply already existing research for event log mining (van der Aalst 2011).

Figure 1 shows a reconstructed instance of a purchasing process that was mined from an ERP system by using the mining algorithm developed by (Gehrke and Müller-Wickop 2010b). This means that a purchasing business transaction was executed and recorded in the examined ERP system. The relevant data was extracted and the process instance was reconstructed. The illustrated example represents an executable colored Petri net that mimics the behavior of the originally processed instance. A colored Petri net can formally be expressed as a tuple $N = (P, T, F, C, cd, W, m_0)$. The specific meaning of each tuple element for the constructed instance is described in Table 1 and can be illustrated by referring to Figure 1. The transitions of the Petri net shown in Figure 1 represent executed transactions that created journal entries in the ERP system. MB01 are transactions that recorded the receipt of ordered material, MR1M processed the received invoices and F110 executed the payment run. FB1S represent automatic clearing transactions. The set of transition T for this Petri net contains all the transitions that represent the executed transactions {MB01_5000004383, MB01_5000004384, …}. Journal entry items are modeled as places with inscriptions identifying the account the corresponding item was posted to. The different colors illustrate whether the item was a debit or credit posting. The created-relationships between items and transactions are illustrated as dotted arrows. The implemented mining algorithm exploits the open-item-accounting structure of journal entries. Journal entries with enabled open-item-accounting are either cleared or not cleared. When a process has been successfully terminated all open items are cleared. Otherwise the process has not been completely executed. A clearing document exists for each cleared item. By following this connection between journal entries the whole process instance can be reconstructed. The cleared-relationships between transactions and items are illustrated as dotted arcs. The inscriptions on the arcs represent the amounts that were posted or cleared on the accounts the items were placed on. The modeling of start places marked with tokens representing the document number that was created by the corresponding transaction ensures that each transition in the Petri net model can actually be executed. The flow relation

F for this Petri net contains all arcs between the transitions and places. The set of places includes all illustrated places (start places and places for journal entry items). The set of colors includes the set of account numbers {310000, 191100, …}, indicators for credit or debit posting {credit, debit} and document numbers {5000004383, 5000004384, …}. The function cd maps the colors to the individual places. The function W maps the arc inscriptions that represent the booking values {13907.17, 10880.29, …} to the arcs for relations between transitions and places representing journal entry items or the document numbers to the arcs between start places and transitions.

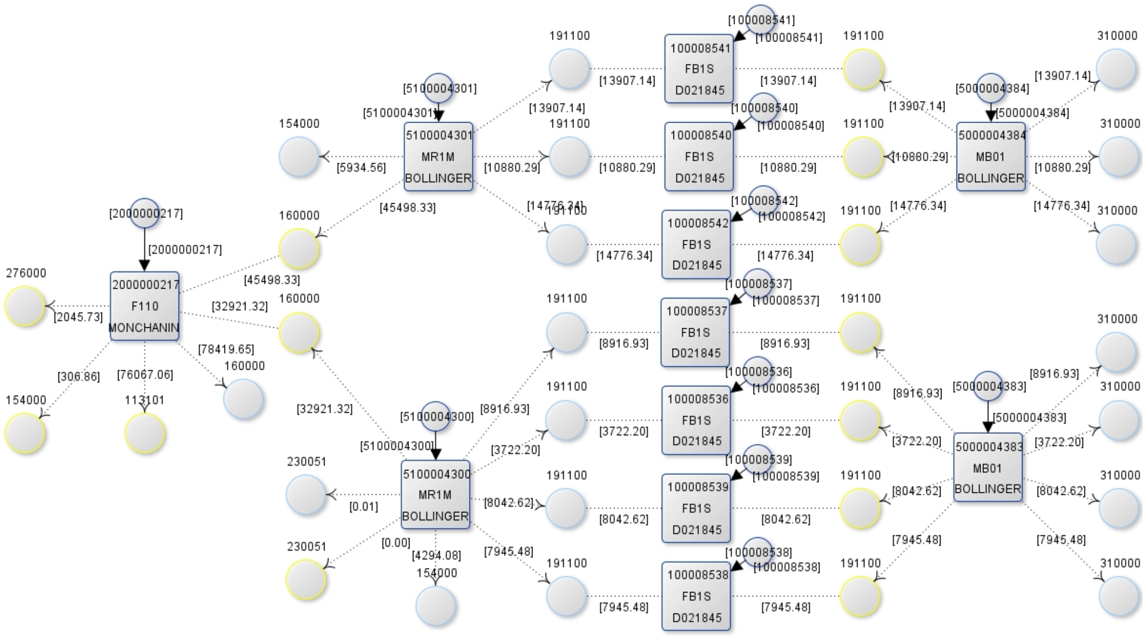


Figure 1. Simple Process Instance A

Table 1. Formal Petri Net Representation

Tuple elements		Specific meaning in the representation context
P	Set of places	Start places and places representing journal entry items
T	Set of transitions	Transactions executed in the ERP system
F	Flow relation	Arcs between the nodes indicating their type of relationship
C	Set of colors	Set of place characteristics, posting values, document numbers
cd	Color domain mapping	Mapping of characteristics to places
W	Arc inscription	Mapping of posting values or document numbers to arcs
m_0	Initial marking	Tokens marking start places with the corresponding document number

Example A contains twelve transitions. The instance is clearly interpretable by observation. It changes for more complex process instances. Example B illustrated in Figure 2 shows a mined process instance like example A and consists of the same Petri net elements as described in Table 1 but it contains 358 transitions. It is still a small instance compared to the most complex instances in our sample database that included up to 27,177 transitions. Table 2 provides an overview of the characteristics of both examples. When looking at Figure

2 it is obvious that the interpretation and analysis of reconstructed instances by simple observation becomes impossible when they include more than a few transactions.

Figure 2. Complex Process Instance B

Table 2. Net Characteristics for Process Instance Examples

Example Instance	A	B
Number of transitions	12	358
Number of places	43	1804
Number of arcs	60	2549
Sum of net elements	115	4711

4 Graph Transformation for Complexity Reduction

The complexity of a graph can be defined differently, depending on the relevant perspective und purpose (Neel and Orrison 2006). A graph is generally defined as a pair of disjoint sets G = (V,E) where V is the set of vertices and E is the set of edges with E ⊆ V2 (Diestel 2010). For the purpose of this paper we consider the complexity of the graph simply as a function of its number of edges and vertices. The mined process instances are modeled as Petri nets. We therefore use the term net elements for the sum of vertices and edges that determine the complexity of the Petri net under review. A promising approach for reducing complexity of voluminous process instances is to aggregate similar net elements and thereby reducing their total number. A key requirement for any type of transformation is that financially relevant information remains unchanged compared to the originally reconstructed process. In the context of auditing it is crucial to be able to trace any transaction from the point of origin to the final ledger posting. The path that allows tracing a transaction through an information systems is called audit trail (Romney and Steinbart 2008 p. 687). In the context of process mining for audit purposes this means that the analyzed data has to mirror exactly the transactions that were actually executed and that no existing paths in the graph may be deleted or new ones be added. Müller-Wickop et al. (2011) point out that the behavior of the process instances may not be changed by any aggregation method. If an aggregation algorithm changed the behavior, this fundamental requirement would be violated. The behavior of a Petri net can be expressed as the set of possible firing sequences. For designing an aggregation method it is necessary to ensure that the set of firing sequences remains unchanged.

Requirement I: The set of firing sequences has to stay constant

Different arc types in the Petri net mean that transactions interact differently with journal entry items. If the arc types between transitions and places were altered the resulting graph would no longer reflect the original relationship between these elements. The arc types and therefore the type of interaction between places and transitions may not be altered.

Requirement II: Different arc types may not be merged

The methods of financial process mining by Gehrke and Müller-Wickop (2010b) have been developed to exploit the structure of accounting data and to capture financially relevant information. This information primarily concerns the value flow within the processes. Analyzing this information actually allows to identify which amounts have been posted on the different accounts by a process instance and to evaluate how significant they are from a materiality perspective.

Requirement III: The arc inscriptions representing the value of posted journal entries have to be preserved

We reconstructed and analyzed processes instances from different data sets originating from two companies that operate in the retail and manufacturing industries and from the SAP IDES test system (SAP 2012). Figure 3 provides an overview of the frequency distribution of approximately 40,000 mined process instances originating from a corporation in the retail industry with a logarithmic scaling on the x- and y-axes. The mean value of the distribution is 2.30, the standard deviation is 3.37, the median value is 2 and the maximum value is 596. The frequency distributions for the other data sets show the same pattern with slightly different values.

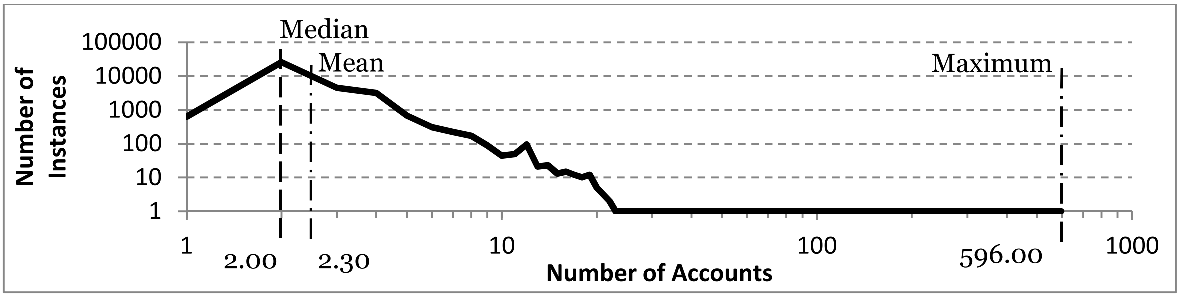
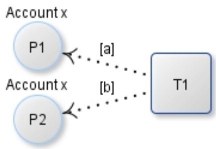
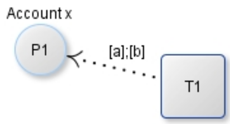
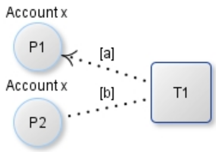
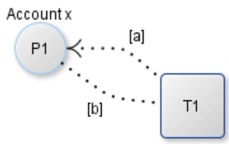
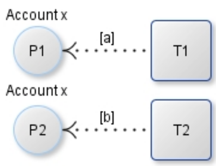
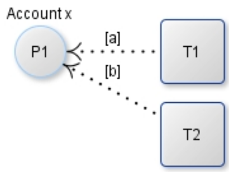
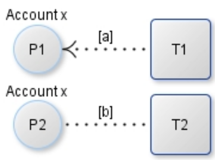
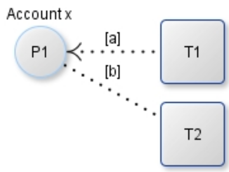


Figure 3. Frequency Distribution of Instances over Accounts

Figure 3 shows that the number of different accounts in process instances is relatively small compared to the instance size. This observation is comprehensible when considering how business processes generally affect accounts. The execution of a certain business process normally produces journal entries only on a very limited subset of the overall available accounts. These are the accounts related to that business process. The execution of a standard procurement process would most likely lead to journal entries on the expense accounts but not on the sales accounts. Although this observation cannot be generalized without further research it is reasonable to assume that it is true also for other industries not covered by the analyzed data sets.

In the Petri net models journal entry items are represented as places carrying an inscription denoting the account they were posted to. Due to the fact that only few accounts are used within the same process instance it is reasonable to assume that the aggregation of places that represent journal entry items on the same accounts might significantly reduce the overall number of net elements. When designing an algorithm for aggregating places requirements I to III have to be considered. The following cases listed in Table 3 illustrate the different constellations that can occur when two places representing journal entry items are to be aggregated. Places are only considered for aggregation if they carry the same account number and credit or debit flag. The case description contains the possible constellations according to the net definition used in this paper and shows the cases for arcs directed from transitions to places (incoming arcs). These constellations also have to be considered for arcs directed from places to transitions (outgoing arcs). Designing an algorithm based on these cases ensures that the set of firing sequences is not changed because the relationship between the transactions is preserved as well as the different arc types connecting the places and transitions. By maintaining inscriptions from merged arcs we ensure that the financially relevant information on posting values does not get lost.

Table 3. Case Distinction for Place Aggregation

	Input	Description	Output
Case 1		P1 and P2 are connected to the same transition T1. The arc types of both arcs are equal. The arcs (T1,P1) and (T1,P2) can be aggregated. The inscriptions [b] is added to [a] resulting in [a];[b]. Place P2 is deleted.	
Case 2		P1 and P2 are connected to the same transition T1. But the arc types are different. The arcs cannot be aggregated. Arc (T1,P2) has to be redirected to P1 resulting in (T1,P1). Place P2 is deleted.	
Case 3		P1 and P2 are connected to different transitions T1 and T2. The arc types of both arcs are equal. The arcs (T1,P1) and (T2,P2) cannot be aggregated in order to preserve the information that the represented items were actually created (or cleared) by different transactions. Arc (T2,P2) has to be redirected to P1 resulting in (T2,P1). Place P2 is deleted.	
Case 4		P1 and P2 are connected to different transitions T1 and T2. The arc types of both arcs are different. The arcs (T1,P1) and (T2,P2) cannot be aggregated. The arc types are different and they originate from different transitions. Arc (T2,P2) has to be redirected to P1 resulting in (T2,P1). Place P2 is deleted.	

The following algorithm in Listing 1 implements the aggregation of places according to the description included in Table 3.

Listing 1. Aggregation Algorithm

```
P_Item ⊆ P      set of all places representing journal entry items in the net
P_ItemAgg= ∅    initially empty set for aggregated places
F_Arcs          set of all arcs in the net
Aggregate Places
While P_Item≠ ∅
      Take p_i ∈ P_Item
      Select all p_j ∈ P_Item with cd(p_i)=cd(p_j)
      Merge arcs for each p_i and p_j
      Add p_i to P_ItemAgg
      Remove p_i and p_j from P_Item
Set P_Item=P_ItemAgg
Merge Arcs
Get incoming arcs F_IncArcsI ⊆ F_Arcs for p_i
Get incoming arcs F_IncArcsJ ⊆ F_Arcs for p_j
For each arc a_i(t_m,p_i) ∈ F_IncArcsI and arc a_j(t_n,p_j) ∈ F_IncArcsJ
      If the arc type of a_i = arc type of a_j
            And if t_m = t_n then add W(a_j) to W(a_i)and      /Case 1
            remove a_j from F_Arcs
      Else set a_j(t_m,p_j)                                    /Case 2,3 and 4
Get outgoing arcs F_OutArcsI ⊆ F_Arcs for p_i
Get outgoing arcs F_OutArcsJ ⊆ F_Arcs for p_j
For each arc a_i(p_i,t_m) ∈ F_OutArcsI and arc a_j(p_j,t_n) ∈ F_OutArcsJ
      If the arc type of a_i = arc type of a_j
            And if t_m = t_n then add W(a_j) to W(a_i)and      /Case 1
            remove a_j from F_Arcs
      Else set a_j(p_j,t_m)                                    /Case 2,3 and 4
```

The algorithm represents a graph transformation production that iteratively substitutes sub-graphs consisting of two nodes and an arbitrary number of arcs into sub-graphs consisting of one node and the same or smaller set of arcs.

5 Evaluation

The application of the aggregation algorithm on example instance A leads to the graph illustrated in Figure 4. The number of net elements was reduced from 115 to 79. We used the software Renew (University of Hamburg 2012) for verification purposes. Renew allows to simulate the execution of Petri nets. The testing of samples showed that the algorithm works properly and generates fully reachable Petri nets. We applied the algorithm to two different data sets. Data set 1 originates from the SAP IDES database. The database is available for universities participating in the SAP University Alliance Program (SAP 2012). The database contains over 115,000 journal entries from approximately 81,000 process instances. We chose the SAP IDES database to enable interested readers to reproduce our results. Data set 2 originates from a corporation operating in the retail industry. The used database contains approximately 90,000 journal entries constituting about 40,000 process instances for a period of one year. Table 4 illustrates the characteristics of the distribution of instance frequency over the number of net elements before and after applying the aggregation algorithm. The mean value for the number of net elements per instance is reduced by 23.2 % from 15.31 to 11.76 for the SAP IDES data set and by 23.4 % from 21.18 to 16.22 in data set 2. We achieve an average complexity reduction of approximately a quarter. The complexity reduction is more effective for complex processes with many net elements. We therefore analyzed the effectiveness of the algorithm by selecting a subset from the available data set 2 containing process instances with 100 or more net elements. The mean value of the original subset was 929.48 net elements per instance. The mean value of the aggregated instances was reduced by 44% to 520.

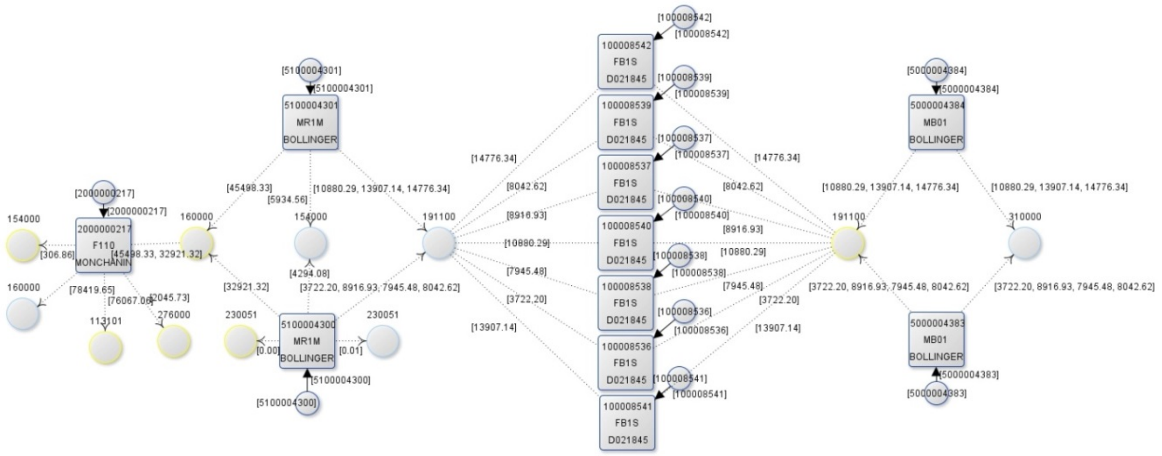


Figure 4. Aggregated Process Instance A

Table 4. Aggregation Results				
	Data Set 1 SAP IDES		**Data Set 2 Retail Company**	
	original	aggregated	original	aggregated
Mean value of net elements per instance	15.31	11.76	21.18	16.22
Median value of net elements per instance	9.00	7.00	9.00	9.00
Maximum of net elements per instance	32,519	18,602	275,870	146,620
Standard deviation of number of net elements	127.83	73.33	1,380.57	743.81

6 Conclusion

Public auditors face the challenge of auditing financial statements that base on the data generated by automated transaction processing in ERP systems. While business intelligence techniques have introduced new ways of analyzing data in the general corporate context such techniques are missing in the auditing industry. The mining and reconstruction of financially relevant processes provides the basis for analyzing data from ERP systems for the purpose of financial audits. We have presented an aggregation algorithm in this paper that operates on mined process instances modeled as Petri nets that can be applied to any process instance mined with the used mining algorithm. It aggregates places within the instances and thereby reduces the number of net elements in the graph. The results are less complex models. The evaluation of the algorithm on the basis of test and real life data shows that significant complexity reductions can be achieved. The reduction effect is higher for process instances encompassing many net elements. With the application of the presented mining methods and representation form it is possible to derive and visualize information about executed processes and their effect on the financial accounts that the public accountant has to audit. They provide a means to overcome the imbalance between mainly manual audit procedures on the auditor side and the automated and system based processing on the company side. The usage of these methods could make financial audits more efficient and set free resources for the evaluation of unusual transactions that commonly constitute higher risks than standard transactions.

When using the methods several limitations should be taken into account. The availability of necessary data is the first. The data needs to be extracted from the ERP system, transformed into a data format that can be processed by the mining algorithm and loaded into a database where it can be accessed for mining purposes. This procedure is called ETL (extract, transform, load) process and requires efficient extraction tools that are able to handle voluminous data. When analyzing the data confidentiality and privacy aspects need to be considered. The use of pseudonymization procedures in the ETL process might address this restriction. A second limitation derives from the current scope of application for the mining algorithm. The algorithm for mining financially relevant processes is only applicable for transactions that affect open-item-operated accounts. A purchase requisition sub-process for example does not directly affect open-item-operated accounts but might be relevant for financial audits. Further development for incorporating financially relevant processes and sub-processes that do not affect open-item-operated accounts might significantly enlarge the scope of application.

The presented methods provide a new approach of analyzing financial data in ERP systems. The presented algorithm achieves significant complexity reductions. But although the number of net elements for the most complex process instance on our sampled data was reduced by almost half from 275,870 to 146,620 the aggregated instance is still too complex for meaningful interpretation and evaluation. First observations show that similar structures among sub-graphs within instances might be common. Frequent sub-graph mining is a well-studied data mining problem (Huan et al. 2003; Yan and Han 2002) The development of aggregation procedures that merge sub-graphs within an instance by considering existing sub-graph mining algorithms might be a promising approach for further complexity reduction. Another obstacle to efficient evaluation is the huge amount of process instances that need to be considered. The analysis of mined process instances leads to the assumption that a great amount of process instances are very similar in structure. Further research is needed to investigate if efficient procedures that create clusters or categories across different process instances can be developed.

We have shown how a process mining approach can be used as a data analysis tool especially in the context of financial audits and how the complexity of mined instances can be reduced effectively. Further research is needed to answer open questions and to overcome existing limitations. But the methods discussed in this paper provide a meaningful milestone on the way for designing system based and automated analysis and audit procedures that cannot only be used in the context of auditing but also in the wider business context for example for performance measurement or optimization purposes.

7 Acknowledgement

The research results presented in this paper were developed in the research project Virtual Accounting Worlds. The project is sponsored by the German Federal Ministry of Education and Research. The authors are responsible for the content of this publication (grant number 01IS10041).

8 References

van der Aalst, W. M. P. 1997. “Verification of Workflow Nets,” In Application and Theory of Petri Nets, P. Azema and G. Balbo (eds.), Berlin: Springer-Verlag, pp. 407–426.

van der Aalst, W. M. P. 1998. “The Application of Petri Nets to Workflow Management,” Journal of Circuits, Systems and Computers (8:I), pp. 21–66.

van der Aalst, W. M. P. 2005. “Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing,” Requirements Engineering Journal (10:3), pp. 198–211.

van der Aalst, W. M. P. 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes, (1st ed,)Springer Berlin Heidelberg.

van der Aalst, W. M. P., and Dongen, B. F. 2002. “Discovering Workflow Performance Models from Timed Logs,” In Engineering and Deployment of Cooperative Information Systems, Y. Han, S. Tai, and D. Wikarski (eds.), (Vol. 2480)Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 45–63.

van der Aalst, W. M. P., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., and Weijters, A. J. M. M. 2002. “Workflow Mining: a Survey of Issues and Approaches,” Beta: Research School for Operations Management and Logistics (Working Paper 74).

van der Aalst, W. M. P., and Weijters, A. 2002. “Rediscovering Workflow Models from Event-Based Data,” In Proceedings of the Third International NAISO Symposium on Engineering of Intelligent Systems (EIS 2002)Presented at the Third International NAISO Symposium on Engineering of Intelligent Systems (EIS 2002).

van der Aalst, W. M. P., Weijters, A. J. M. M., and Maruster, L. 2002. “Workflow Mining: Which Processes can be Rediscovered?,” BETA Working Paper Series (Working Paper 74).

Agrawal, R., Gunopulos, D., and Leymann, F. 1998. “Mining Process Models from Workflow Logs,” In Proc. Sixth Int’l Conf. Extending Database Technology, pp. 469–483.

Anandarajan, M., Anandarajan, A., and Srinivasan, C. A. 2004. Business intelligence techniques : a perspective from accounting and finance, Berlin, Germany; New York: Springer-Verlag.

Brinkkemper, S. 1996. “Method engineering: engineering of information systems development methods and tools,” Information and Software Technology (38:4), pp. 275–280.

Cook, J. E., and Wolf, A. L. 1998a. “Discovering models of software processes from event-based data,” ACM Trans. Softw. Eng. Methodol. (7:3), pp. 215–249.

Cook, J. E., and Wolf, A. L. 1998b. “Event-based detection of concurrency,” SIGSOFT Softw. Eng. Notes (23:6), pp. 35–45.

Cook, J. E., and Wolf, A. L. 1999. “Software process validation: quantitatively measuring the correspondence of a process to a model,” ACM Trans. Softw. Eng. Methodol. (8:2), pp. 147–176.

Diestel, R. 2010. Graph theory, (4th ed,)Heidelberg; New York: Springer.

Gehrke, N., and Müller-Wickop, N. 2010a. "Rekonstruktion von Geschäftsprozessen im Finanzwesen mit Financial Process Mining," In Lecture Notes in Informatics, Proceedings der Jahrestagung Informatik Presented at the Jahrestagung Informatik 2010, IT-supported Service Innovation and Service Improvement, Leipzig.

Gehrke, N., and Müller-Wickop, N. 2010b. "Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems," In Proceedings of the 16th Americas Conference on Information SystemsPresented at the Americas Conference on Information Systems, Lima, Peri.

Heckel, R. 2006. "Graph Transformation in a Nutshell," Electronic Notes in Theoretical Computer Science (148:1), pp. 187–198.

Herbst, J. 2000a. "A Machine Learning Approach to Workflow Management," In Proceedings ofMachine Learning: 11th European Conference Machine Learning, R. López de Mántaras and E. Plaza (eds.), (Vol. 1810)Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 183–194.

Herbst, J. 2000b. "Dealing with concurrency in workflow induction," In European Concurrent Engineering Conference. SCS Europe.

Herbst, J. 2003. Ein induktiver Ansatz zur Akquisition und Adaption von Workflow-Modellen, Tenea Verlag Ltd.

Herbst, J., and Karagiannis, D. 1998. "Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models," In Proceedings 9th International Workshop on Database and Expert Systems Applications (DEXA'98), pp. 745–752.

Herbst, J., and Karagiannis, D. 1999. "An inductive approach to the acquisition and adaptation of workflow models," In Proceedings of the IJCAI (Vol. 99), pp. 52–57.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," Mis Quarterly (28:1), pp. 75–105.

Huan, J., Wang, W., and Prins, J. 2003. "Efficient mining of frequent subgraphs in the presence of isomorphism," In Data Mining, 2003. ICDM 2003. Third IEEE International Conference on, pp. 549–552.

March, S. T., and Smith, G. F. 1995. "Design and natural science research on information technology," Decis. Support Syst. (15:4), pp. 251–266.

Maruster, L., Van Der Aalst, W. M. P., Weijters, A., van den Bosch, A., and Daelemans, W. 2001. "Automated discovery of workflow models from hospital data," In Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001), pp. 183–190.

Maruster, L., Weijters, A., van der Aalst, W., and van den Bosch, A. 2001. "Process mining: Discovering direct successors in process logs," In Discovery Science, pp. 32–36.

Maxeiner, M. K., Küspert, K., and Leymann, F. 2001. "Data Mining von Workflow-Protokollen zur teilautomatisierten Konstruktion von Prozessmodellen," In Datenbanksysteme in Büro, Technik und Wissenschaft (BTW), 9. GI-Fachtagung, Springer-Verlag, pp. 75–84.

Müller-Wickop, N., Schultz, M., Gehrke, N., and Nüttgens, M. 2011. “Towards Automated Financial Process Auditing: Aggregation and Visualization of Process Models,” In Proceedings of the Enterprise Modelling and Information Systems Architectures Presented at the EMISA 2011, Germany.

Neel, D. L., and Orrison, M. E. 2006. “The linear complexity of a graph,” the electronic journal of combinatorics (13:R9), pp. 1.

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., and Sinz, E. J. 2010. “Memorandum on design-oriented information systems research,” European Journal of Information Systems (20:1), pp. 7–10.

Romney, M. B., and Steinbart, P. J. 2008. Accounting Information Systems, (11th Revised edition (REV),)Prentice Hall.

Rozenberg, G. 1997. Handbook of Graph Grammars and Computing by Graph Transformation, (illustrated ed, , Vols. 1-3, Vol. Volume 1 Foundations)Singapore: World Scientific Pub Co.

SAP. 2012. “SAP-UCC,” available at http://www.sap-ucc.com accessed on 29th May 2012.

Schimm, G. 2000. “Generic Linear Business Process Modeling,” In Conceptual Modeling for E-Business and the Web, S. W. Liddle, H. C. Mayr, and B. Thalheim (eds.), (Vol. 1921)Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 31–39.

Schimm, G. 2001a. “Process Mining Elektronischer Geschäftsprozesse,” In Proceedings Elektronische Geschäftsprozesse.

Schimm, G. 2001b. “Process Mining linearer Prozessmodelle - Ein Ansatz zur Automatisierten Akquisition von Prozesswissen,” In Proceedings 1. Konferenz Professionelles Wissensmanagement.

Schimm, G. 2002. “Process Miner — A Tool for Mining Process Schemes from Event-Based Data,” In Logics in Artificial Intelligence, S. Flesca, S. Greco, G. Ianni, and N. Leone (eds.), (Vol. 2424)Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 525–528.

Turban, E., Aronson, J. E., Liang, T.-P., and Sharda, R. 2007. Decision support and business intelligence systems, Upper Saddle River, N.J.; London: Pearson Education International.

University of Hamburg. 2012. “Renew - The Reference Net Workshop,” available at http://www.renew.de accessed on 29th May 2012.

Valk, R. 2008. “Lecture Notes: Formale Grundlagen der Informatik II (FGI 2) Modellierung & Analyse paralleler und verteilter Systeme”, Universität Hamburg.

Vercellis, C. 2009. Business intelligence, Chichester: Wiley.

Weijters, A., and Van der Aalst, W. M. P. 2001. “Process mining: discovering workflow models from event-based data,” In Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2001), pp. 283–290.

Werner, M., and Gehrke, N. 2011. “Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen,” In Forschung für die Wirtschaft Shaker Verlag.

Werner, M., Gehrke, N., and Nüttgens, M. 2012. "Business Process Mining and Reconstruction for Financial Audits," In Proceedings of the 45th Hawaii International Conference on System Sciences Presented at the 45th Hawaii International Conference on System Sciences, pp. 5350–5359.

Yan, X., and Han, J. 2002. "gSpan: graph-based substructure pattern mining," IEEE Comput. Soc, pp. 721–724.

10.8 Improving Structure: Logical Sequencing of Mined Process Models

Number	**8**
Title	**Improving Structure: Logical Sequencing of Process Models**
Appendix	10.8
Primary Related Chapters	5.4
Type	Conference Paper
Conference	47th Hawaii International Conference on System Sciences (HICSS 2014)
Reference	(Werner and Nüttgens, 2014)
Acceptance Rate	56 %
VHB JQ 2.1 Ranking	C (6.44)
WKWI Ranking	B
ERA 2010	A
CORE 2013	A
Review Procedure	Double Blinded
Number of Reviews	4
Authors	1. Michael Werner 2. Markus Nüttgens
Dissertation Points	0.67
Authorship	
Overall	95%
Design	95%
Realization	95%
Writing	95%
Status	Published
Part of other Dissertations	No
Link	http://www.computer.org/csdl/proceedings/hicss/2014/2504/00/2504d888.pdf

Improving Structure: Logical Sequencing of Mined Process Models

MICHAEL WERNER
University of Hamburg, Germany
michael.werner@wiso.uni-hamburg.de

MARKUS NÜTTGENS
University of Hamburg, Germany
markus.nuettgens@wiso.uni-hamburg.de

Abstract: The increasing availability of digital data offers new opportunities for analyzing business processes. Process aware information systems like Enterprise Resource Planning systems store data in the course of transaction processing. This data can be exploited by using process mining techniques. Process mining algorithms produce process models by analyzing recorded event logs. A fundamental challenge in process mining is the creation of purpose-oriented and useful process models. Process mining algorithms commonly refer to the temporal ordering of events for determining the control flow in reconstructed process models. We show how the logical sequence of events can be used instead of the temporal for reconstructing the control flow in mined process models. The exploitation of the logical structure of available event log records opens up new ways to receive purpose-oriented, less complex, and more informative process models.

1 Introduction

Data explosion [1] and information overflow [2] are well known phenomena that accompany the increasing integration of information technology into society and business. The availability of huge amounts of digital data provides extensive opportunities for novel data analyses. But the handling and examination of voluminous data sets also creates new challenges that need to be addressed thoroughly to provide well-researched solutions that can be applied in practice. The analysis of voluminous data has recently gained increased attention in the academic community under the term Big Data [3]. Its implications are not limited to the field of information systems or computer science but are also relevant for other scientific disciplines like biology [4] or physics [5]. Business intelligence (BI) is a research domain that provides methods and tools to deal with big data. Process mining is a BI approach that uses recorded event log data to provide information about business processes. It can be seen as a bridge between BI and business process management [6]. Process mining algorithms produce process models. They reconstruct models by analyzing the available source event logs. A fundamental challenge in process mining is the creation of process models that are useful for the intended purpose. Mined process models are often too complex for simple interpretation and analysis. Extremely complex process models are referred to as lasagna and spaghetti processes in the process mining community due to their graphical characteristics of layered or intertwined arcs in the respective process models [2]. Complexity reduction of process models has been addressed by scholars following different approaches. Advanced mining algorithms like the Fuzzy Miner [7] provide functionality to abstract from infrequent events and execution paths and are able to deliver less complex process models. But infrequent behavior might indeed be relevant for compliance [8] and conformance checking purposes [9] and their omission in the process model might lead to process models that are of little use in these contexts. Other scholars suggest abstraction methods like aggregation and reduction for reducing complexity in process models [10], [11]. Complexity is just one criterion that is relevant for evaluating the quality of a process model. The quality and value of a process model has to be interpreted in the context of the intended purpose and cannot be generally defined [12]. A process model is purpose-oriented if it fulfills the requirements that can be derived from the aimed goal or objective of a process mining attempt.

We suggest an alternative approach. Our aim is to provide process models that are less complex and that provide information on the logical structure of business processes by exploiting the logical structure of recorded events in an event log. This approach creates more informative models according to relevant requirements of the application domain compared to traditional approaches that use the temporal structure of events. Common process mining algorithms like deterministic [2], heuristic [13] and genetic [14] algorithms use the temporal ordering of recorded events for reconstructing the control flow in a process model. We illustrate in this article how information from the application domain can be used to enable the reconstruction of process models based on the logical ordering.

The selected application domain refers to the auditing industry. Companies publish financial information for informing stakeholders about the financial performance. These financial statements are subject to audits due to their significance for the well-functioning of economic markets. An important step in financial audits is the auditing of business processes. It is assumed that well-controlled business processes most likely lead to correct recording of financially relevant transactions in the financial accounts. It is more efficient to audit the overall business process structure than inspecting individual business transactions. A critical requirement for the application of process mining in financial audits is the reliability of produced process models and the preservation of the audit trail [15]. The audit trail describes the path in an information system that allows following an entry on the financial accounts back to its point of origin and vice versa. Auditors use process models to gain an understanding of the audited business processes and to identify incorrect transaction processing. These process models are commonly created using traditional audit procedures like interviews and inspections of available documents which are highly time-consuming and error-prone. Process mining can be used to produce reliable process models very efficiently. But errors or omissions in the models that can be a result of the mining process might lead to misinterpretations in the audit. Mining algorithms that produce unfitting and imprecise [16] process models are therefore generally not suitable for the application in financial audits.

We illustrate in this article how the logical structure of recorded events can be used for reconstructing the control flow in mined process models. Entries in financial accounts exhibit a specific structure. Enterprise Resource Planning (ERP) systems produce entries on financial accounts for each financially relevant transaction that is processed in the system [17]. The relationships between journal entries can be exploited to mine process models [18], [19].

The research presented in this paper follows a design science research approach [20]–[22]. The reason for choosing such an approach is the proximity of the research question to the practical challenge of creating useful process models and the objective to deliver design artifacts that are valuable for the application domain. We have collected extensive real world data from different companies for experiments and test purposes to develop a mining algorithm that is able to reconstruct the logical ordering of events.

The following section includes a brief review of related work to provide an overview of the positioning of the presented work in the context of already published research. Section three provides an example of a simple business process instance that is used for illustrating the differences between temporal and logical ordering of events in section four and five. Section six describes the evaluation efforts that were integrated into the research work to demonstrate the rigor of achieved results. The article closes with a brief summary and outlook in the last section.

2 Related Work

Process mining is a research area that emerged in the context of software engineering [23], [24]. It was first applied to workflow management in the late 1990s [25]–[27] and has matured significantly in recent years.

It would be out of the scope of this paper to provide an all-embracing overview on process mining. The Process Mining Manifesto [28] provides a comprehensive overview of contemporary challenges in process mining, and an overview of basic and advanced concepts on process mining can be found in [2].

Four dominant application domains for process mining are process discovery, enhancement, conformance [29] and compliance checking [8]. Process discovery, conformance and compliance checking are areas that are especially relevant for audit purposes. Process mining has already been successfully applied in the context of internal audits [30]–[33] and financial audits [34]. But we are not aware of research approaches that investigate the logical structure of events for the reconstruction of the control flow in the context of process mining. Approaches for organizational mining and social network analysis [35] focus on the resources that interact in a business process and exploit data referring to the relationship between process participants and activities to create models that illustrate organizational structures and social networks. Organizational mining also uses information from the event log other than the temporal ordering of events for dis-covering models but with the objective to discover the interaction of process participants. It therefore differs from the approach used in this paper that aims to reconstruct the control flow based on the logical structuring of events.

We use the Fuzzy Miner algorithm [7] for illustrating the traditional reconstruction of the control flow that bases on the temporal sequence of recorded events. The Fuzzy Miner is an advanced heuristic mining algorithm that is able to create less complex process models by abstracting from infrequent behavior and events. It has been implemented in academic [36] and commercial software tools [37].

The reconstruction of control flows based on the logical sequence of events is demonstrated by using the Financial Process Mining (FPM) algorithm that was developed specifically for the context of financial audits [18]. Financial accounts and journal entries are key concepts for process audits [38]. Each execution of a financially relevant activity creates an accounting journal entry that is recorded in an ERP system.

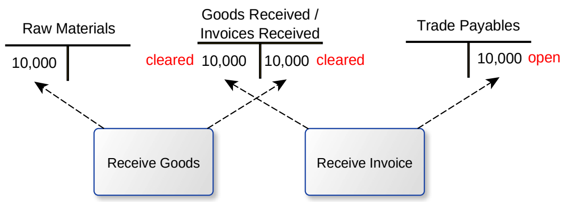


Figure 1 Accounting Structure

A journal entry consists of at least two journal entry items, one on the debit and one on the credit side of a financial account. Open items on an account are cleared by items belonging to other journal entries. These are created by the execution of activities that belong to the same process instance. This structure is illustrated in Figure 1. It shows the financial accounts with the corresponding journal entry items and the activities that created them. The analysis of the relationships between activities, journal entries and journal entry items allows the reconstruction of the logical control flow which will be discussed in detail in the following sections.

3 Process Example

We start with an example of a simple process instance for demonstrating the differences between temporal and logical ordering. Table 1 provides the event log for this example that was derived from a company operating in the manufacturing industry.

Table 1 Example event log

Event ID	Timestamp	Activity
0050155443	2010/01/02	Post Received Goods
0050155250	2010/02/08	Post Received Goods
0015975223	2010/02/17	Post Received Invoice
0015975224	2010/02/18	Post Received Invoice
0015975221	2010/02/19	Post Received Invoice
0095348327	2010/02/20	Clear Postings
0095348517	2010/02/21	Clear Postings
0050157332	2010/08/16	Post Received Goods
0015980342	2010/09/03	Post Received Invoice
0012490379	2010/09/04	Payment
0007904673	2010/09/05	Post with Clearing
0095359370	2010/09/07	Clear Postings

Process mining algorithms usually produce process models. A business process is a set of connected activities that in combination realize a specific business goal [39]. A business process model is an abstraction of a business process and consists of a set of activity models and execution constraints between them [40]. A single execution of a business process is called process instance. Process models represent the behavior of a set of process instances that belong to the same business process. A model representing a single process instance is called process instance model. Process models and process instance models generally include every activity only once. The represented activity models in process instance and process models are already abstractions of a set of executed activities. A process instance graph [41] resides on the lowest level of abstraction. Each event is represented as a single activity in the model. The differences between the abstraction levels of process models, instance models and instance graphs are important for the interpretation of the used example and mining algorithm outputs in the subsequent sections. We will only refer to the instance graph and process instance level in this article for ease of illustration. But it is important to note that the same results are also valid for the process model level.

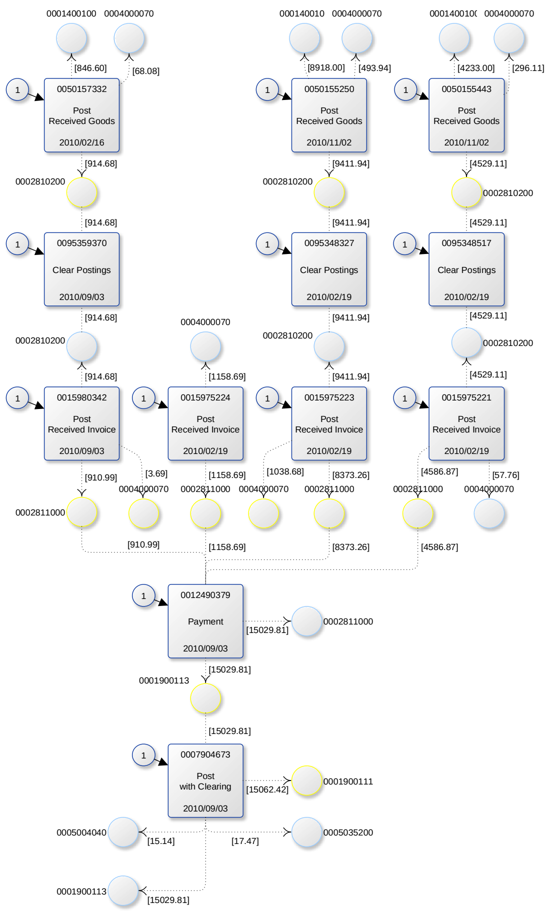


Figure 2 Example process instance

Table 1 includes the event log of a single process instance. Figure 2 shows an instance graph of the used example. The picture shows a Colored Petri Net (CPN) [42]. It provides information on the activities that were executed and involved financial accounts. The transitions (rectangles) represent activities that were executed in the process. The CPN includes two different types of places (circles). The yellow and blue colored places represent financial accounts. The dotted arrows leading from a transition to an account place denote that the corresponding activity posted a journal entry item on the connected account. The values of the posted items are included as inscriptions for each arrow. The dotted edges (test arcs) without arrows denote that an entry item was cleared on the respective account by the connected activity. The value of the cleared item is also displayed as an inscription for the corresponding edge. Each transition is further connected to a control place that carries a token indicating how often the activity was executed in the process instance.

The explicit modeling of these places creates life CPNs that adequately represent the observed behavior of the process instance. It ensures that each activity in the net can only fire once as recorded in the event log.

The model represents an instance of a purchase process. It shows that three different events occurred for the recording of received goods. The receipt of goods recorded by the activity *Post Received Goods* with the event ID *0050157332* for example led to journal entry item postings on the accounts *0001400100*, *0004000070* and *0002810200*. An invoice was received for each obtained good. An additional invoice processed by the activity with event ID *0015975224* was received with no corresponding recording of received goods. The open items on the related accounts were cleared by using the dedicated activity *Clear Postings*. All received invoices were subsequently cleared by the same payment. Intermediate postings were finally shifted to the final accounts by using the activity *Post with Clearing*.

The diagram provides a detailed overview of the structure of the process instance and illustrates information on financial accounts as well as clearing and posting relationships between account places and transitions that are relevant to reconstruct the logical ordering of events.

4 Temporal Sequence

If we use the Fuzzy Miner to discover a model based on the event log in Table 1 we obtain a model as displayed in Figure 3. The mining algorithm produces an instance model. It includes every activity only once (contrary to the instance graph in Figure 2) and therefore provides a higher level of abstraction than the model represented in Figure 2. The shown figure is not a process model because it only illustrates the behavior of a single process instance.

The mined model shows a different structure than expected when compared to the graph in Figure 2. Figure 2 actually shows four different branches that represent subprocesses of received goods and invoices that were all paid by the same payment run. The invoice in each branch was received after the receipt of goods was recorded. We would therefore expect a process model that shows the corresponding sequence of *Post Received Goods* → *Post Received Invoice* → *Clear Postings* → *Payment* → *Post with Clearing*. The model in Figure 3 instead shows several short loops indicating that the execution of an activity was followed by the execution of the same activity.

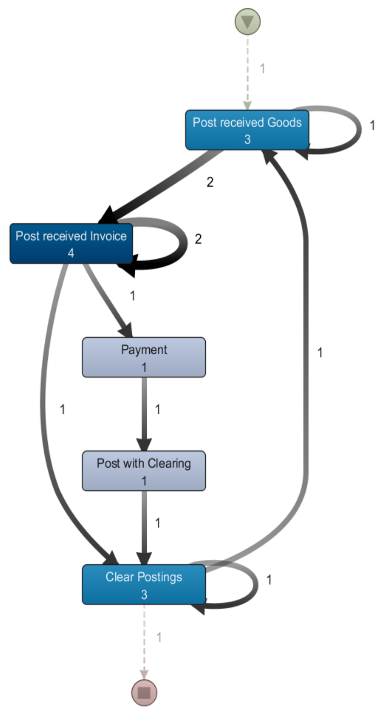


Figure 3 Discovered Fuzzy Miner process instance model

We can also observe a back loop from the *Clear Postings* activity to *Post Received Goods*. The reason for the difference becomes obvious when comparing Figure 4. It shows the temporal dependencies between the different events based on the recorded timestamps.

The Fuzzy Miner algorithm reconstructs the control flow according to the temporal sequence of events. The event *Post Received Goods* with the event ID *0050155443* was the first event that occurred at the *2010/01/02*.

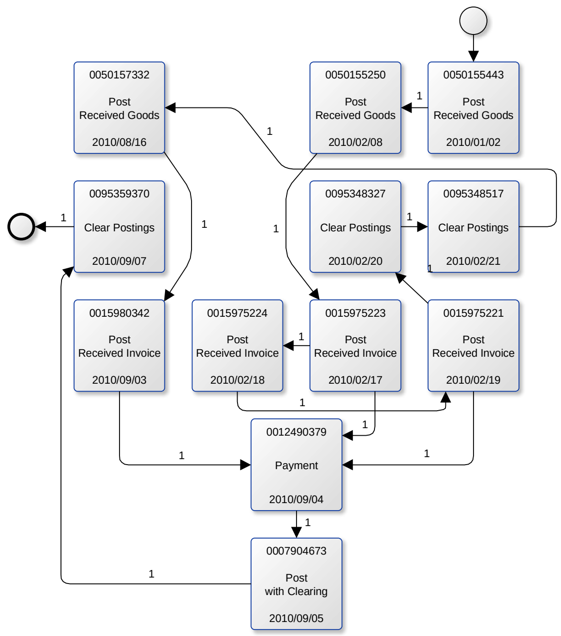


Figure 4 Temporal sequence

This event was followed by *Post Received Goods* with the event ID *0050155250* at the *2010/02/08*. The algorithm interprets this temporal dependency for reconstructing the control flow from event *0050155443 → 0050155250*. This approach produces a process model that adequately illustrates the temporal ordering of events. But it depends on the application scenario if this information is indeed useful for the intended purpose.

5 Logical Sequence

5.1 Defining Logical Dependencies

In financial audits it is important to get an understanding of the control flow of processes and the relation of business processes to the financial statements [38], [43]. The auditor needs to get an understanding what the different processes do and how they interact with the financial accounts. The first step in a process audit is to understand the process. Information on the logical structure is in this context more useful than the temporal sequence of events. The key requirement is to create process models that are purpose-oriented and easy to interpret.

We satisfy this requirement by modeling the logical control flow. It can be reconstructed by analyzing which journal entry item was cleared by another activity. An open item can only be cleared if it has been posted before the clearing item. It is therefore possible to derive the causal relationship between events by analyzing which activities cleared items created by other activities.

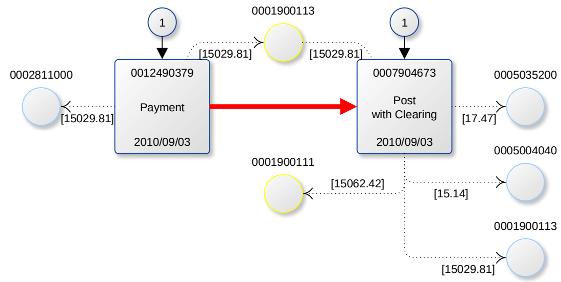


Figure 5 Logical dependency

Figure 5 provides an illustration of this approach. The activity Payment posted an item with the value of *15,029.81* on the account *0001900113*. This item was cleared by the activity Post with Clearing. The logical sequence for these activities can therefore be derived as *Payment → Post with Clearing*.

This procedure can be applied to all activities in the event log for determining the complete logical control flow of the process instance. It is further necessary to identify end and start nodes for creating a complete process model. Start nodes can be determined by identifying activities that do not have any incoming control arcs and end nodes do not have any outgoing control arcs.

5.2 Clearing Deadlocks

A problem arises with activities that do not post any items but only clear items posted by other activities. This constellation, which we call a clearing dead-lock, is illustrated in Figure 6.

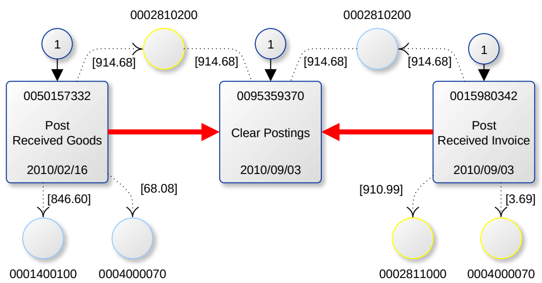


Figure 6 Clearing deadlock

Due to the fact that the *Clear Postings* activity did not create any other posted item the logical control flow ends at this activity and it would be defined as an end node. This is a problem. The clearing activity is actually not an end node because the process does not end before the activity *Post with Clearing*.

The results from our evaluations show that this constellation is very common at least for data extracted from SAP systems. It cannot be neglected because the resulting models would be of little value for the user. The additional end nodes would confuse the process model and imply invalid information about the process structure.

This outcome can be prevented by using graph transformations [44]. The idea is to identify clearing deadlocks and to apply graph transformations for removing them. Clearing activities first get identified by selecting all activities that did not post any items but only cleared items from other activities. Clearing activities commonly clear items from two or more other activities. At least one of these activities must have posted an additional item that was cleared by another activity different from the clearing activity for a deadlock to appear. Otherwise the clearing activity would represent a valid end node and such a constellation would not be considered a clearing deadlock.

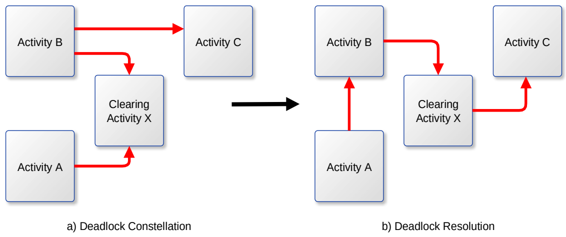


Figure 7 Deadlock resolution

Figure 7a) provides an illustration of a typical constellation. Items from activities A and B were cleared by clearing activity X. One or more additional items from activity B were also cleared by activity C. Because activity A has no further outgoing control arcs it is reasonable to assume that it is logically ordered before activity B and that the process continued with activity C after A, B and X took place. The related sub-graph can be substituted by an amended sub-graph as illustrated in Figure 7b).

The logical sequence can be rearranged unambiguously if only one of the posting activities has posted an item that was cleared by another activity. It is not possible to define a definite logical sequence locally by observing the direct neighbor activities if more than one of the posting activities has posted items that were cleared by other activities. Such constellations require non-local searches in the graph. The evaluation results showed that such constellations rarely occur and can be neglected for applications in practice.

Although the presented transformation improves the usefulness of the produced models it violates an important requirement from the application domain that relates to the preservation of the audit trail and postulates that mining algorithms should not alter the original data or provide information that is not reflected in the recorded event log [11]. The transformation of the modeled logical sequence is indeed an alteration of the original sequence. But negative impacts should be minimal. The presented approach is intended to be applied for the discovery of process models to provide models that allow auditors to easily understand relevant processes. The amendment of the logical control flow for resolving clearing deadlocks is not critical for this purpose but it might not be suitable for detecting conformance or compliance violations that require completely unchanged process models.

5.3 Logically Sequenced Process Models

The original FPM algorithm [18] is only able to reconstruct simple graphical representations of process instances. We used an amended algorithm that produces CPN and is able to create models on process instance and process level [11].

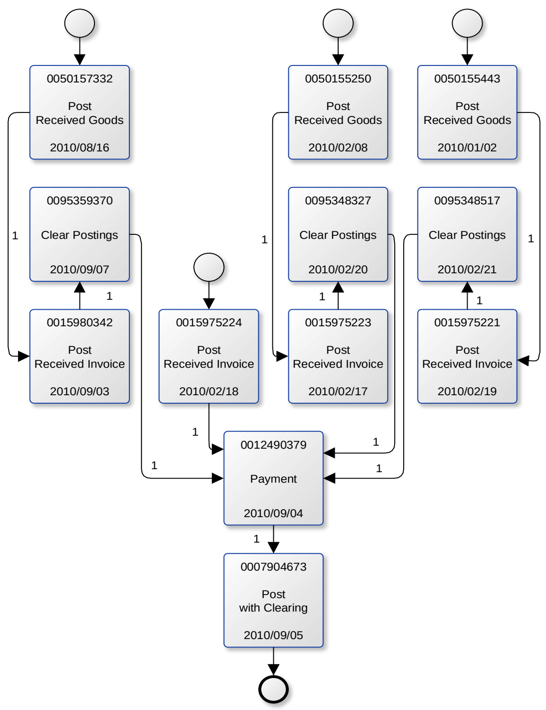


Figure 8 Logical sequence

The formerly explained procedures for (1) defining causal dependencies, (2) removing clearing deadlocks and (3) defining start and end nodes were implemented in a software prototype.

Figure 8 shows the effect of the logical sequencing. The positioning of the activities has been kept constant compared to Figure 2 and 4. The first, third and fourth branch show the logical sequence *Post Received Goods* → *Post Received Invoice* → *Clear Postings*. The second branch only consists of the sequence *Post Received Invoice*. All branches follow the subsequent sequence *Payment* → *Post with Clearing*. The illustrated instance therefore contains the logical sequences:

α: Post *Received Goods* → *Post Received Invoice* → *Clear Postings* → *Payment* → *Post with Clearing*

β: *Post Received Invoice* → *Payment* → *Post with Clearing*

Figure 9 shows the model for the example instance produced by the FPM algorithm. It only exhibits the same logical sequences α and β that also characterize the control flow in the model displayed in Figure 8.

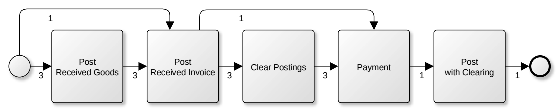


Figure 9 Discovered FPM process instance model[48]

The provided model in Figure 9 that was mined using the logical sequence of events is less complex and provides more useful information for the purpose of understanding the structure of a process than the model derived from using the traditional approach of determining the control flow by analyzing the temporal sequence that is illustrated in Figure 3.

[48] Process models produced by the FPN usually include information on the data perspective by modeling account places and their connections to the activities similar to the graph presented in Figure 2. We have omitted this information and only present a simplified dependency graph that can be easily compared with the output of the Fuzzy Miner illustrated in Figure 3.

6 Evaluation

Evaluation is a significant part of design science research [22], [45]. We have chosen an experimental setup for evaluating the mining results for the described mining approach. A laboratory experiment is a suitable evaluation method for artificial ex post evaluations [46] and was therefore chosen for the presented research work. We extracted event logs from productive SAP systems from three different companies operating in the manufacturing, media and retail industries.

An overview of the used data sets is provided in Table 2. The event logs served as input for the Fuzzy Miner and FPM algorithm. We evaluated the proper functioning of the implemented FPM and checked if the mining results fitted the expected outcome. The adjusted FPM should create sound [47] CPNs. Produced CPNs were tested for soundness by using the academic software Renew [48]. The graph editor yEd [49] provides powerful automatic layout functionality and was used to graphically represent mined models. The produced models were inspected and compared with the models produced by the Fuzzy Miner. The FPM models generally showed a lower complexity due to fewer loops in the models compared to the Fuzzy Miner models and complied with the expected mining outcomes.

Table 2 Overview of evaluation data sets

Data Set	Industry	Process Instances	Process Models
1	Manufacturing	1,035,805	841
2	Media	18,975	516
3	Retail	40,634	307

The evaluation revealed that unexpected, unsound process models were created under certain circumstances. This occurred when an activity created a posted item that was cleared by a subsequent activity carrying the same activity label. This constellation leads to a short loop and deadlock in the model. But it only occurred in very complex process models and very rarely. Research for addressing this problem is currently in progress. A solution could be the prevention of aggregating activities if this would lead to a short loop and to allow subsequent activities in the model that carry the same activity label.

The aim of using the logical structure of event logs for process mining is the provision of less complex and more informative process models according to the requirements of the application domain. Several metrics exist for measuring the complexity of process models [16]. The metric *structural appropriateness* measures the complexity of a process model by counting the number of included tasks [50]. *Structural precision* and *structural recall* measure the amount of causality relations that a mined model has in common with a reference model. The metrics *duplicates precision* and *duplicates recall* measure how many duplicate tasks a mined model has in common with a reference model [14]. The metrics *structural precision*, *structural recall, duplicates precision*, and *duplicates recall* are not suitable for measuring the complexity in our experimental setup because they require a reference model for comparison. The metric *structural appropriateness* only considers the number of modeled activities. Using the logical structure of event logs for process mining has no effect on the number of represented activities because only the sequence of activities is different. We therefore used the number of arcs as a criterion for measuring the complexity of a process model. The vast majority of mined models from our data sets represent trivial process instances that consist of only one activity [34]. Trivial instance models do not differ in complexity if the temporal or logical structure is used because the sequence is always the same if just one activity is involved. We therefore focused on nontrivial models that consist of more than four activities assuming that the complexity reduction is higher for more complex models. A quantitative evaluation for measuring the complexity reduction was carried out by using a specifically designed software artifact. It produced instance models using the FPM algorithm with the traditional temporal structuring and the alternative logical structuring as presented in this paper. Using the same algorithm ensured comparability of the results.

Table 3 Mean values of the number of arcs included in mined instance models

Data Set	Nontrivial Models			Complex Models		
	Temporal	Logical	%	Temporal	Logical	%
1	9.67	9.55	1	24.90	16.33	34
2	12.25	11.26	8	23.17	20.19	13
3	12.23	10.70	12	32.90	21.70	34

Table 3 summarizes the evaluation results for the used data sets.[49] It shows the mean values of the number of arcs that are included in the mined instance models. The table is split into two parts. The left columns show the mean values as a representative for the model complexity for nontrivial models that contain five or more activities. The overall complexity reduction is modest ranging from 1 to 12 percent. The right columns show the results for complex instance models that contain eight or more activities. The reduction is significantly higher for more complex models ranging from 13 to 34 percent. The complexity reduction for the most complex models was 51 percent for data set 1, 36 percent for data set 2 and 55 percent for data set 3.

It remains to discuss if the resulting models are more informative compared to those using a temporal ordering. A key requirement for process audits in the context of financial audits is the understanding of the structure of a process and its interaction with the financial accounts. Figure 3 shows the temporal sequence of recorded activities. The model is only of limited use because the intertwined temporal sequence of activities that belong to different subprocesses as indicated in Figure 4 leads to loops that make it difficult to identify the main sequence of activities in the process. This is different for the model shown in Figure 9. It clearly shows the logical structure and facilitates the interpretation of the process. It will be evaluated in future research work if this argument based assessment is also supported by experts from the application domain.

7 Conclusion and Outlook

The value of a process model depends on its ability to satisfy the requirements that are relevant to the application domain. A fundamental challenge in process mining is to create process models that are useful for the user. Mined models are often too complex for the intended purpose. A common approach to produce more suitable process models is the application of complexity reduction methods. This includes the development of mining algorithms like the Fuzzy Miner that are able to abstract from infrequent behavior and events to provide more abstract and therefore less complex models. Other approaches use abstraction techniques like aggregation or reduction methods for reducing complexity.

[49] A random sample consisting of 100,000 instances was used for data set 1 due to computational constraints.

The research presented in this paper follows a different approach. The aim is to use information from the application domain to discover the control flow in process models by exploiting the logical structure of events rather than the temporal structure that is used in traditional approaches. The exploitation of the logical structure of events provides the opportunity to mine process models that fit the requirements from the application domain better than traditional approaches and that produce less complex models. We have illustrated how information on the structure of event data can be used by referring to the application domain of financial audits. Information on the structure of accounting data can be used to reconstruct the logical control flow in process models that are useful for financial audits. This application domain should only be seen as an example of how such additional information from the event data can be used to improve the produced process models. A prerequisite for using domain specific information is its availability. The domain knowledge has to be explicit and formalized. Ontologies are known as a suitable representation form for such purposes. The development of an ontology for formalizing key concepts in process audits in the context of financial audits will be addressed in future research. It cannot be guaranteed that similar results can also be achieved for different application scenarios. But our research shows a new route that can be chosen for innovative improvements of mined process models.

The presented methods have been evaluated by using extensive data for testing the implemented methods and for inspecting the achieved outcomes. The data was derived from different companies. Although the data sets are extensive they only included event data from SAP systems. It can therefore not be concluded that the results are also valid for other ERP systems. But due to the fact that the chosen methods exploit the generic structure of accounting entries which need to be supported by all information systems used for accounting it is very likely that they are also applicable for other systems. Our evaluation proofed that the designed methods work correctly. But we did not gain information if they will also be accepted and useful in real world organizational settings. Additional research like field experiments will be conducted in future research to address this aspect.[50]

8 References

[1] H. Krcmar, Informationsmanagement. Berlin; Heidelberg: Springer, 2010.

[2] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st Edition. Berlin Heidelberg: Springer, 2011.

[3] H. Chen, R. H. L. Chiang, and V. C. Storey, “Business Intelligence and Analytics: From Big Data to Big Impact,” MIS Q., vol. 36, no. 4, pp. 1165–1188, Dec. 2012.

[4] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Han-nick, W. Hide, D. P. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Twigger, O. White, and S. Yon Rhee, “Big data: The future of biocuration,” Nature, vol. 455, no. 7209, pp. 47–50, Sep. 2008.

[50] The research results presented in this paper were developed in the research projects VAW (grant number 01IS10041) and EMOTEC (grant number 01FL10023) sponsored by the German Federal Ministry of Education and Research. The authors are responsible for the content of this publication.

[5] C. Lynch, “Big data: How do your data grow?,” Nature, vol. 455, no. 7209, pp. 28–29, Sep. 2008.

[6] W. M. P. van der Aalst, “Using Process Mining to Bridge the Gap Between BI and BPM,” Computer, vol. 44, no. 12, pp. 77–80, 2011.

[7] C. Günther and W. van der Aalst, “Fuzzy mining–adaptive process simplification based on multi-perspective metrics,” Bus. Process Manag., pp. 328–343, 2007.

[8] J. Becker, P. Delfmann, M. Eggert, and S. Schwittay, “Generalizability and Applicability of Model- Based Business Process Compliance-Checking Approaches – A State-of-the-Art Analysis and Research Roadmap,” BuR - Bus. Res., vol. 5, no. 2, pp. 221–247, Nov. 2012.

[9] W. M. van der Aalst and A. K. A. de Medeiros, “Pro-cess mining and security: Detecting anomalous process executions and checking process conformance,” Electron. Notes Theor. Comput. Sci., vol. 121, pp. 3–21, 2005.

[10] M. Reichert, “Visualizing Large Business Process Models: Challenges, Techniques, Applications,” in 1st Int’l Workshop on Theory and Applications of Process Visualization, Tallin, 2012.

[11] M. Werner, M. Schultz, N. Müller-Wickop, N. Gehrke, and M. Nüttgens, “Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits,” in Proceedings of 33rd International Conference on Information Systems, Orlando, 2012.

[12] W. M. P. van der Aalst, “What makes a good process model?,” Softw. Syst. Model., vol. 11, no. 4, pp. 557–569, 2012.

[13] A. Weijters, W. M. P. van der Aalst, and A. K. A. de Medeiros, “Process mining with the heuristics miner-algorithm,” Tech. Univ. Eindh. Tech Rep WP, vol. 166, 2006.

[14] A. K. A. de Medeiros, “Genetic Process Mining,” Eindhoven University of Technology, Eindhoven, 2006.

[15] M. B. Romney and P. J. Steinbart, Accounting Information Systems, 11th Revised edition (REV). Prentice Hall, 2008.

[16] A. Rozinat, A. K. A. de Medeiros, C. W. Günther, A. Weijters, and W. M. van der Aalst, “The Need for a Process Mining Evaluation Framework in Research and Practice,” in Business Process Management Workshops, 2008, pp. 84–89.

[17] M. Werner, N. Gehrke, and M. Nüttgens, “Business Process Mining and Reconstruction for Financial Audits,” in Hawaii International Conference on System Sciences, Maui, 2012, pp. 5350–5359.

[18] N. Gehrke and N. Müller-Wickop, “Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems,” in Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010.

[19] N. Gehrke and N. Müller-Wickop, “Rekonstruktion von Geschäftsprozessen im Finanzwesen mit Financial Process Mining,” in Lecture Notes in Informatics, Proceedings der Jahrestagung Informatik, Leipzig, 2010.

[20] A. R. Hevner, S. T. March, J. Park, and S. Ram, "De-sign Science in Information Systems Research," MIS Q., vol. 28, no. 1, pp. 75–105, Mar. 2004.

[21] S. T. March and G. F. Smith, "Design and natural science research on information technology," Decis Support Syst, vol. 15, no. 4, pp. 251–266, 1995.

[22] H. Österle, J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, and E. J. Sinz, "Memorandum on design-oriented information systems research," Eur. J. Inf. Syst., vol. 20, no. 1, pp. 7–10, 2010.

[23] J. E. Cook and A. L. Wolf, "Discovering models of software processes from event-based data," ACM Trans Softw Eng Methodol, vol. 7, no. 3, pp. 215–249, 1998.

[24] J. E. Cook and A. L. Wolf, "Software process validation: quantitatively measuring the correspondence of a process to a model," ACM Trans Softw Eng Methodol, vol. 8, no. 2, pp. 147–176, 1999.

[25] R. Agrawal, D. Gunopulos, and F. Leymann, "Mining Process Models from Workflow Logs," in Proc. Sixth Int'l Conf. Extending Database Technology, 1998, pp. 469–483.

[26] Herbst and Karagiannis, "Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models," Int'l J Intell. Syst. Account. Finance Manag., vol. 9, no. 2, pp. 67–92, Jun. 2000.

[27] J. Herbst and D. Karagiannis, "Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models," in Proceedings 9th International Workshop on Database and Expert Systems Applications (DEXA'98), 1998, pp. 745–752.

[28] W. M. P. van der Aalst, A. Andriansyah, A. K. de Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. van den Brand, R. Brandtjen, and J. Buijs, "Process Mining Manifesto," in BPM 2011 Workshops Proceedings, 2012, pp. 169–194.

[29] W. M. P. van der Aalst, "Process Mining: Overview and Opportunities," ACM Trans. Manag. Inf. Syst., vol. 99, no. 99, pp. 1–16, Feb. 2012.

[30] M. Jans, J. M. van der Werf, N. Lybaert, and K. Vanhoof, "A business process mining application for internal transaction fraud mitigation," Expert Syst. Appl., vol. 38, no. 10, pp. 13351–13359, 2011.

[31] M. Jans, N. Lybaert, K. Vanhoof, and J. M. Van Der Werf, "Business Process Mining for Internal Fraud Risk Reduction: Results of a Case Study," 2008.

[32] M. Jans, M. Alles, and M. Vasarhelyi, "Process Mining of Event Logs in Internal Auditing: A Case Study," in 2nd International Symposium on Accounting Information Systems, 2011.

[33] M. Jans, M. Alles, and M. Vasarhelyi, "Process mining of event logs in auditing: opportunities and challenges," Working paper. Hasselt University. Belgium, 2010.

[34] M. Werner, N. Gehrke, and M. Nüttgens, "Towards Automated Analysis of Business Processes for Financial Audits," presented at the Wirtschaftsinformatik, Leipzig, 2013.

[35] M. Song and W. M. P. van der Aalst, "Towards comprehensive support for

organizational mining," Decis. Support Syst., vol. 46, no. 1, pp. 300–317, 2008.

[36] Process Mining Group, "ProM," 2013. [Online]. Available: http://www.processmining.org/prom/start. [Accessed: 08-Apr-2013].

[37] fluxicon, "Process Mining and Process Analysis - Fluxicon," 2013. [Online]. Available: www.fluxicon.com. [Accessed: 01-Feb-2013].

[38] N. Müller-Wickop, M. Schultz, and M. Peris, "To-wards Key Concepts for Process Audits – A Multi-Method Research Approach," in Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics, Utrecht, 2013.

[39] M. Reichert and B. Weber, Enabling flexibility in process-aware information systems challenges, methods, technologies. Berlin; New York: Springer, 2012.

[40] M. Weske, Business process management concepts, languages, architectures. Berlin; New York: Springer, 2012.

[41] B. F. van Dongen and W. M. P. van der Aalst, "Multi-phase process mining: Aggregating instance graphs into EPCs and Petri nets," in PNCWB 2005 workshop, 2005, pp. 35–58.

[42] K. Jensen and L. M. Kristensen, Coloured petri nets. Springer, 2009.

[43] M. Schultz, N. Müller-Wickop, and M. Nüttgens, "Key Information Requirements for Process Audits – an Expert Perspective.pdf," in Proceedings of the 5th International Workshop on Enterprise Modelling and Information Systems Architectures, Vienna, 2012.

[44] G. Rozenberg, Handbook of Graph Grammars and Computing by Graph Transformation, Illustrated edition., vol. Volume 1 Foundations, 3 vols. Singapore: World Scientific Pub Co, 1997.

[45] S. Gregor and A. R. Hevner, "Positioning and Present-ing Design Science Research for Maximum Impact," MIS Q., vol. 37, no. 2, pp. 337–355, 2013.

[46] J. Venable, J. Pries-Heje, and R. Baskerville, "A comprehensive framework for evaluation in design science re-search," Des. Sci. Res. Inf. Syst. Adv. Theory Pr., pp. 423–438, 2012.

[47] W. M. P. van der Aalst and C. Stahl, Modeling business processes : a petri net-oriented approach. Cambridge, Mass.: MIT Press, 2011.

[48] University of Hamburg, "Renew - The Reference Net Workshop," 2013. [Online]. Available: http://www.renew.de/. [Accessed: 03-May-2012].

[49] yWorks GmbH, "yEd - Graph Editor," 2013. [Online]. Available: http://www.yworks.com/de/products_yed_about.html. [Accessed: 12-Aug-2012].

[50] A. Rozinat and W. M. P. van der Aalst, "Conformance Checking of Processes Based on Monitoring Real Behavior," Inf. Syst., vol. 33, no. 1, pp. 64–95, Mar. 2008.

10.9 Towards Automated Analysis of Business Processes for Financial Audits

Number	**9**
Title	**Towards Automated Analysis of Business Processes for Financial Audits**
Appendix	10.9
Primary Related Chapters	6
Type	Conference Paper
Conference	11th International Conference on Wirtschaftsinformatik (WI 2013)
Reference	(Werner et al., 2013)
Acceptance Rate	26%[51)]
VHB JQ 2.1 Ranking	C (6.73)
WKWI Ranking	A
ERA 2010	C
CORE 2013	C
Review Procedure	Double Blinded
Number of Reviews	5
Authors	1. Michael Werner 2. Nick Gehrke 3. Markus Nüttgens
Dissertation Points	0.50
Authorship	
Overall	90%
Design	90%
Realization	90%
Writing	90%
Status	Published
Part of other Dissertations	No
Link	http://www.wi2013.de/proceedings/WI2013%20-%20Track%203%20-%20Werner.pdf

[51] Acceptance rate for the relevant track 21.6 %

Towards Automated Analysis of Business Processes for Financial Audits

MICHAEL WERNER[1], NICK GEHRKE[2], AND MARKUS NÜTTGENS[1]

[1] University of Hamburg, Germany
{michael.werner, markus.nuettgens}@wiso.uni-hamburg.de

[2] NORDAKADEMIE, Elmshorn, Germany
nick.gehrke@nordakademie.de

Abstract. Financial audits play a significant role in the economy by safeguarding the correctness of published financial information. Public auditors face the challenge to audit financial statements that are created by increasingly integrated and complex information systems. This paper addresses a specific problem in the auditing process. A major challenge in this process is the analysis and audit of business processes that produce financial entries. We illustrate results from applying business process mining techniques to extensive test and real life data and discuss gained insights from the application for the development of automated business process analysis methods in the context of financial audits.

Keywords: Process Mining, Financial Audits, Business Process Analysis

1 Introduction

Financial audits play a significant role in modern economies. Companies publish financial statements in order to inform relevant stakeholders. For preventing misinformation of the addressees financial statements are subject to audits that are mandated by law and specified in regulatory requirements. The audits are carried out by public auditors who follow specific audit approaches for planning and executing their audits. Audit standards require that auditors consider and test relevant business processes during the audit [1]. The requirement derives from the assumption that well controlled transaction processing will lead to valid entries on the balance sheet and profit and loss statements. When business transactions are carried out in a correct and controlled manner they will most likely lead to complete and accurate journal entries.

With increasing integration of the execution of business processes in information systems and the accompanied progress in automation of transaction processing it becomes more and more challenging to audit these processes. Contemporary audit approaches take into account the relevance of business processes, supporting information systems and internal control frameworks, but they basically rely on manual audit procedures to analyze and test them. The manual procedures primarily include interviews for obtaining information and manual test activities for evaluating relevant controls. With increasing integration of information systems for supporting and automating transaction processing audit activities like interviews and manual audit activities become inefficient or even ineffective due to the increasing complexity and the mere volume of processed transactions [2].

An alternative would be the application of automated analysis and audit procedures as a business intelligence tool that supports the auditor in the auditing process. [3] conceptually illustrate how process mining methods can be combined with automated application control testing methods for designing automated audit methods. A requisite for such a development are methods that allow an automated analysis of business processes. The analysis results can then be used for automated testing purposes.

When information systems are used to support or automate the transaction processing they also provide information that can be used for an automated analysis. By using process mining techniques [4] and specific mining algorithms for financially relevant business processes [5] executed process instances can be mined, reconstructed and analyzed.

In this paper we focus on the aspect of automated analysis. We apply an existing mining algorithm for financially relevant business processes to test data and real life data. The aim of this research is to evaluate which insights can be derived by analyzing the application of the implemented algorithm. We statistically analyze the mined business processes instances that are reconstructed from the available data to identify which further research and improvement is needed on the path towards automated analysis methods.

We start with an illustration of related work in section two, followed by a brief description of the applied research methodology in section three. The mined process instances are represented as Petri nets. The used representation, the chosen mining method and the experimental setup are explained in section four. Section five provides the results from analyzing the process instances that were mined from the test and real life data. A discussion of the gained results and an illustration of identified limitations followed by a brief summary and conclusion close the paper.

2 Related Work

Of particular interest for the research laid out in this paper are publications from the field of process mining. Research on process mining started in the late 1990s by [6] and has gained extensive attention in the last decades. Significant research work has been published by van der Aalst et al. leading to a comprehensive basic publication on process mining that covers major aspects of the research domain [4].

From a financial accounting and auditing perspective requirements are outlined in relevant audit standards. The major international standard setting body is the International Auditing and Assurance Standards Board (IAASB) which publishes the International Standards on Auditing (ISA). The ISA 315 (Revised) “Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment“ outlines the requirement to consider business processes and related internal controls in order to assess the risk for material misstatement (ISA 315.18) [1].

The role of information systems for accounting is well researched but few authors address the role of information systems in the context of auditing. [7] describes techniques to audit enterprise resource planning (ERP) systems, but the exploitation of information that is available in information systems for the purpose of automated analyses is a relatively novel field of research as illustrated by [8].

Specific research on process mining for auditing purposes has gained increased attention over the last two to three years. [9] offers an overview of current limitations and future challenges of process mining in the context of audits whereas [10] illustrates opportunities of online auditing. [11–13] focus on fraud and outline possibilities of process mining for fraud detection and auditing thereby highlighting the potential of process mining as a new toolkit for internal audits. [5, 14] developed a mining algorithm that is able to exploit the structure of financial journal entries for the purpose of process mining in the context of financial audits. [15] further introduce automated audit methods for testing application controls in ERP systems. [2, 3] finally conceptually describe how process mining techniques for financially relevant business processes can be combined with methods for automated control testing.

For the purpose of the research work of this paper we implemented the mining algorithm introduced by [5, 14]. Their mining technique includes the extraction of financially relevant information of journal entry values that are relevant for the purpose of auditing.

An alternative approach is used by [16]. They provide an interesting case study about the examination of mined instances of a procurement process. Their approach differs from the research presented in this paper as they actually perform a deviation analysis of the mined process instances with a manually evaluated ideal process. They base their analysis on a predefined set of process instances. The aim of the research work illustrated in this paper does not focus on providing a case study for auditing mined business processes but intends to reveal general possibilities and limitations to discover and analyze business processes from event log data without further knowledge of the underlying processes in the context of financial audits. As illustrated in [3] the ultimate aim is to develop methods that show which processes are mirrored in the available logs and how they affect the financial statements.

For analyzing mined process instances these need to be modeled in a purposeful modeling language. [17] suggest using a BPMN representation for mined process instances in the context of financial audits. Although BPMN process models might be easier to interpret for end users we have chosen Petri nets as a modeling language for the research presented in this paper. On the one hand a broad variety of the aforementioned research work from the field of process mining relies on Petri nets as the choice of modeling language [18]. And referring to Petri nets opens up the opportunity to incorporate these already existing research results and techniques for the purpose of mining and analyzing. On the other hand Petri nets have a mathematical foundation and offer a formal graphical notation. These characteristics allow the development of sophisticated analysis methods. They therefore constitute the preferred modeling language for the research outlined in this paper. In the context of this paper we primarily refer to the publications of [19] and [20] concerning the theoretical foundation for the application of Petri nets.

3 Research Methodology

The research presented in this paper follows a design science approach [21–23]. A common critic in the academic arena refers to the perceived lack of rigor concerning design science oriented research. In order to address this aspect we have obtained extensive test and real life data for testing the designed artifacts. Actually the key aspect of this paper is to illustrate the results of evaluating already designed methods against this data. The illustrated work follows a research process as suggested by [23] consisting of the phases analysis, design, evaluation and diffusion. The requirements for an adequate representation and modeling of mined processes were investigated by considering specific, already existing literature [17] and by analyzing available test and real life data. The used mining methods were engineered [24] by assembling parts of already available methods and by developing new concepts where no adequate solutions were available yet.[52] The analysis results and requirements for further development constitute the primary outputs produced in the research process that is laid out in this paper. The engineered methods were implemented in a software prototype for evaluation purposes. We rigorously tested the software artifact with test and real life data in order to validate it against the relevant research questions addressed by the research work [25]. The content of this article discusses the results and insights that have been generated by applying the designed methods to this voluminous data.

4 Representation and Experimental Setup

[5] introduced a simple, deterministic and unsupervised mining algorithm that is suitable for extracting data from information systems and for reconstructing executed process instances. When using process mining in the context of financial audits it is necessary to mine information that is relevant from an audit perspective and to ensure that the received information precisely reflects the executed transactions. Financial transactions in ERP systems create journal entries when they are executed. The chosen algorithm exploits the open-item-accounting structure of journal entries that can be used to link transactions to a process instance.[53] Journal entries consist of an accounting document and at least two entry items that are posted as credits and debits. When open-item-accounting is enabled each cleared item has a reference to the accounting document that cleared it. The algorithm starts with an arbitrary journal entry and reconstructs the links between journal entries that cleared each other. It matches the events in the event log to cases that represent process instances.[54]

The original mining algorithm produced directed graphs representing the mined process instances. We extended the mining algorithm with a function for mapping the mined cases to Petri nets and implemented it in a software artifact. The software prototype was written in Java using the Java NetBeans IDE [26]. It provides functionality to export Petri net models in different data formats for visual representation. The open source software Renew [27] was used for verifying that the software artifact reconstructs reachable and therefore correct Petri nets. The yEd Graph Editor [28] was used for graphical representation and automatic layout of mined process instances.

Figure 1 displays a colored Petri net (CPN) model of a reconstructed process instance. The example shows an instance of a purchasing process. Executed transactions are modeled as

---

[52] The engineering of the applied methods is not part of this paper. Details are available in [5].

[53] The open-item-accounting is a fundamental concept of the double-entry bookkeeping which needs to be supported by every information system used for double-entry bookkeeping.

[54] Compared to other mining algorithms like the α-Algorithm [4] the implemented algorithm does not rely on the temporal ordering of events but on their logical structure.

rectangles (Petri net transitions). The journal entry items produced by executing the transactions are modeled as circles (Petri net places). The places are colored by the account number according to the account the item was posted to. Two different types of connections are possible between transactions and journal entry items. A dotted arrow (Petri net arc) means that a transaction has posted the connected journal entry item. A dotted line (Petri net test arc) illustrates that a journal entry item was cleared by the connected transaction. Arc inscriptions play a significant role. They denote the values that are associated to the connection between transactions and journal entry items. Each transition is accompanied by a start place containing a token colored with the original document number of the journal entry. They are connected to the corresponding transaction with a simple arrow (Petri net arc). This actually leads to an enabled CPN that mimics the behavior of the originally executed process instance.

The example in Figure 1 shows that a transaction for receiving goods (MB01) was processed by user 2. It led to entries on the raw material account (310000) and on the goods received / invoices received account (191100) with the amount of 17,874.76. The invoice for these received goods were processed (MR1M) and a payment run (F110) executed that cleared the items posted by the MR1M transactions. The FB1S transaction was executed for clearing the items that were created by MB01 and MR1M.

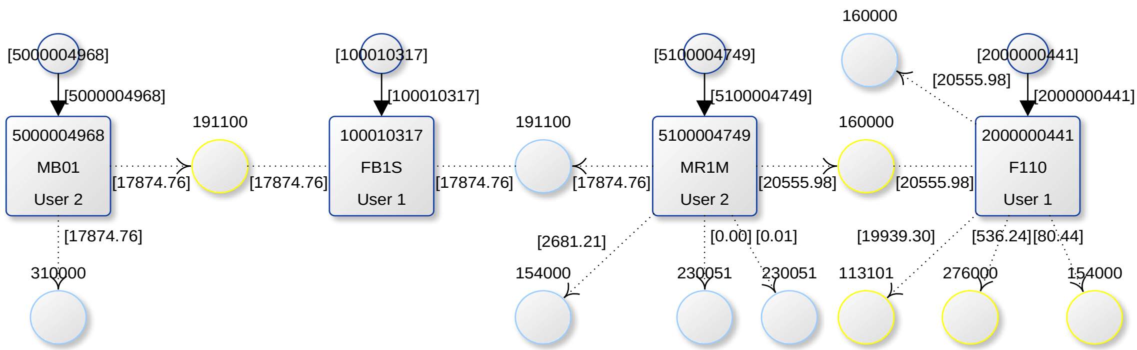


Fig. 1. Example of a Reconstructed Purchase Process Instance

For applying the mining algorithm the necessary data of the executed process instances was extracted from the available ERP systems. The setup of the experiment including conducted activities, involved software modules and input and output for each activity is illustrated in Figure 2. The relevant data was extracted by using a configurable extraction module. It retrieves the event log from the ERP system by extracting data from relevant database tables. The usage of a separate module provides the benefit that only the extraction component needs to be adjusted when data is extracted from different ERP systems. The mining algorithm operates independently from the underlying data structure of the individual ERP systems. The extraction module loads the extracted data into an event log database that can be accessed by the mining module. The mining module matches events in the log to cases, reconstructs executed instances and provides functionalities for analyzing them. It also produces output files in different formats (Extensible Graph Modeling Language (XGML) and Petri Net Markup Language (PNML)) that can be imported into subsequent software for verification (Renew) and graphical presentation (yEd Graph Editor) purposes. The reconstructed instances are modeled as Petri nets and stored as separate data objects in the mining module.

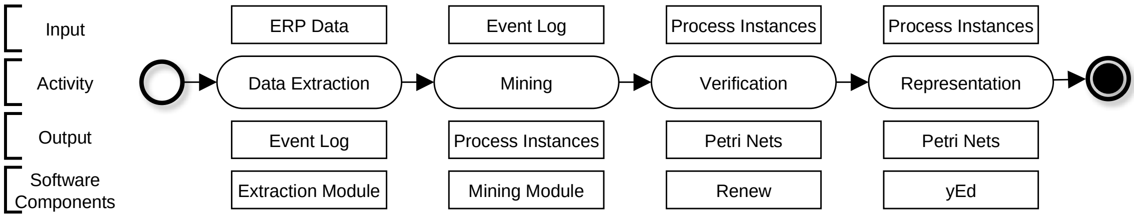


Fig. 2. Experimental Setup

We used three different data sets for analysis. The first set was extracted from the SAP IDES test system. The test system is available for universities participating in the SAP University Alliance Program [29]. Postings in ERP systems are stored as data entries with information relating to the whole posting (journal entry) and the single entries that were posted to different accounts (journal entry items). The data set contained 115,060 journal entries and 419,106 journal entry items. 81,171 process instances could be reconstructed by executing the implemented mining algorithm. The data set included all transaction data that was available in the test system covering a period of 17 years.

The second data set was extracted from a SAP system of a retail company. The set included the data of all executed transactions but only for a time period of one year. The volume of 92,487 journal entries and 222,901 related journal entry items that can be traced back to 40,130 process instances over just one year illustrates the high amount of transactions that are processed in real life environments.

This observation becomes even more evident for the third data set. It originated from a SAP system of a manufacturing company in the health sector. It contains 1,764,773 journal entries and 7,395,434 journal entry items. 1,035,805 process instances could be reconstructed using the mining algorithm. Table 1 provides an overview of the different data sets.

Table 1. Overview of Data Sets

Data Set	#1 SAP IDES	#2 Retail	#3 Manufacturing
Number of journal entries	115,060	92,487	1,764,773
Number of journal entry items	419,106	222,901	7,395,434
Number of process instances	81,171	40,130	1,035,805
Covered period	17 years	1 year	1 year

5 Mining Results Analysis

The mining and reconstruction of process instances from the available data sets provide the basis for analyzing the created Petri net models. The aim of this analysis is the identification of patterns that might help in further improvement of the mining algorithm, the gaining of insights concerning which further research is needed for developing automated analysis methods that can be applied in real life scenarios, and what kind of limitations for developing such methods might exist.

The following sub-sections illustrate results from statistical analyses of the mined process instances for all three used data sets. Due to place restrictions we limit the presentation of results to those aspects that we consider relevant for the aforementioned aim.

5.1 Distribution of Net Size

Figures 3, 4 and 5 show the distribution of the number of process instances over the number of net elements with logarithmic scaling on the x- and y-axes for the different data sets.[55] Net elements include transitions, places and arcs. The number of net elements gives an impression of the size and complexity of a process instance.

The charts illustrate that the distribution of the number of net elements over the number of instances exhibit the same pattern for all data sets. Only very few instances consist of very many net elements. The vast majority of instances consist of relatively few net elements. Table 2 provides an overview of specific characteristic values of the distributions.

The shown distributions by themselves do not allow drawing conclusions concerning the design of automated business process analysis procedures. But the observation that the majority of instances actually consist of relatively few elements and that this is the case for all evaluated data sets constitutes a useful insight in combination with the analysis results highlighted in the following sub-sections.

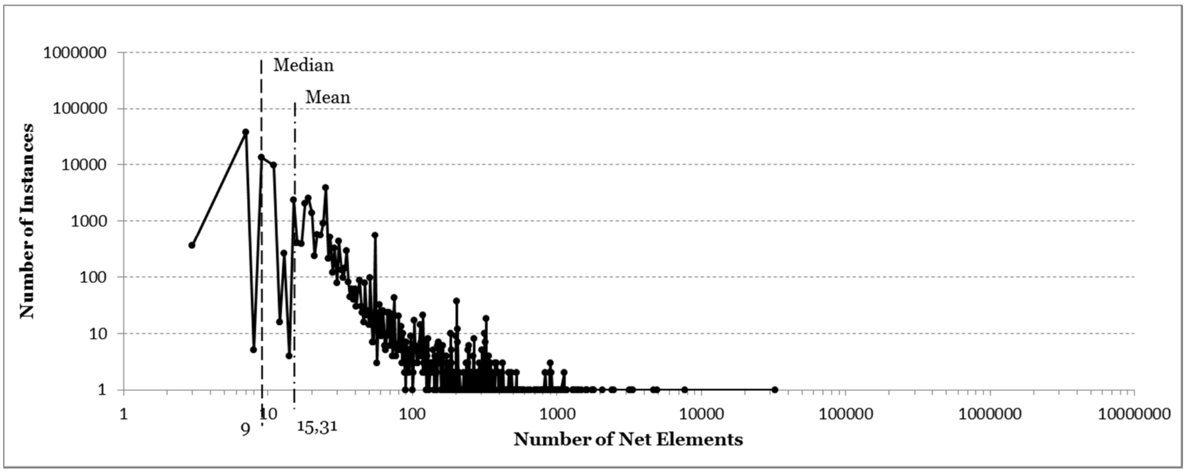


Fig. 3. Data Set #1 Distribution of Number of Net Elements over the Number of Instances

[55] The diagrams do not show values for instances with less than three net elements. Each net consists at least of one transaction represented by a transition in the model. Each transition is accompanied by a start place that is connected to the transition and that enables the transition to fire. Therefore the minimum number of net elements per net is three.

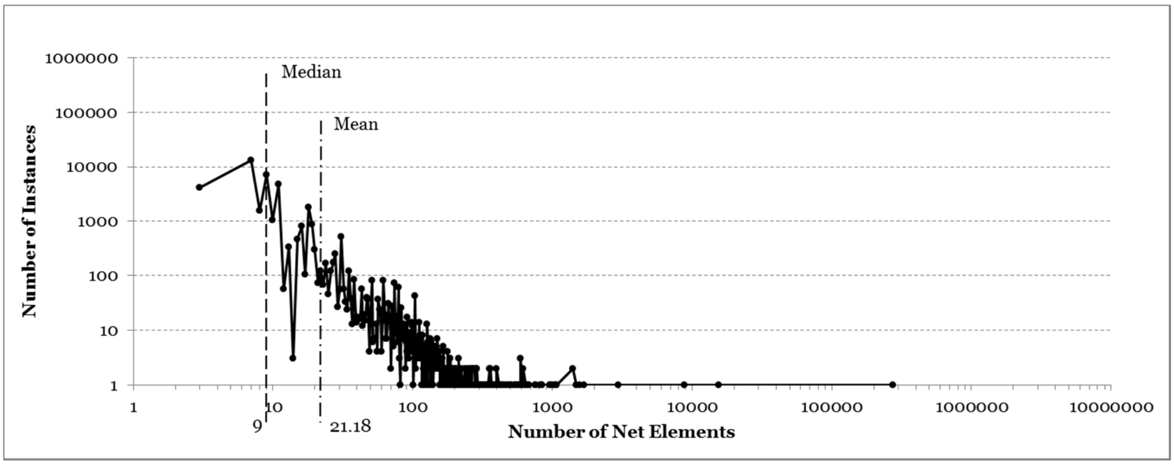


Fig. 4. Data Set #2 Distribution of Number of Net Elements over the Number of Instances

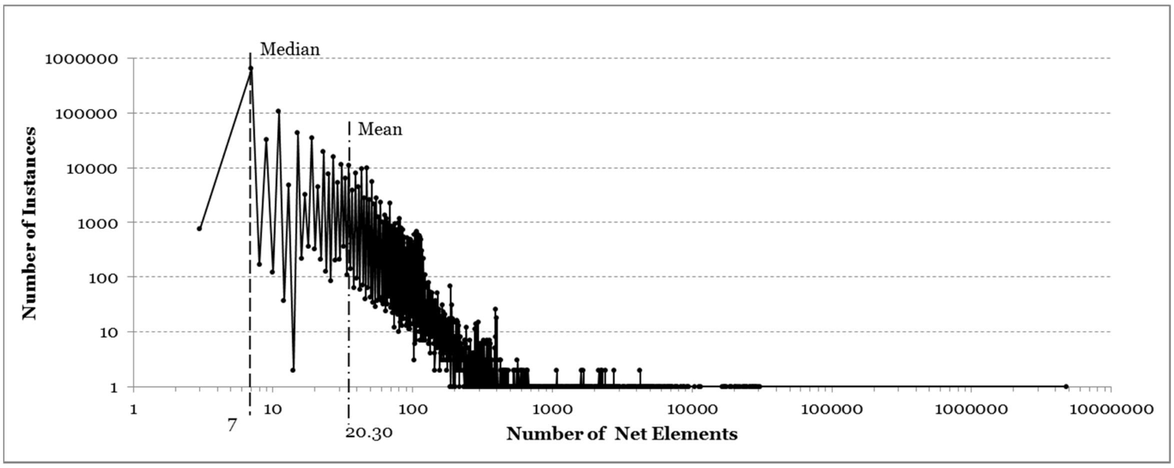


Fig. 5. Data Set #3 Distribution of Number of Net Elements over the Number of Instances

Table 2. Overview of Net Size Distribution Characteristics

Data Set	#1 SAP IDES	#2 Retail	#3 Manufacturing
Mean value of net elements per instance	15.31	21.28	20.30
Median value of net elements per instance	9	9	7
Maximum of net elements per instance	32,519	275,870	4,769,379
Standard deviation of number of net elements	127.83	1,380.41	4,688.30

5.2 Distribution of Transaction Code Combinations

Figures 6, 7 and 8 show the distribution of transaction code combinations over the number of instances. The y-axes follow a logarithmic scaling. Each number on the x-axis represents a transaction code combination. This means for example that the transaction code combination number 27 (FB1S, MB01, F110, FB05, MIRO) was executed in 1,069 process instances in data set three. These instances did not include any other transaction codes.

The distributions for all data sets show the same pattern. The majority of instances only contain very few different transaction code combinations. Taking into account the results from analyzing the distribution of net sizes from the previous section it is reasonable to assume that the majority of instances are very limited in size and reveal the same transaction code combinations.

A clustering of instances that exhibit the same size and the same transaction code combination could be a starting point for automatically analyzing a large amount of the mined instances. Each cluster could be reviewed for the value that it contributes to the financial statements and if the cluster constitutes a material process flow that needs to be further evaluated from a materiality perspective. Such an analysis would provide useful information to the auditor about how the processes in a company actually affect the financial statements.

The clustering of isomorphic graphs into clusters would also enable to search for application controls that affect the cluster under review. In combination with automated application control testing the isomorphic process instances in a cluster could be automatically audited.

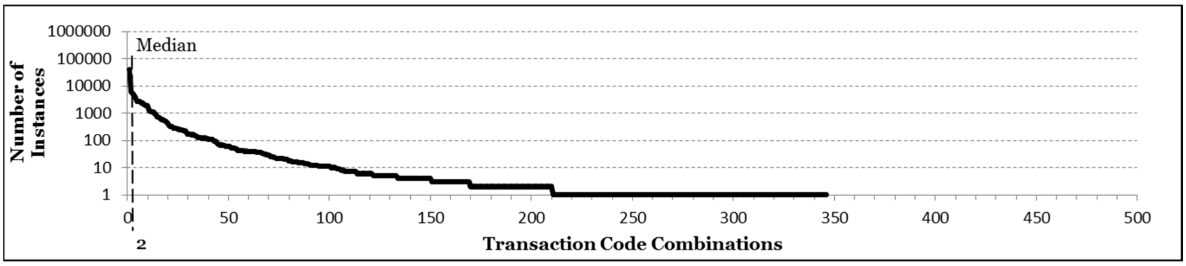


Fig. 6. Data Set #1 Distribution of the Number of Instances for Different Transaction Code Combinations

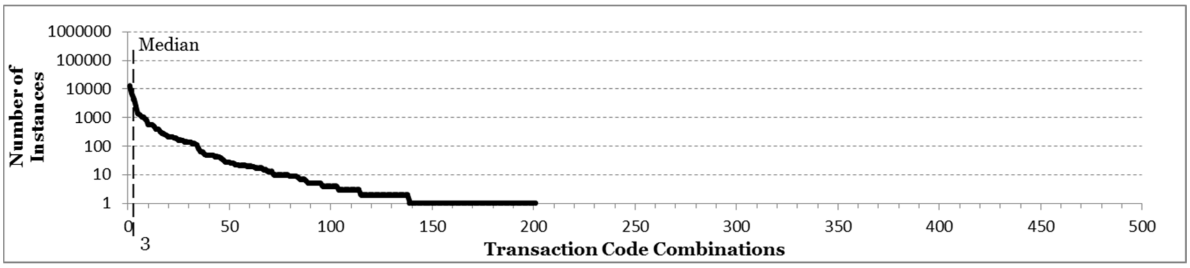


Fig. 7. Data Set #2 Distribution of the Number of Instances for Different Transaction Code Combinations

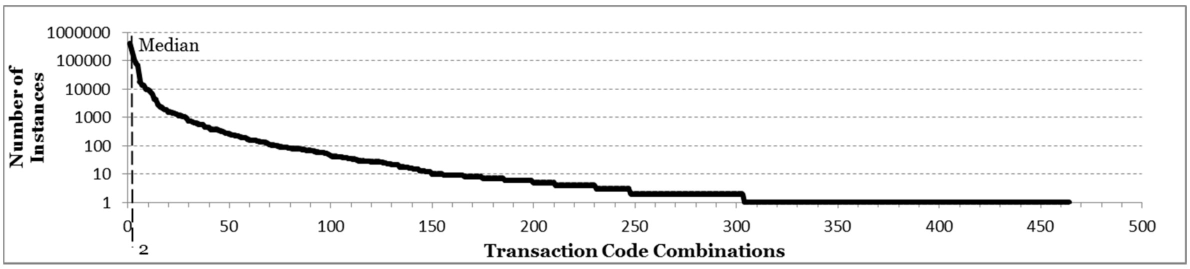


Fig. 8. Data Set #3 Distribution of the Number of Instances for Different Transaction Code Combinations

5.3 Distribution of Accounts

We suggested in the previous section that clustering could be a promising solution for analyzing similar and small instances. This leads to the question how large and complex instances should be handled. The instance in Figure 1 contains four transitions and consists of 38 net elements. An instance of this size and complexity can be evaluated by simple observation. This is not the case anymore for more complex instances. Figure 9 shows a process instance containing 3,057 transitions and consisting of 15,319 net elements. These kinds of instances cannot be evaluated without further consideration.

Figures 3 to 5 show that the number of net elements extending 100 net elements is relatively small in all the data sets. Only 691 instances in data set one consist of 100 or more net elements, 428 in data set two and 7,470 in data set three. Although the number of complex instances is relatively low they might represent a material amount of transactions that affect the financial statements and can therefore not be neglected. For analyzing and evaluating these instances it is necessary to reduce their complexity. The analysis of the accounts that are actually used within an instance provides a starting point for investigating complexity reduction possibilities. The distributions of the accounts over the number of process instances in Figures 10 to 12 show on how many accounts journal entry items were posted to in a single process instance. For example 222 instances in data set two used seven accounts to post items.[56]

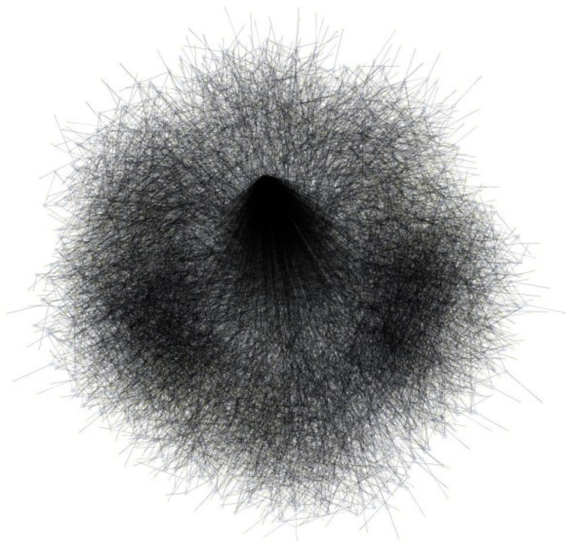


Fig.9. Complex Process Instance

The distributions for all three data sets again display the same pattern. Table 3 provides an overview of characteristic values of these distributions. The maximum number of used accounts is relatively low compared to the maximum net sizes. For interpreting the illustrated distributions and their characteristic values it is necessary to consider that only one instance in data set two uses the maximum number of 596 accounts and one in data set three the maximum of 399 accounts. All other instances do not use more than 63 accounts in data set two and no more accounts than 45 in data set three.

The observation of the distributions reveals that journal entry items are posted to relatively few accounts. This is reasonable because a specific process generally only uses a subset of the available set of accounts. The execution of a purchase process for example would likely lead to journal entries on expense accounts but not on sales accounts.

Based on the observation that the number of used accounts is relatively small it might be useful to aggregate journal entry items that were posted to the same accounts and thereby reducing the net size and complexity. Items are modeled as places in the Petri net and colored by the account number reflecting the financial account the item was posted on. The

[56] A logarithmic scaling is again used for the x- and y-axes.

places carrying the same color (account number) could be folded leading to Petri net models that contain significantly less net elements and therefore a reduced complexity. Further research would actually be needed for verifying if a sufficient complexity level can be reached by this approach.

Table 3. Overview of Account Distribution Characteristics

Data Set	#1 SAP IDES	#2 Retail	#3 Manufacturing
Mean value	2.95	2.30	2.33
Median value	2	2	2
Maximum value	36	596	399
Standard deviation	1.61	3.37	0.92

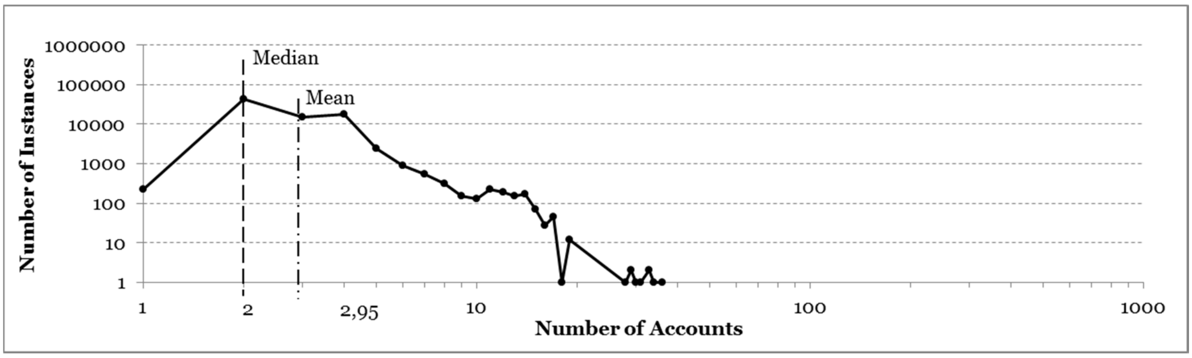


Fig. 10. Data Set #1 Distribution of the Number of Instances over Number of Accounts

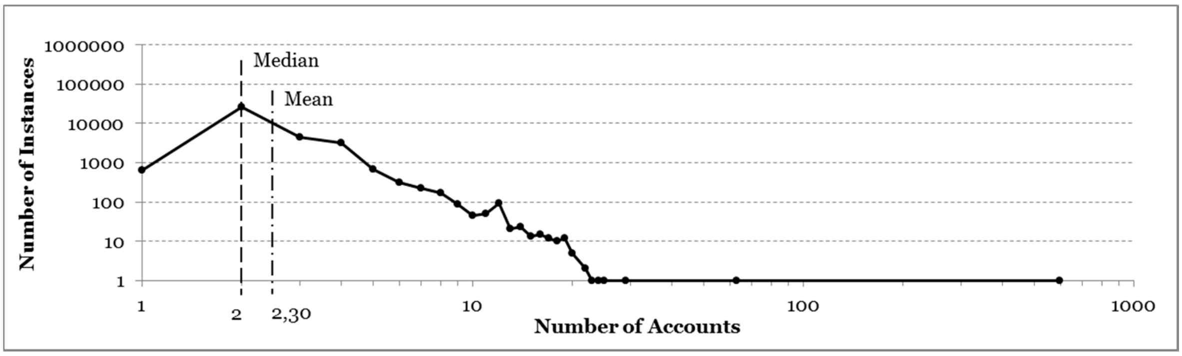


Fig. 11. Data Set #2 Distribution of the Number of Instances over Number of Accounts

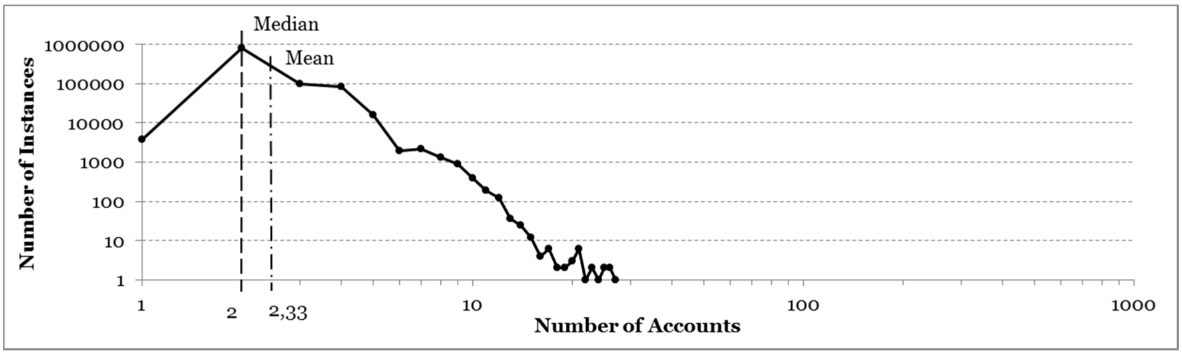


Fig. 12. Data Set #3 Distribution of the Number of Instances over Number of Accounts

6 Discussion and Limitations

The previous sections lay out insights that can be derived from analyzing results from the application of a mining algorithm. When considering these results it is necessary to keep in mind that they only provide starting points for further research and that no designed methods and their implementations do exist yet that could prove or disprove the postulated assumptions.

A further restriction is the number of used data sets. [16] relied on an event log representing 26,185 procurement process instances from a single company in the financial sector. We therefore consider the size of the data sets as being sufficient and we also did not limit our analysis to a specific type of business process but we actually only cover two industries – retail and manufacturing. We do not know if the identified characteristics also hold true for other industries.

As a result of the analysis of used accounts we proposed the folding of places in a Petri net carrying the same account number for complexity reduction purposes. From an audit perspective it is crucial that the received information exactly mirrors the transaction that really took place and therefore to maintain the audit trail [30]. Applying a folding approach would actually lead to a graph transformation of the mined process model and it would need to be ensured that the behavior of the net remains stable [17].

A further limitation relates to the execution of the programmed software and the time that is needed to calculate the mined instances. Dedicated research on the performance of the implemented software prototype is currently outstanding but our experience from analyzing the data sets presented in this paper lets us assume that the complete reconstruction of all process instances might be unrealistic in a real life scenario and that a sampling approach based on a materiality perspective might be needed.

7 Summary and Conclusion

Research on process mining has advanced and matured significantly over the past two decades and is now increasingly applied to specific application domains. The focus of the research presented in this paper lies on the domain of financial audits. The audit of business processes is a mandatory step in the auditing process that becomes increasingly challenging with the ongoing integration of information systems and automation of transaction processing. The application of process mining for supporting the auditor comprises a promising alternative to counter the growing complexity and amount of processed transactions that need to be evaluated during an audit.

We implemented a mining algorithm in a software artifact and evaluated it against voluminous test and real life data. Based on the results derived from the analysis of reconstructed process instances we gained insights that can be used for developing automated process analysis methods in the context of financial audits.

The majority of process instances is small and consists of only few net elements. The clustering of instances that exhibit the same size and the same transaction code combination could be a starting point for automatically analyzing a large amount of the mined instances. Each cluster could be reviewed for the value that it contributes to the financial statements.

Such an analysis would provide useful information to the auditor about how the processes in a company actually affect the financial statements.

Process instances containing more than a few processed transactions cannot be evaluated manually by simple observation. Although the overall number of complex process instances is very limited they cannot be neglected from an audit perspective. A single instance may already contain extremely high volumes of transactions that might be material. A starting point for reducing complexity could be the consideration of used accounts. Even for large and complex process instances the number of used accounts in a single process instance is relatively small. The folding of equally colored places in process instances could lead to significant complexity reduction especially for large process instances.

The research presented in this paper can be seen as a step towards the development of automated business process analysis methods. The accounting scandals of major companies over the last years illustrates that the audit industry is currently lacking adequate solutions for safeguarding the correctness of published financial statements. The usage of automated analysis and audit methods constitutes a necessary requirement to overcome the existing imbalance between automated processing on the companies' side and manual audit procedures on the auditors' side. The introduction of automated audit procedures has the potential to leverage this imbalance.

Limitations concerning the applicability of the implemented algorithm exist especially in regard to processing time and the amount of process instances that can be mined and analyzed. At this point of time it also remains unclear if the identified starting points can actually be transferred successfully into the design of automated business process analysis methods. But first results from consecutive research that bases on the results presented in this paper that will be published in forthcoming articles provide a positive indication.[57]

References

1. International Federation of Accountants: ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment, (2012).
2. Werner, M., Gehrke, N.: Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen. Forschung für die Wirtschaft. Shaker Verlag (2011).
3. Werner, M., Gehrke, N., Nüttgens, M.: Business Process Mining and Reconstruction for Financial Audits. Hawaii International Conference on System Sciences. pp. 5350–5359. Maui (2012).
4. Van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin, Heidelberg (2011).
5. Gehrke, N., Müller-Wickop, N.: Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems. Proceedings of the 16th Amer-icas Conference on Information Systems. Lima, Peru (2010).

[57] The presented results were developed in the research project Virtual Accounting Worlds. The project is sponsored by the German Federal Ministry of Education and Research (grant number 01IS10041). The authors are responsible for the content of this publication.

6. Agrawal, R., Gunopulos, D., Leymann, F.: Mining Process Models from Workflow Logs. Proc. Sixth Int'l Conf. Extending Database Technology. pp. 469–483 (1998).

7. Musaji, Y.F.: Integrated Auditing of ERP Systems. John Wiley & Sons (2002).

8. Alles, M., Jans, M., Vasarhelyi, M.: Process Mining: A New Research Methodology for AIS. CAAA Annual Conference 2011 (2011).

9. Jans, M.J.: Process Mining in Auditing: From Current Limitations to Future Challenges. In: Daniel, F., Barkaoui, K., and Dustdar, S. (edr.) Business Process Management Workshops. pp. 394–397. Springer, Berlin, Heidelberg (2012).

10. Aalst, W.M.P. van, Hee, K.M. van, Werf, J.M. van, Verdonk, M.: Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. Computer. 43, pp. 90–93 (2010).

11. Jans, M., Lybaert, N., Vanhoof, K., Van Der Werf, J.M.: Business process mining for in-ternal fraud risk reduction: Results of a case study. (2008).

12. Jans, M., Alles, M., Vasarhelyi, M.: Process mining of event logs in auditing: opportunities and challenges. Working paper. Hasselt University, Belgium (2010).

13. Jans, M., Van der Werf, J.M., Lybaert, N., Vanhoof, K.: A business process mining application for internal transaction fraud mitigation. Expert Systems with Applications. 38, pp. 13351–13359 (2011).

14. Gehrke, N., Müller-Wickop, N.: Rekonstruktion von Geschäftsprozessen im Finanzwesen mit Financial Process Mining. Lecture Notes in Informatics, Proceedings der Jahrestagung Informatik. Leipzig (2010).

15. Gehrke, N.: The ERP Auditlab-A Prototypical Framework for Evaluating Enterprise Re-source Planning System Assurance. 43rd Hawaii International Conference on System Sciences (HICSS) (2010).

16. Jans, M., Alles, M., Vasarhelyi, M.: Process Mining of Event Logs in Internal Auditing: A Case Study. 2nd International Symposium on Accounting Information Systems (2011).

17. Müller-Wickop, N., Schultz, M., Gehrke, N., Nüttgens, M.: Towards Automated Financial Process Auditing: Aggregation and Visualization of Process Models. Proceedings of the Enterprise Modelling and Information Systems Architectures. Germany (2011).

18. Tiwari, A., Turner, C., Majeed, B.: A review of business process mining: state-of-the-art and future trends. Business Process Management Journal. 14,pp. 5–22 (2008).

19. Van der Aalst, W.M.P., Stahl, C.: Modeling business processes : a petri net-oriented approach. MIT Press, Cambridge, Mass. (2011).

20. Valk, R.: Lecture Notes: Formale Grundlagen der Informatik II (FGI 2) Modellierung & Analyse paralleler und verteilter Systeme. University of Hamburg (2008).

21. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems re-search. Mis Quarterly.pp. 75–105 (2004).

22. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decision support systems. 15,pp. 251–266 (1995).

23. Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J.: Memorandum on design-oriented information systems research. European Journal of Information Systems. 20, pp.7–10 (2010).

24. Brinkkemper, S.: Method engineering: engineering of information systems development methods and tools. Information and Software Technology. 38, pp.275–280 (1996).

25. Riege, C., Saat, J., Bucher, T.: Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik. Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik. pp.69–86 (2009).

26. Oracle: Welcome to NetBeans, http://netbeans.org/.

27. University of Hamburg: Renew - The Reference Net Workshop, http://www.renew.de/.

28. yWorks GmbH: yEd - Graph Editor, http://www.yworks.com/de/products_yed_about.html.

29. SAP: SAP-UCC, http://www.sap-ucc.com/.

30. Romney, M.B., Steinbart, P.J.: Accounting Information Systems. Prentice Hall (2008).

10.10 Multilevel Process Mining for Financial Audits

Number	**10**
Title	**Multilevel Process Mining for Financial Audits**[58]
Appendix	5.5, 6
Primary Related Chapters	10.10
Type	Journal Paper
Journal	IEEE Transactions on Services Computing
Reference	(Werner and Gehrke, 2015)
Acceptance Rate	[59]
VHB JQ 2.1 Ranking	-
WKWI Ranking	A
ERA 2010	-
CORE 2013	-
Review Procedure	Blinded
Number of Reviews	3[60]
Authors	1. Michael Werner 2. Nick Gehrke
Dissertation Points	0.67
Authorship	
Overall	95%
Design	95%
Realization	95%
Writing	95%
Status	Published[61]
Part of other Dissertations	No
Link	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=7277120

[58] The following article is the first revised version of the originally submitted manuscript. It reflects the status at the time of the submission of this dissertation. The final version was published in December 2015 and partly differs from the version presented in this thesis.

[59] The journal does not communicate acceptance rates for special or regular issues.

[60] The number of reviews refers to the first review round, three more reviews were received during the second review round.

[61] Paper was accepted for publication for the IEEE Transactions on Services Computing special issue on Processes Meet Big Data and published in December 2015, initial submission in August 2013, passed the first review round in December 2013 and second review round in July 2015.

Multilevel Process Mining for Financial Audits

MICHAEL WERNER, NICK GEHRKE

Abstract: The relevance of business intelligence increases with the growing amount of recorded data. The research on business intelligence has led to a mature set of methods and tools that are used in many application areas, but they are almost absent in the auditing industry. The audit of business processes is a critical part in financial audits. Public accountants face the challenge of being obligated to audit increasingly complex business processes that process huge amounts of transaction data. Process mining can be used as a business intelligence approach in the context of process audits to exploit this data. Key requirements for the application of process mining in financial audits are the reliability of the mining results, the integration of a data flow perspective and the ability to inspect data from the point of origin to the final output on the financial accounts. We introduce a process mining algorithm that integrates the control flow and data flow perspective. It operates on different abstraction levels, creates precise and fitting process models, accepts specific unlabeled event logs as input, and considers data relationships for inferring the control flow.

Index Terms: Business intelligence (BI), Financial Audits, Business Process Intelligence, Process Mining, Data Mining, Data Analysis, Business Process Modeling, ERP Systems, Design Science

1 Introduction

Financial audits are important for the smooth functioning of economic markets. They are a control mechanism to prevent the publication of false financial information. The reliability of published financial statements is crucial for stakeholders to direct their decisions. Governments have issued laws and regulations to ensure that companies prepare their financial statements truthfully and fairly. Public accountants act as referees ensuring the adherence to laws, regulations and accounting standards by auditing financial statements. A significant part of a financial audit is the auditing of business processes. The rationale of auditing business processes is the assumption that well-controlled business processes will lead to complete and correct postings on the financial accounts. The International Standard on Auditing (ISA) 315 (Revised) mandates the consideration of business processes and related information systems for financial audits: “The auditor shall obtain an understanding of the information system, including the related business processes, relevant to financial reporting (…)” [1, p. 7]. It is much more efficient to investigate the structure and embedded controls in business processes than to inspect single transactions. The task of auditing business processes is getting more and more challenging with the increasing integration of information systems for the automation of transaction processing and the growing amount of produced data. Traditional audit procedures like interviews and inspections of selected documents become inefficient or even ineffective in such audit environments [2]. Interview partners may no longer have overall information about a business process if part of it is operated automated and nontransparent in the information system without any human interaction. Furthermore it is questionable if the inspection of relatively few samples is a sufficient audit procedure when millions of transactions are processed.

The current situation in auditing engagements leads to an imbalance. Companies on the one hand use highly integrated information systems to support and automate the operation of business transactions leading to huge amounts of processed data. The auditors on the other hand apply traditional and mainly manual audit procedures that are not appropriate anymore in highly automated environments. The results are inefficient or even ineffective audits.

Problems arising from the increase of processed and available data are not idiosyncratic to the field of financial audits. Data explosion is a phenomenon that accompanies the expanding availability of digital data storage capacity [3]. The challenges and prospects that arise from the handling and analysis of huge data amounts are currently being discussed and intensively investigated under the umbrella of the term Big Data [4]. Big Data does not only occur in the field of information systems or computer science but also in very diverse scientific disciplines like biology [5] and physics [6]. Human recipients are only able to handle a certain amount of information before information overload hinders any additional information reception [7, pp. 53–57].

Business intelligence (BI) provides solutions to handle Big Data and to prevent information overload. It is a scientific domain that traditionally researches how digital data can be analyzed and presented to enhance decision making processes [8]. Chen et al. discuss the evolution of business intelligence and analytics (BI&A) and state that it is commonly "referred to as the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand its business and market and make timely business decisions." [4, p. 1166].

BI provides mature methods and tools that are used in many application areas. But they are almost absent in the auditing industry. Software tools that assist the auditor are called computer assisted audit tools (CAAT). Braun and Davis state that CAAT "include any use of technology to assist in the completion of an audit" [9, p. 726]. This is a very broad definition and it would mean that simple project management and documentation tools could also be considered as CAAT. Gehrke describes a more specific approach by illustrating how information on controls that are embedded in information systems can be tested using a prototypical software tool for the purpose of financial audits [10]. But the presented approach only uses a very small part of the data that is stored in information systems and merely focusses on the control perspective. The richest source of information is the recorded transaction data. This source remains mostly untouched by existing CAAT. Transaction data stored in Enterprise Resource Planning (ERP) systems exhibit characteristics of Big Data due to its volume and the velocity of data accumulation [4].

A promising solution to exploit the available transaction data in financial audits is the use of process mining. Process mining deals with the discovery, monitoring and enhancement of business processes by extracting information from event logs [11]. The usage of process mining would enable an auditor to receive reliable information for the audited business processes very efficiently and effectively. Van der Aalst provided a conceptual model for the integration of process mining into auditing [12] and scholars have applied process mining in the context of internal audits [13]. But process mining tools still have not been widely accepted and applied in the auditing industry.

A main reason is the specific characteristic of the application domain. Process mining algorithms can only be applied usefully if they fit the requirements of the application domain. The research area of process mining has matured during the last decade with the development of powerful general purpose mining algorithms such as heuristic [14], fuzzy [15] or genetic [16] mining algorithms. However, a significant aspect has not been investigated intensively yet. The vast majority of mining algorithms focusses on the control flow perspective. Other perspectives like the data flow perspective are neglected [17]. But the data flow is very important for financial audits. If a process is not compliant the auditor needs to assess the impact on the financial accounts. This requires the integration of financially relevant information. Process mining is commonly used to condense information by abstraction. Process models abstract from the observed behavior of single process executions. It is generally necessary to weigh competing quality criteria against each other when mining process models. Simple process models are normally preferred to complex ones even if this means that these simple models are not as precise and fitting. Auditors have to rely on the correctness of mined models to identify and to assess compliance violations. They therefore require perfectly fitting and precise process models. The auditor also has to be able to inspect individual process executions to find out which business transactions caused a violation. Common mining algorithms use labeled event logs. They contain ordered events that are mapped to cases [18]. Financially relevant transaction data stored in ERP systems cannot be used to create such an event log without prior preparation. But specific data relationships in this data can be exploited for process mining purposes [19]. The mining algorithm presented in this paper combines and visualizes the control flow and the data flow perspective. It accepts unlabeled event log data from ERP systems as input, produces perfectly fitting and precise process models, and uses data dependencies to determine the control flow. It is especially suitable as a special purpose mining algorithm for financial audits. But the presented approach for combining the control flow and the data flow perspective in a single model might also be valuable for other application domains that exhibit similar requirements.

2 Research Methodology and Structure

The research presented in this paper follows a design science research approach (DSR). This approach was chosen because of the proximity of the investigated research questions to the practical problems and the intention to develop artifacts that have a high contribution and relevance for the application domain. Österle et al. suggest following a research process that consists of the four phases analysis, design, evaluation and diffusion [20]. The structure of this article follows these phases. The analysis phase is represented in the subsequent sections that illustrate the state-of-the-art of related scientific work and discuss the requirements for the development of the presented mining algorithm. The design of the mining algorithm as the core contribution of the paper is presented in section five. Scholars like Hevner et al. stress the importance of research rigor in DSR [21]. We have therefore included an evaluation section that illustrates the results that have been achieved by applying the presented algorithms to extensive real world data. The paper closes with a discussion of the research results, relevant limitations, outlook to future research and a brief summary.

We used different research methods with varying paradigmatic orientation during the research process to reduce the risk of paradigmatic bias and to investigate the research problem from different angles [22]. The results presented in this paper are part of larger research effort that has been carried out by several researchers over the recent years. The development of the presented algorithm bases on the application domain requirements that were identified in empirical studies using expert interviews and surveys [23]. The main research methods that were used to develop the presented algorithm were method engineering [24] and prototyping [25]. Method engineering is an approach for assembling novel methods based on already existing or newly developed method fragments. Research outcomes from prior research have been incorporated as input for the presented research to develop a novel algorithm. Relevant research results that have been considered in the course of method engineering include solutions for the calculation of instance graphs, projections, aggregation graphs [26], and causal matrices [16] as well as previously published research results from the authors concerning the integration of the data perspective into process models [27], case matching [19], complexity reduction of mined models [28], and data dependent control flow inference [29]. The designed algorithm was instantiated in a software prototype to enable laboratory simulation experiments. Data sets from industry partners were used as input for the prototype to evaluate the proper functioning of the implemented algorithm and to analyze the mining results.

3 Related Work

The research presented in this paper deals with the question as to how process mining can be used as a BI approach for financial audits. Process mining is a research domain that has matured over the past decades. It would go beyond the scope of this paper to provide a complete overview of process mining research. Instead we will refer to literature that provides a good overview. Tiwari et al. provide a survey of the state-of-the-art and future trends in process mining until the year 2008 [30]. Basic and advanced process mining concepts have been comprehensively summarized by van der Aalst [3]. He provides an extensive collection of main research results that have been achieved in recent years and presents an overview of contemporary opportunities and challenges [31].

Compliance checking is a process mining application area that is of particular interest to the research at hand. Compliance can be defined as the adherence to internal or external rules. The main objective of financial audits is to ensure that companies adhere to accounting standards, laws and regulations. Debreceny and Gray suggest the application of data mining methods for analyzing journal entries and provide an extensive case study [32]. Becker et al. have researched the applicability of model-based business process compliance-checking approaches and developed a classification framework [33]. They provide a literature review and distinguish between forward- and backward-compliance checking approaches. Process mining for financial audits is a backward-compliance checking approach because its objective is the disclosure of compliance violations after they have occurred and been recorded in the event log. The application of process mining as a compliance checking approach has already been addressed by different scholars.

Alles et al. propose the application of process mining in accounting information systems [34]. Jans et al. highlight opportunities and challenges for using process mining as an audit tool [35] and provide interesting case studies [13], [36]. They focus on the control flow and organizational perspective. An important difference between internal and external audits is the relevance of the data perspective. Müller-Wickop et al. conducted empirical research which shows that internal and external auditors do not necessarily share the same perspective on the importance of different application domain constructs [23], [37]. We will discuss the requirements needed to use process mining in financial audits in greater detail in the subsequent section. But it is important to mention that the inclusion of the data perspective is a key requirement which has not been considered by prior research. The lion's share of process mining research deals with the discovery of the control flow whereas the integration of the data perspective in process mining has generally not been investigated extensively in the academic community yet, apart from very few scientific publications. This observation is supported by Stocker [38] and de Leoni and van der Aalst [17]. De Leoni and van der Aalst use the data flow perspective to discover rules that explain why instances of the same process follow different execution paths. They introduce variables as net components for the extension of Petri Nets (DPN-nets). Trcka et al. follow a different research question to discover data flow errors but apply a similar approach by using extended workflow nets (WFD-nets). Accorsi and Wonnemann choose a different approach to identify information leaks in process models. They include data objects as colored tokens in Colored Petri Nets (CPN) [39]. We follow a similar approach and use CPN to include the data perspective by modeling data objects as colored tokens [27]. This allows us to present the control flow and data flow perspective in a single model.

A fundamental challenge for process mining is the balancing between competing quality criteria [11]. Process mining is generally used to reduce complexity by visual representation and abstraction. A process model represents a set of process executions which are called process instances.[62] Multiple executions of a business process commonly do not occur in exactly the same manner. Variety in the execution leads to differing process instances. Every organization needs flexibility to adapt business activities to changing customer demands and market influences. A certain degree of deviation is therefore neither surprising nor damaging. But it is an obstacle to process mining because the objective of using process mining is to discover process models that describe the real business processes in the best possible way. Variance in the course of execution means that it can get impossible to create a model that unambiguously describes the represented process. This leads to the phenomenon of miss-fitting process models. Models are either under- or over-fitting. A model is under-fitting if it allows execution paths in the process model that are not represented in the event log and over-fitting if they do not allow for any additional behavior that is not included in the event log. Rozinat et al. provide a framework for the evaluation of process mining algorithms. They identify four quality criteria for the evaluation: fitness, preci-

[62] The term "process instance" and "case" are used ambiguously among scholars. We refer to "process instances" as real world executions of business processes whereas "cases" represent a record of a process instance in an event log. A case is therefore a purposeful abstraction of a process instance represented as a data record.

sion, generalization and structure [40]. Fitness indicates if a model is able to represent all cases in the event log. Precision is the complementary criterion. It indicates if a process model does not allow additional behavior that was not observed in the log. Generalization addresses the capability of a model to express more behavior than recorded in the log. It is generally desirable to create a process model that shows an adequate degree of generalization. Structure refers to the graphical representation of a business process and depends on the graphical components of the target language. Other scholars use the closely related criterion of simplicity instead of structure [11]. The following section will show that contrary to many other application domains financial audits require perfectly fitting and as precise process models as possible. De Medeiros et al. suggest the clustering of cases that exhibit similar traces in the event log.[63] Process models are then generated for each cluster [41]. This approach prevents over-generalization and is very useful because it does not require using a specific mining algorithm. A similar trace clustering approach is used by Song et al. [42]. Van der Aalst et al. provide a powerful mining algorithm for balancing between over- and under-fitting by using the theory of regions to create Petri Nets from transition systems [43]. The approaches by de Medeiros et al., Song et al. and van der Aalst et al. are very valuable but they do not consider the data perspective and the idiosyncratic structure of journal entries that form the basis for the event log in financial audits.

4 Requirements

This section describes the requirements for using process mining algorithms in financial audits. A business process consists of activities. Information systems that support or automate the execution of activities create journal entries that are posted to the relevant financial accounts. Fig. 1 illustrates this relationship and provides an example of a simple purchase process that consists of four activities.

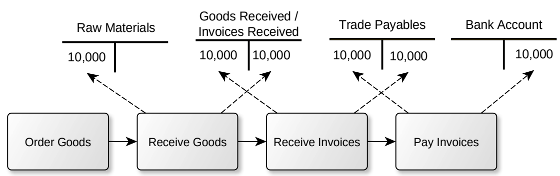


Fig. 1. Simple purchase process

Three of the four illustrated activities create journal entries on different accounts. Müller-Wickop et al. conducted empirical investigations using expert interviews and surveys to identify key concepts and information requirements for process audits [23], [44]. They show that the process flow is a central concept and highly relevant from an audit perspective in practice. To audit a process it is necessary to understand how a process is structured and which control activities are included that safeguard the correct processing of transactions. This requirement can be satisfied by modeling the control flow perspective. Two further important concepts for external auditors are financial statements and materiality[64]. Knowledge about the interaction among activities alone is not sufficient. For the audi-

[63] A trace is the recorded sequence of executed activities in a process instance. Every case has a specific trace but different cases can exhibit identical traces.

[64] Materiality is defined in ISA 320: “Misstatements, including omissions, are considered to be material if they, individually or in the aggregate, could reasonably be expected to influence the economic decisions of users taken on the basis of the financial statements.” [45].

tor it is necessary to understand how the activities relate to the financial accounts as illustrated in Fig. 1.

Only those business transactions are inspected in a financial audit that can have a material effect on the financial statements. It is therefore necessary to receive information on the value flow that is created by the audited business process to decide if it needs to be audited from a materiality perspective or if it can be neglected.

Another critical requirement in financial audits is the preservation of the audit trail. The audit trail is a fundamental concept in financial accounting. It is a path in an information system that allows tracing a transaction from the point of origin to the final output. It is used to verify the accuracy and validity of journal entries [46]. Translating this requirement into the context of process mining implies that a mining algorithm may not alter the original data during the mining process. A suitable mining algorithm must further be able to present the unchanged source data to the auditor for investigation purposes. But on the other hand the mining algorithm should also be able to present information at an adequate abstraction level to provide an overview of the control and data flow as discussed before. If process mining is used in financial audits the generated models are used to discover incompliant behavior. The provided process models should therefore be as precise and fitting as possible. If the produced process models are over-fitting certain behavior recorded in the event log is not represented in the process model. The auditor would therefore assume that no incompliant behavior has occurred. But in reality it is just not represented in the miss-fitting process model which would eventually lead to false positive audit results. If the process model is too general, process behavior is illustrated that actually did not occur. This would lead to false negative audit results and unnecessary investigations by the auditor. The quality criteria identified by Rozinat et al. [40] are useful to express the requirement of accurate process models in terms that are applicable for the process mining research domain. Simplicity of a process model is a preferred characteristic but it is not a key requirement in financial audits. Auditors currently spent weeks trying to understand a business process with the use of traditional audit procedures and by reviewing hundreds of documents. It is therefore acceptable if process models are complex and in extreme cases only com comprehensible to experts. Nevertheless a mining algorithm should be able to deliver process models as simple as possible. Generalization should be minimized and precision maximized to prevent false negative compliance testing results. Process models should be perfectly fitting to possibly represent all recorded behavior and to prevent that incompliant behavior that actually occurred remains undetected and therefore reduce the audit effectiveness. Different metrics can be used to measure fitness (*completeness* [47], $PF_{complete}$ [16], *fitness* (f) [48], *parsing measure (PM)* and *continuous parsing measure (CPM)* [14]. The metrics *completeness* and *PM* measure calculate the percentage of traces in the log that can be replayed by the model. The other three metrics consider both traces and tasks in a model. A process model has a perfect fitness if the metrics have a value of one. The precision dimension can also be measured by using different metrics (*soundness* [47], *behavioral appropriateness* [48], and *behavioral precision* [16]). A perfect precision is reached if the relevant metrics also take on the value of one.

5 Multilevel Process Mining

5.1 Mining Algorithm

The previous section discussed the requirements for using process mining in financial audits. These requirements are partially conflictive. Listing 1 shows an abstract and simplified version of the mining algorithm that was developed based on the identified requirements. A fundamental aspect of the presented solution is to satisfy different requirements at different abstraction levels. The algorithm is called Multilevel-Process-Mining algorithm (MLPM) due to its main feature of being able to produce process models at different abstraction levels. The MLPM first matches event data to cases, then creates instance graphs, calculates process instance models and aggregates instance models to process models. It is divided into four main sections. Each section is discussed in detail in the following subsections.

Listing 1 Multilevel Process Mining Algorithm

```
46. Mine Cases (Section 1)
47.   D      set of all posting document numbers
48.   J      set of all journal entry item numbers
49.   ID     set of all case IDs
50.   D_i = ∅ initially empty set of document numbers belonging to case i∈ID
51.   J_i = ∅ initially empty set of journal entry item numbers for case i∈ID
52.  While D ≠ ∅
53.     Remove d ∈ D from D and insert d into D_i
54.         Insert all j ∈ J posted by d into J_i and remove j from J
55.         Insert all d ∈ D that cleared j ∈ J_i into D_i and remove d from D
56.         Repeat 54. and 55. for all d ∈ D_i and j ∈ J_i

57. Reconstruct Instance Graphs (Section 2)
58.   IG = ∅   initially empty set of instances graphs
59.   IG_i      instance graph (N_i,E_i,L_i,l_i) for case i with the set of nodes
60.             N_i ≠ ∅, E_i ⊆ N_i × N_i is the set of arcs, L_i the set of task labels
61.             and l_i: N_i → L_i is a labeling function mapping nodes onto L_i
62.  For all i ∈ ID
63.     Create n ∈ N_i for each d ∈ D_i with l_i(n)= transaction code of d
64.     Create e(n_j,n_k) ∈ E_i for each d_j and d_k ∈ D_i if d_k cleared an item
65.         j ∈ J_i that was posted by d_j
66.     Insert IG_i into IG

67. Reconstruct Instance Models (Section 3)
68.   IM = ∅   initially empty set of instances models
69.   IM_i      instance model (T_i,P_i,A_i,Σ_i,V_i,C_i,G_i,E_i,I_i)for case i∈ID
70.  For all I i ∈ ID
71.     Set T_i = N_i
72.     For each e(n_j,n_k)∈ E_i create p ∈ P_i, a(t_j,p), a(p,t_k)
73.     For each d ∈ D_i
74.         Create p ∈ P_i for each j ∈ J_i that was posted by d
75.         Create a(t,p) ∈ A_i for each j ∈ J_i that was posted by d and
76.            a(t,p), a(p,t)∈ A_i for each j ∈ J_i that was cleared by d
77.     Aggregate all places p_k and p_j ∈ P_i if C_i(p_k) = C_i(p_j)
78.     Aggregate all transitions t_k and t_j ∈ T_i if l_i(t_k) = l_i(t_j)
79.     Insert IM_i into IM
```

```
80. Mine Process Models (Section 4)
81.   PM = ∅  initially empty set of process models
82.  Compute causal matrix CM(IM_i) for all i ∈ ID
83.  While IM ≠ ∅
84.      Remove IM_j from IM and insert IM_j into PM
85.      For each IM_k ∈ IM
86.          If CM(IM_j) = CM(IM_k)
87.             Merge IM_j and IM_k with T_jk=T_j∪T_k, P_jk=P_j∪P_k, A_jk=A_j∪A_k, Σ_jk=Σ_j∪ Σ_k
88.             Aggregate all places p_l and p_m ∈ P_jk if C_jk(p_l) = C_jk(p_m)
89.             Aggregate all t_l and t_m ∈ T_jk if l_jk(t_l) = l_jk(t_m)
90.             Remove IM_k from IM
```

5.2 Case Mining and Event Log Structure

The mining algorithm uses recorded transaction data from ERP systems as the data source for the mining. This kind of data only has a medium maturity level from a process mining perspective because ERP systems like SAP or JD Edwards do not use a systematic approach to link and store process relevant data [11]. The event log data in such systems is stored in different database tables and needs to be composed in a meaningful manner before it can be used as input for process mining algorithms. Data related to financially relevant business transactions exhibits specific characteristics that can be used for process mining purposes [27]. The execution of financially relevant transactions in an ERP system creates specific data records. Every execution of an activity creates a posting document. Each document contains at least two journal entry items. This connection is illustrated by *the contains* relationship in the ER model presented in Fig. 2. Journal entries that follow an open-item-accounting principle are linked to each other if they belong to the same process instance. If a process instance is terminated each open journal entry item has been cleared by another posting document. This dependency is modeled in Fig. 2 via the *is cleared* relationship.

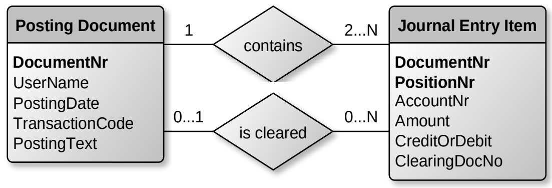


Fig. 2. Entity-relationship-model for accounting data structure

It is possible to exploit this data chain for each process instance. The procedure is illustrated in line 1 to 11 in Listing 1. The algorithm maps event log entries to cases. It starts with an arbitrary document number and mines all related journal entry items by using the *contains* relationship illustrated in Fig. 2. It then searches for all document numbers that have cleared these items by using the *is cleared* relationship. All posted items for every clearing document are then searched in further iterations. The loop terminates when all documents and items that belong to the same process instance have been found. The loop itself is repeated until all case IDs have been mined. The outcome is an event log $\bigcup_{i=1}^{n}\{D_i, J_i\}$ with $D_i = \{DocumentNr_{d1}, \dots, DocumentNr_{dm}\} \; \forall \; posting\ documents\ d_1 \dots d_m \xRightarrow{match} i$. The event log also includes the data attributes associated to each posting document and journal entry item that are listed as entity attributes in Fig. 2. In contrast to traditional event logs the events belonging to a single case do not follow a strict linear order. The causal dependencies between the events have to be determined in a separate step.

5.3 Instance Graphs and Abstraction Levels

The application of process mining for financial audits requires algorithms that enable the investigation of a financially relevant transaction from its point of origin to the posting on the financial accounts. On the other hand an auditor should be able to receive an overview to get an all-encompassing understanding of the process structure and its effect on the financial accounts. Both requirements can be met by using different abstraction levels. Scholars commonly distinguish between four levels of horizontal abstraction in business process management [49]. The instance level represents tangible entities that are involved in business processes which include executed activities, resources, concrete data values etc. A set of similar business processes are represented as business process models on the model level. A model is expressed using constructs of a meta-model. These can themselves again be defined at the meta-meta-model level. Process mining algorithms commonly operate on the model level. They generate process models that are abstractions of the individual process instances they represent.[65]

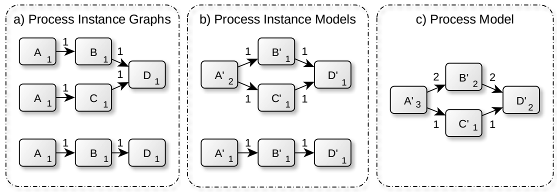


Fig. 3. Abstraction levels in the context of process mining[66]

Fig. 3 illustrates how the different abstraction levels in the context of business process models relate to each other. Process instance graphs reside on the lowest level of abstraction. They are graphical representations of the source data from executed and recorded process instances. Fig. 3a shows two instance graphs. Each rectangle represents the execution of an activity and the arcs between two activities denote the causal relationships between executed activities. The numbers indicate how often an activity was executed and how often the path from one activity to another was chosen. The process instance model is an abstraction of the instance graph (Fig. 3b). Executions of identical activities are aggregated into activity models but an instance model still represents just a single execution of a business process. A process model is an abstraction of a set of similar process instances (Fig. 3c). The process model is important for the auditor to get an overview about the structure of the process, its relationships to the financial accounts and to assess the overall materiality of a business process. Process instance models and process instance graphs are useful for following the audit trail and for inspecting individual

[65] An exception is the multi-phase process mining approach that was developed to generate event-driven-process-chains [26]. It operates on the instance and model level. A similar approach is used for the design of the MLPM.

[66] The models represent simple directed and labeled graphs for illustration purposes without logical operators. It is not necessary to model choice at the instance graph level because all decisions have already been made and all relationships in Fig. 3a have the semantic of an AND split or join. The models presented in Fig. 3b and 3c represent abstractions of the lower level models. The relationships on these level can generally represent AND, OR and XOR splits and joins. An algorithm to transform these models into EPC or Petri Nets is described in [26].

process instances with respect to involved activities, users, data values etc.

The second part of the mining algorithm ranging from line 12 to 21 creates instance graphs. The algorithm first creates a node for every document number labeled with the transaction code that was used to create the posting document in the ERP system (line 18). Different nodes in the instance graph can carry the same label (compare Fig. 2a) if they were created by the same transaction of the ERP system. The algorithm then infers the causal dependency between activities (line 19 and 20). Traditional mining algorithms rely on the time stamp of events to infer the control flow. This is not suitable for the event log created by the MLPM algorithm. The same posting document can clear journal entry items belonging to different other posting documents leading to parallelism in the event log. Using the temporal ordering of events can lead to intertwining in parallel branches resulting in models that are of little use to the user.[67] The algorithm uses instead a data dependent approach for determining the control flow. Sun and Zhao introduced an approach to derive the control flow by modeling the data flow [51]. The data dependencies of an activity v can be expressed as $\lambda^{d}{}_{v}(I^{d}{}_{v}, O_{v})$, where I_{v} is the input for v, and O_{v} the output. d denotes the type of data dependency. The algorithm analyses in lines 19 and 20 how activities relate to each other based on their data relationships. It checks if activity A has cleared $Item_B$ posted by activity B. If the condition is true a control arc from B to A is inserted. A can only have occurred if B took place before. Otherwise there would have been no $Item_B$ that could have been cleared by A. This is equivalent to a mandatory dependency between A and B denoted as $B \rightarrow {}_{m}A$ because of $O_B \cap I_A^u \neq \emptyset$ with $Item_B \in O_B \wedge Item_B \in I_A^u$. The results of these operations are instance graphs in form of directed graphs equivalent to the graph shown in Fig 2a.

5.4 Instance Models and Colored Petri Nets

The third section of the mining algorithm reconstructs process instance models as CPN. Petri nets are the predominant modeling language in the process mining research domain [30]. They are suitable for the modeling of business processes and offer a formal as well as graphical notation that can even be understood by non-experts [52]. They provide a sound mathematical foundation for the simulation and verification of Petri Net models [53]. The majority of process mining algorithms rely on low-level Petri Nets. An exception is the approach used by Accorsi and Wonnemann [39]. They use CPN and model data objects as colored tokens. We use a similar approach to generate process models that model the control flow and data flow perspective simultaneously in a single model. A Colored Petri Net is formally expressed by the tuple CPN = $(T, P, A, \Sigma, V, C, G, E, I)$ [53, p. 87], with:

1) $\boldsymbol{T}$ is a finite set of transitions
2) $\boldsymbol{P}$ is a finite set of places
3) $\boldsymbol{A} \in \boldsymbol{P} \times \boldsymbol{T} \cup \boldsymbol{T} \times \boldsymbol{P}$ is a set of directed arcs
4) $\boldsymbol{\Sigma}$ is a set of non-empty color sets
5) $\boldsymbol{V}$ is a finite set of typed variables such that Type$[\boldsymbol{v}] \in \boldsymbol{\Sigma}$ for all variables $\boldsymbol{v} \in \boldsymbol{V}$
6) $\boldsymbol{C}: \boldsymbol{P} \rightarrow \boldsymbol{\Sigma}$ is a color set function that assigns a color set to each place
7) $\boldsymbol{G}: \boldsymbol{T} \rightarrow \boldsymbol{EXPR_V}$ is a guard function that assigns a guard to each transition $\boldsymbol{t}$ such that $\boldsymbol{Type}\ [\boldsymbol{G}(\boldsymbol{t})] = \boldsymbol{boolean}$

[67] A discussion of this aspect is beyond the scope of this paper but it is illustrated in detail in [29]. The event logs can be converted into linear event logs [50]. But this transformation is accompanied with undesired side-effects like the duplication of events in the log, and it is not able to deal with the data perspective.

8) $\boldsymbol{E}: \boldsymbol{A} \rightarrow \boldsymbol{EXPR_V}$ is an arc expression function that assigns an arc expression to each arc $\boldsymbol{a}$ such that $\boldsymbol{Type}\ [\boldsymbol{E}(\boldsymbol{a})] = \boldsymbol{C}(\boldsymbol{p})_{\boldsymbol{MS}}$, where $\boldsymbol{p}$ is the place connected to the arc $\boldsymbol{a}$.
9) $\boldsymbol{I}: \boldsymbol{P} \rightarrow \boldsymbol{EXPR_\emptyset}$ is an initialization function that assigns an initialization expression to each place $\boldsymbol{p}$ such that $\boldsymbol{Type}\ [\boldsymbol{I}(\boldsymbol{p})] = \boldsymbol{C}(\boldsymbol{p})_{\boldsymbol{MS}}$

We integrate the data perspective by modeling colored places in the CPN that represent financial accounts [27]. An instance graph produced by section *Reconstruct Instance Graphs* is first transformed into a CPN. The nodes of the instance graph become the transitions of the CPN (line 26).[68] We distinguish between two types of places. *Control places* model the control flow and *account places* model the data flow. Each arc from the instance graph is transformed into a combination of a *control place* and two connecting *control arcs* in the instance model. A source place p_{source} is inserted with arcs $a_j\ (p_{source}, t_j)\ \forall\ t_j \in T \wedge \bullet\ t_j = \emptyset$ and a sink place p_{sink} is with arcs $a_k\ (t_k, p_{sink})\ \forall\ t_k \in T \wedge t_k \bullet = \emptyset$. The data perspective is integrated and visualized by creating *account places* for each journal entry item in the event log (line 29). The color set function C assigns different color sets to places depending on whether they belong to the group of control or account places.[69] The set of color sets Σ includes the color sets for all possible journal entry values, account numbers, account types, credit or debit indicators and execution numbers.

Account places are connected to related transitions (line 30 and 31). A simple arc $a(t, p)$ is inserted from transition t to the place p if the transition posted a journal entry item on the represented financial account. They are referred to as *posting arcs*. Two arcs $a(t, p) \wedge a(p, t)$ are inserted if the transition has cleared an item on the respective account. These arcs have the semantic of testing arcs because they do not consume the tokens on the connected places. Double-headed arcs are used for visualization as a syntactical abbreviation for two arcs $a(t, p) \wedge a(p, t)$. They are called *clearing arcs*. The arc inscriptions are modeled as constants ($V = \{\}$). The arc expression function E assigns to each *posting* and *clearing arc* a set of constants that denote the posted or cleared value, the account type, account number and an indicator whether it is a credit or debit posting. Each *control arc* is assigned a number that indicates how often this represented control path was chosen in the instance.[69] The source place is initialized by the initialization function I. The initialization expression for p_{source} generates n tokens in the initial marking $M_0(p)$, one for each connected start transition.

After having integrated the data perspective by creating a CPN it is necessary to aggregate model components to create instance models. The first step is aggregating places representing financial accounts. Two places p_k and p_j can be aggregated if they represent the same account. This is the case if they carry the same color $C(p_k) = C(p_j)$ (line 32). The relationships between the places and connected transitions need to be maintained in respect to the type and inscription of each arc.[70] The second step aggregates the transitions (line 33). The used method for aggregating transitions is based on the algorithm described by van Dongen

[68] Guards are not needed and the guard function G is therefore defined as $G(t) = \text{true}$ for all $t \in T$.
[69] A detailed description of the color set function C and arc expression function E is provided in [27].
[70] A complete description of this procedure is beyond the scope of this paper. It is described in [28].

and van der Aalst [26]. Transitions are aggregated if they carry the same label $l(t_j) = l(t_k)$. The result of the third section of the mining algorithm is a set of instance process models represented as CPN.

5.5 Process Models and Clustering

The final section of the mining algorithm produces process models (lines 35 to 45). A key requirement for the application of process mining in the context of financial audits is the creation of perfectly fitting and precise process models. Some researchers have developed methods to create precise and fitting process models by clustering traces in the event log [41],[42]. Their approaches are not directly applicable because they do not take into account the data perspective and because they are not suitable for the given event log structure. But clustering is an approach that is also used for the creation of process models in the final section of the presented mining algorithm. But instead of clustering traces we cluster and aggregate process instance models. We use causal matrices to identify isomorph process instances. Causal matrices are key components for heuristic [14] and genetic mining algorithms [54, p.6]. A causal matrix describes the causal relation between activities in a model. An entry in a causal matrix in column j and row k designates that a causal relation exists between the activities j and k. The causal matrices of two instance models are identical if they show exactly the same behavior. The algorithm computes the causal matrices for all instance models (line 37) that were created by the previous operations. It then searches for identical causal matrices (line 41) and aggregates two models if they have the same causal matrix (lines 42 to 44). The aggregation procedures are identical to the procedures used at the instance model level (line 32 and 33). The aggregation procedure is repeated for all instance models until the set of instance models is empty (line 38).

Aggregating process instance models that exhibit identical control flow patterns does not mean that the resulting process model is identical to the source process instance models. The data values in each process instance model are unique and are preserved during the aggregation process. A process model therefore shows the aggregated data values and data flow of all represented process instances. The final outcome of the mining algorithm is a set of process models where each model represents process instances with the same control flow pattern. Each process model provides the aggregate view of the data flow and relationship of activities with the financial accounts.

6 Mining Results and Evaluation

Many scholars stress the importance of evaluation for rigorous DSR [20], [21], [55]. Riege et al. provide a categorization of evaluation methods and point out that it is not sufficient in DSR to evaluate research artifacts against the epistemological objective but also against the design objective [56]. Venable et al. provide a framework for the evaluation in DSR that can be used to support the selection of adequate evaluation methods [57]. The aim of the conducted evaluation is to test whether the created artifact is able to produce the estimated results. An artificial evaluation is appropriate for the research at hand because it is not necessary to observe the behavior in organizational contexts at this stage. We have chosen an artificial laboratory experiment as one of the evaluation methods suggested for this setting by Venable et al. [57]. Such an evaluation requires the existence of instantiated artifacts that can be used for the experiment. The different sections of the mining algorithms were therefore implemented in a software artifact in iterative cycles. The experiment itself was divided into the phases data extraction, mining and results analysis. We used a separate extraction module for extracting relevant data from ERP systems that can be adjusted to different source systems. We checked the data for first and second order defects [8, pp. 29–32].

Three data sets were used for the evaluation. Their characteristics are listed in Table 1. We extracted data from the productive SAP systems of three companies operating in the retail, manufacturing and media industry The MLPM was used to mine all process instance graphs, process instance models and process models for the three data sets. The models were inspected on a sample basis by observation and comparison with the original event log data. The software yEd – Graph Editor [58] provides powerful automatic layout functionality and is free for use for noncommercial purposes. It uses the GXML and GraphML formats as input and was used in the experiment to graphically represent the process models. The models were further tested for soundness (proper completion, option to complete, absence of dead transitions, and safeness for all places except account places) by using the CPN simulation tool Renew [59].

TABLE 1 Evaluation Data Sets

Set	Industry	Journal Entries	Journal Entry Items	Process Instances	Process Models
1	Manufacturing	1,764,773	7,395,434	1,035,805	841
2	Media	156,604	559,506	18,975	516
3	Retail	92,487	222,901	40,634	307

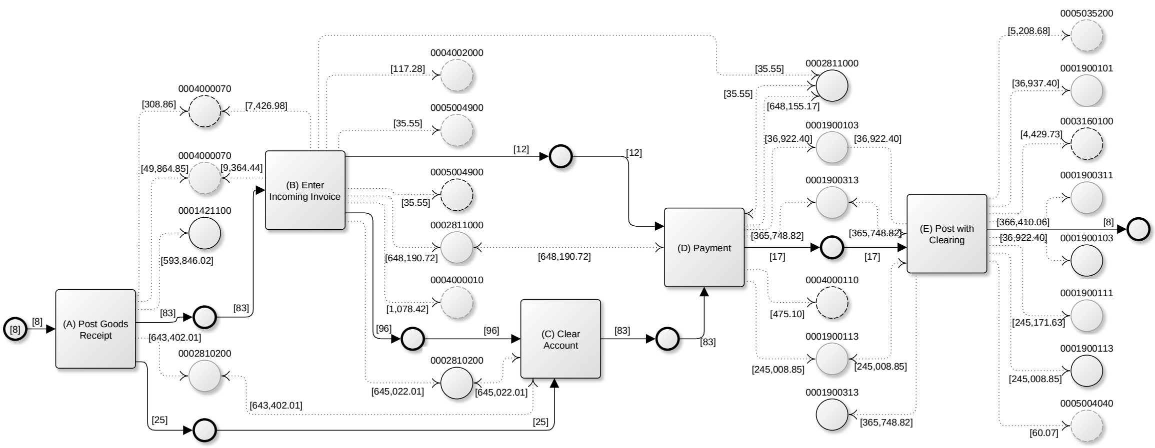


Fig. 4. Example of a mined process model[71]

Fig. 4[71] shows an example of a mined process model from data set 1. It is modeled as a CPN as specified in section 5.4. The transitions illustrated as rectangles represent the activities. The mined model represents a

[71] The arc inscriptions for posting and clearing arcs only display the assigned constant for the posted or cleared value. The inscriptions for the account type, account number and credit or debit indicator are omitted for better readability. The same is the case for the inscriptions of the connected account places that only show the account number.

purchasing process that was executed by 8 identical process instances. It contains the activities *Post Goods Receipt*, *Enter Incoming Invoice*, *Clear Account*, *Payment* and *Post with Clearing* that were executed in a sequential order. Control places are rendered with a bold black border. Simple arrows between transitions and control places model the control flow and sequence of activities. The inscriptions for these arcs show how often the path was chosen in the represented instances. Account places have a solid or dashed border. The type and color of the border line indicate if the represented account is a balance sheet or profit and loss account and if it represents the debit or credit side.[72] Account places and transitions are connected by dotted arcs. A simple dotted arrow is a posting arc. Double-headed dotted arcs are clearing arcs. Posting and clearing arcs visualize the data flow in the process model. The *Post Goods Receipt* activity for example creates a token with the value of *593,842.00* representing a journal entry item posting on the raw materials account *0001421100*. Another token is created on the account *0002810200*. This is cleared by the subsequent activity of *Clear Account* but without consuming the token on the respective account. The modeled CPN mimics the behavior of the represented process. Transitions create colored tokens on the account places representing the posted journal entries. The control places define the control flow in the model.

The model is perfectly fitting and precise because it can replay all process instances and does not allow any additional behavior. Fitness and precision can be measured using different metrics such as *fitness* (f) and *behavioral appropriateness* (a_B) [48]. These measures can be calculated by replaying cases from the event log and by counting if tokens are missing in the CPN in order to execute the simulation or remain unconsumed in the model after the simulation is terminated. These matrices are not directly applicable to the used CPN. Colored data tokens remain in the model by definition even after all transitions have fired. The used events in the event log do not follow a strict linear order for each case. It is therefore unclear which sequence of events should be used for replaying a case. The traces can be transformed into strict linear sequences [50] but these sequences would not fit to the mined model anymore. The possible transformation into linearized traces would introduce choice at the instance level that did not occur in reality. The model illustrated in Fig. 4 does not show any choice. The paths $p_1 = B \rightarrow D$ and $p_2 = B \rightarrow C \rightarrow D$ are parallel and not optional paths. The trace $A \rightarrow B \rightarrow D \rightarrow E$ for example, which would be an output of the transformation, cannot be replayed by the model from Fig. 4.[73]

Calculating token based measures like (f) and (a_B) for the produced process model would require many artificial adjustments which are not reflected by the actual source data. We alternatively use metrics that are suitable to take into account possible parallelism at the instance level and that directly compare the execution paths in the different models. We use the percentage of the control flow paths from $p_{source} \rightarrow p_{sink}$ that are present in the instance graphs to those in the process model for measuring the fitness denoted as f_{Path}. And we compute the percentage of the control flow

[72] Solid line = balance sheet account, dashed line = profit and loss account, black line = debit side of an account, gray line = credit side of a account.

[73] The model itself can be transformed in such a way that it is able to replay all linearized traces by using the procedures described in [28]. But this would result in a much more complex models and a negative effect on the model precision.

paths in the process model that are not present in the instance graphs for measuring the precision denoted as p_{Path} with:

$$f_{Path} = \frac{|\{p \mid p \in P_{PM} \wedge \in P_{IG}\}|}{|P_{IG}|}$$

$$p_{Path} = \frac{|P_{PM}| - |\{p \mid p \in P_{PM} \wedge \notin P_{IG}\}|}{|P_{PM}|}$$

$P_{PM} = \{p_1, \dots, p_n\}$ is the set of distinct control flow paths in the process model PM and $P_{IG} = \{p_1, \dots, p_m\}$ is the set of all distinct control flow paths from all instance graphs IG that belong to PM.

Van Dongen and van der Aalst proof that the used aggregation procedures are path preserving [26]. This means that all control flow paths are also represented in the process model with f_{Path} = 1. We calculated p_{Path} for the mined models from data set 1 to 3. The average value for this measure was 0.81.[74] This means that 19 % of the paths in the process models may actually represent behavior that was not recorded in the event log. It is an acceptable result compared to outcomes of artificial simulations from other scholars that use comparable metrics [40]. But it has to be validated in further research if this precision is high enough for the application in real world scenarios.

The models produced by the MLPM from data set 1 to 3 were analyzed using descriptive statistics. Fig. 5 to 8 show selected results for data sets 1 and 2. Fig. 5 and 7 show the frequency distributions of the mined process models depending on their model complexity measured as the number of included transitions. They show that the process models are distributed comparably to a normal distribution. Fig. 6 and 8 present scatter diagrams for data set 1 and 2. They illustrate the distribution of the number of represented instance models in a process model depending on the model size. The diagrams show that the vast majority of process instances actually belong to very simple process models that contain only a few transitions. This observation confirms preliminary results from prior research work [60].

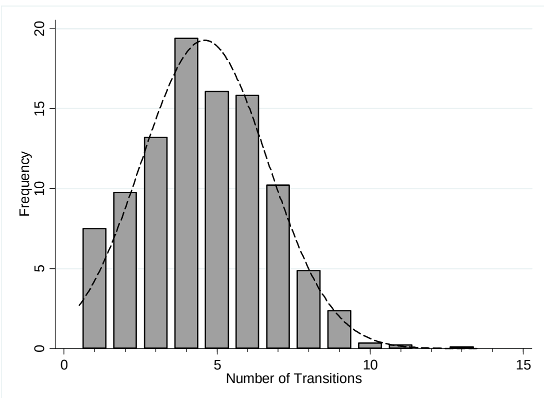


Fig. 5. Frequency distribution for data set 1[75]

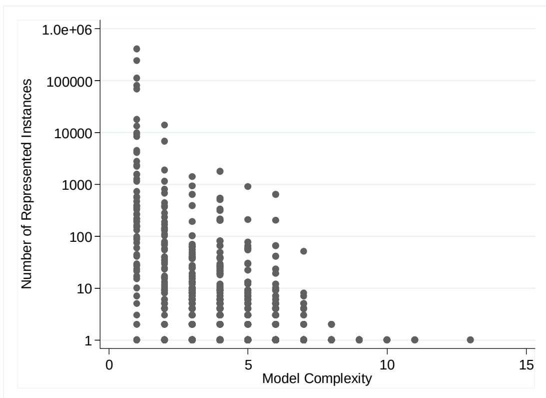


Fig. 6. Scatter diagram for data set 1[76]

[74] Process models including loops were excluded from the calculation

[75] Process models including loops were excluded from the calculation

[76] The dependent variables in Fig. 6 and 8 use a logarithmic scaling.

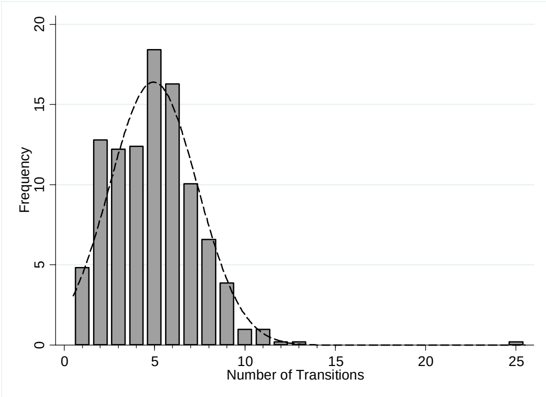


Fig. 7. Frequency distribution for data set 2

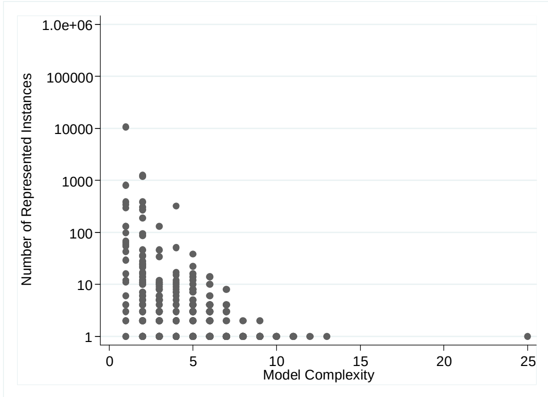


Fig. 8. Scatter diagram for data set 2[76]

Table 2 Descriptive Statistics

Parameter	Data Set 1	Data Set 2	Data Set 3
Mean value of transitions per process model	4.61	4.95	3.98
Median value of transitions per process model	5	5	4
Standard deviation of transitions per process model	2.10	2.43	2.24
Maximum value of transitions per process model	15	25	21
Minimum value of transitions per process model	1	1	1

The diagrams for the third data set are not included due to space restrictions. They follow similar patterns. Some relevant descriptive statistical values are listed in Table 2 for all data sets.

7 Discussion

The main contribution of this paper is the introduction of a process mining algorithm that is able to discover the control flow and the data flow perspective in process models by simultaneously providing perfectly fitting and precise models at different abstraction levels. The innovation of the presented artifact is achieved by combining already existing and newly developed methods that lead to a novel solution for a new application area. It is often questioned if research results derived from DSR can be equally important as results provided by other research approaches. March and Smith emphasize that both, design and natural science, have to coexist and benefit from each other to promote the progress of information systems research [55]. DSR has the potential to create artifacts that are of practical relevance and prescriptive nature. These artifacts can in return be the subject of descriptive science. Gregor and Hevner provide a useful framework for the categorization of knowledge contribution by DSR [61]. They categorize research work into the four domains routine design, improvement, exaptation and invention. The presented research work can be assigned to the exaptation quadrant because the main objective is to provide a solution for a new application area by partly using already existing knowledge. But it also affects the improvement quadrant by introducing a new method to model the control flow and data simultaneously in mined process models. Gregor and Hevner further differentiate between three levels of contribution types that range from abstract, complete and mature knowledge on the highest level to more specific, limited and less mature knowledge on the lowest level. The research results presented in this paper are mainly located on the second level providing constructs and methods for the mining of process models and on the first level presenting an instantiated software artifact. The results that can be achieved by analyzing the mining outcomes can also be input for the third and highest knowledge contribution level. The distributions of the number of instances over the number of transitions in Fig. 6 and 8 for example can lead to the assumption that their distribution curves are very similar. But the descriptive statistics in Table 2 shows that the mean values for the number of transitions in the process model differ quite significantly from each other with 4.61 for data set 1, 4.95 for set 2 and 3.98 for set 3. It could hypothesized that the complexity of the mined process models relates to the maturity of the mined business processes following the assumption that a mature process is more integrated into information systems than a less mature. This research question surely needs further investigation, but it highlights how the presented results can be the starting point for further theoretical research.

The presented mining algorithm is able to discover process models in accordance with the identified requirements to a large extent. The mined models are not absolutely precise. It needs to be validated in further research if the achieved level of precision is sufficient in practice. Several other limitations need to be taken into account. The process models do not represent sound workflow nets according to commonly used definitions [3, p. 39]. This handicap is not too severe because the objective of process mining for financial audits is the adequate modeling of the control and data flow perspective with precise and fitting process models. Formally well-structured process models are of minor interest. But if it should be necessary soundness could be achieved when account places are neglected. The remaining models would then represent sound workflow nets but without representing the data flow perspective. The data sets were all extracted from SAP systems. It can therefore not be concluded that the research results also hold true for other data sources. But an important advantage of the used mining algorithm is its independence from the implemented data structures of a particular ERP system because it bases on the general structure of accounting entries. Some process models showed loops that occur when a transaction has cleared a journal item that was posted by the same transaction or by a transaction located in the subsequent execution path. This constellation leads to a deadlock in the process model which is not critical for the interpretation of the model but generally not desired for the modeling of correct process models. A solution could be the prevention of aggregating transitions carrying the same label if this would result in a loop.

The mining algorithm produces precise and fitting process models at the cost of lacking generalization. It is therefore not applicable for scenarios with highly variable business processes. In the worst case scenario all process instances show a different behavior. The mining algorithm would then not be able to aggregate any instance models and the set of process instances models would be identical to the set of process models resulting in no or little information gain. The data presented in Table 1 shows that this risk is not acute for the given application area. Business processes are usually standardized to a certain degree when they are supported by ERP systems. The data shows that the number of process models ranges from 307 for the smallest data set to 841 models for the largest. This may still seem to be a big number. 63 process models in data set 1 only consist of one transition.

These represent trivial processes. The activities in these processes were mostly carried out by using a single general purpose transaction. They are of little interest from a process perspective but highly important from an audit perspective because this category of process models represent 96% of process instances in data set 1, 70% in set 2 and 51% in set 3. It is clear that further analytical procedures are necessary to address this category. A starting point could be the clustering of process models that use the same accounts. The process models that reflect the major business processes are those that contain many transitions and represent a high number of process instances. Data set 1 contains 135 process models consisting of 5 transitions. But just two of them already represent 62% of the instances of this category. It can therefore be assumed that the majority of instances for more complex process models only represent very infrequent behavior and can be tested traditionally by inspecting individual journal entries. The process models that represent many process instances and create a high value flow are interesting from a materiality perspective and can be audited by including the testing of embedded application controls [62].

8 Conclusion

The amount of available data increases with the integration of information systems for the support and automation of business activities. BI is a research domain that provides mature methods and tools that can be used to exploit and handle the growing amount of data. While it is commonly used in many application scenarios it is almost absent in the auditing industry. Traditional audit procedures are not efficient and effective in audit environments with highly integrated information systems and an increasing amount of processed data. The audit of business processes is a significant part in the financial audit. Process mining can be applied as a BI approach to support and automate the audit of business processes. The selection of a process mining algorithm should be founded on the analysis of relevant application domain requirements. For the case of financial audits it is crucial that a mining algorithm is able to model both the control flow and data flow. The algorithm should preserve the audit trail and produce perfectly fitting and as precise process models as possible. We have designed and evaluated a multilevel process mining algorithm that meets these requirements to a large extent. It introduces novel constructs and methods for the mining of process models and an instantiated software artifact. The results derived by exposing the designed artifact to extensive real life data can be the starting point for future theory building.

Data sets from the SAP systems of three different companies operating in diverse industries were used for the evaluation of the designed artifact. It cannot be concluded that the results hold true for other ERP systems and industries but the implemented mining algorithm exploits the structure of accounting entries that is system-independent and should therefore be generally applicable. The extension to other ERP systems will be covered in future research.

Public accountants face the challenge to audit increasingly complex and integrated business processes that process huge amount of data. The presented mining algorithm exploits large data sets that are created during the operation of business processes. It provides a suitable solution for analyzing business processes in financial audits but it can also be applied in application contexts that exhibit similar requirements.

A basic limitation of the presented algorithm is the lack of generalization. This is a desired characteristic for financial audits but it is not appropriate in settings with highly variable business processes. Other mining approaches using a two-step approach [43] or fuzzy mining [15] might be more appropriate in these settings. Some restrictions still exist in certain process constellations that create deadlocks in the process models. This phenomenon and adequate solutions still need to be researched in the future.

References

[1] International Federation of Accountants, "ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment." 2012.

[2] M. Werner and N. Gehrke, "Potentiale und Grenzen automatisierter Prozessprüfungen durch Prozessrekonstruktionen," in Forschung für die Wirtschaft, G. Plate, Ed. Aachen: Shaker Verlag, 2011.

[3] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st Edition. Berlin Heidelberg: Springer, 2011.

[4] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, vol. 36, no. 4, pp. 1165–1188, Dec. 2012.

[5] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Han-nick, W. Hide, D. P. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Twigger, O. White, and S. Yon Rhee, "Big data: The future of biocuration," Nature, vol. 455, no. 7209, pp. 47–50, Sep. 2008.

[6] C. Lynch, "Big data: How do your data grow?," Nature, vol. 455, no. 7209, pp. 28–29, Sep. 2008.

[7] H. Krcmar, Informationsmanagement. Berlin; Heidelberg: Springer, 2010.

[8] H.-G. Kemper, W. Mehanna, and H. Baars, Business intelligence - Grundlagen und praktische Anwendungen : eine Einführung in die IT-basierte Managementunterstützung. Wiesbaden: Vieweg + Teubner, 2010.

[9] R. L. Braun and H. E. Davis, "Computer-assisted audit tools and techniques: Analysis and perspectives," Managerial Auditing Journal, vol. 18, no. 9, pp. 725–731, 2003.

[10] N. Gehrke, "The ERP Auditlab - A Prototypical Framework for Evaluating Enterprise Resource Planning System Assurance," in Proceedings of the 43th Hawaii International Conference on System Sciences, Kauai, 2010, pp. 1–9.

[11] W. M. P. van der Aalst, A. Andriansyah, A. K. de Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. van den Brand, R. Brandtjen, and J. Buijs "Process Mining Manifesto," in BPM 2011 Workshops Proceedings, 2012, pp. 169–194.

[12] W. M. P. van der Aalst, K. M. van Hee, J. M. van Werf, and M. Verdonk, "Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor," Computer, vol. 43, no. 3, pp. 90–93, Mar. 2010.

[13] M. Jans, M. Alles, and M. Vasarhelyi, "Process Mining of Event Logs in Internal Auditing: A Case Study," in 2nd International Symposium on Account-ing Information Systems, 2011.

[14] A. Weijters, W. M. P. van der Aalst, and A. K. A. de Medeiros, "Process mining with the heuristics miner-algorithm," Technische Universiteit

Eindhoven, Tech. Rep. WP, vol. 166, 2006.

[15] C. Günther and W. van der Aalst, "Fuzzy mining–adaptive process simplification based on multi-perspective metrics," Business Process Management, pp. 328–343, 2007.

[16] A. K. A. de Medeiros, "Genetic Process Mining," Eindhoven University of Technology, Eindhoven, 2006.

[17] M. de Leoni and W. M. van der Aalst, "Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments," 2013.

[18] D. Ferreira and D. Gillblad, "Discovering Process Models from Unlabelled Event Logs," Business Process Management, pp. 143–158, 2009.

[19] N. Gehrke and N. Müller-Wickop, "Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems," in Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru, 2010.

[20] H. Österle, J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, and E. J. Sinz, "Memorandum on design-oriented information systems research," European Journal of Information Systems, vol. 20, no. 1, pp. 7–10, 2010.

[21] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," MIS Quarterly, pp. 75–105, 2004.

[22] J. Mingers, "Combining IS research methods: to-wards a pluralist methodology," Information systems research, vol. 12, no. 3, pp. 240–259, 2001.

[23] N. Müller-Wickop, M. Schultz, and M. Peris, "To-wards Key Concepts for Process Audits – A Multi-Method Research Approach," in Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics, Utrecht, 2013.

[24] S. Brinkkemper, "Method engineering: engineering of information systems development methods and tools," Information and Software Technology, vol. 38, no. 4, pp. 275–280, 1996.

[25] T. Wilde and T. Hess, "Forschungsmethoden der Wirtschaftsinformatik," Wirtschaftsinformatik, vol. 49, no. 4, pp. 280–287, 2007.

[26] B. F. van Dongen and W. M. P. van der Aalst, "Multi-phase process mining: Aggregating instance graphs into EPCs and Petri nets," in PNCWB 2005 workshop, 2005, pp. 35–58.

[27] M. Werner, "Colored Petri Nets for Integrating the Data Perspective in Process Audits," in Proceedings of 32nd International Conference on Conceptual Modeling (ER 2013), Hong Kong, China, 2013, pp. 387–394.

[28] M. Werner, M. Schultz, N. Müller-Wickop, N. Gehrke, and M. Nüttgens, "Tackling Complexity: Process Reconstruction and Graph Transformation for Financial Audits (Research in Progress)," in Proceedings of 33rd International Conference on Information Systems, Orlando, 2012.

[29] M. Werner and M. Nüttgens, "Improving Structure - Logical Sequencing of Process Models," in Proceedings of the 47th Hawaii International Conference on System Sciences, Big Island, 2014.

[30] A. Tiwari, C. J. Turner, and B. Majeed, "A review of business process mining:

state-of-the-art and future trends," Business Process Management Journal, vol. 14, no. 1, pp. 5–22, 2008.

[31] W. M. P. van der Aalst, "Process Mining: Overview and Opportunities," ACM Transactions on Management Information Systems, vol. 99, no. 99, pp. 1–16, Feb. 2012.

[32] R. S. Debreceny and G. L. Gray, "Data mining journal entries for fraud detection: An exploratory study," International Journal of Accounting Information Systems, vol. 11, no. 3, pp. 157–181, Sep. 2010.

[33] J. Becker, P. Delfmann, M. Eggert, and S. Schwittay, "Generalizability and Applicability of Model- Based Business Process Compliance-Checking Approaches – A State-of-the-Art Analysis and Research Roadmap," BuR - Business Research, vol. 5, no. 2, pp. 221–247, Nov. 2012.

[34] M. Alles, M. Jans, and M. Vasarhelyi, "Process Mining: A New Research Methodology for AIS," in CAAA Annual Conference 2011, 2011.

[35] M. J. Jans, "Process Mining in Auditing: From Current Limitations to Future Challenges," in Business Process Management Workshops, vol. 100, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer, 2012, pp. 394–397.

[36] M. Jans, N. Lybaert, K. Vanhoof, and J. M. Van Der Werf, "Business Process Mining for Internal Fraud Risk Reduction: Results of a Case Study," 2008.

[37] N. Müller-Wickop and M. Schultz, "Modelling Concepts For Process Audits - Empirically Grounded Extension Of BPMN," in Proceedings of the 21st European Conference on Information Systems, Utrecht, 2013.

[38] T. Stocker, "Data flow-oriented process mining to support security audits," in Service-Oriented Computing-ICSOC 2011 Workshops, 2012, pp. 171–176.

[39] R. Accorsi and C. Wonnemann, "InDico: Information Flow Analysis of Business Processes for Confidentiality Requirements," in Security and Trust Management, Springer, 2011, pp. 194–209.

[40] A. Rozinat, A. K. A. de Medeiros, C. W. Günther, A. Weijters, and W. M. van der Aalst, "The Need for a Process Mining Evaluation Framework in Research and Practice," in Business Process Management Workshops, 2008, pp. 84–89.

[41] A. K. A. de Medeiros, A. Guzzo, G. Greco, W. M. P. van Der Aalst, A. Weijters, B. F. Van Dongen, and D. Saccà, "Process Mining Based on Clustering: A Quest for Precision," in Business Process Management Workshops, 2008, pp. 17–29.

[42] M. Song, C. W. Günther, and W. M. Van der Aalst, "Trace Clustering in Process Mining," in Business Process Management Workshops, 2009, pp. 109–120.

[43] W. M. P. van der Aalst, V. Rubin, H. M. W. Verbeek, B. F. Dongen, E. Kindler, and C. W. Günther, "Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting," Software & Systems Modeling, vol. 9, no. 1, pp. 87–111, Nov. 2008.

[44] N. Müller-Wickop, M. Schultz, N. Gehrke, and M. Nüttgens, "Towards

Automated Financial Process Auditing: Aggregation and Visualization of Process Models," in Proceedings of the Enterprise Modelling and Information Systems Architectures, Germany, 2011.

[45] International Federation of Accountants, "ISA 320 Materiality in Planning and Performing an Audit." 2009.

[46] M. B. Romney and P. J. Steinbart, Accounting Information Systems, 11th Revised edition (REV). Prentice Hall, 2008.

[47] G. Greco, A. Guzzo, L. Pontieri, and D. Sacca, "Mining expressive process models by clustering workflow traces," in Advances in Knowledge Discovery and Data Mining, Springer, 2004, pp. 52–62.

[48] A. Rozinat and W. M. P. van der Aalst, "Conformance Checking of Processes Based on Monitoring Real Behavior," Information Systems, vol. 33, no. 1, pp. 64–95, Mar. 2008.

[49] M. Weske, Business process management concepts, languages, architectures. Berlin; New York: Springer, 2012.

[50] N. Mueller-Wickop and M. Schultz, "ERP Event Log Preprocessing: Timestamps vs. Accounting Logic," in Design Science at the Intersection of Physical and Virtual Design, Berlin, Heidelberg, 2013, vol. 7939, pp. 105–119.

[51] S. X. Sun and J. L. Zhao, "Formal workflow design analytics using data flow modeling," Decision Support Systems, vol. 55, no. 1, pp. 270–283, Apr. 2013.

[52] W. M. P. van der Aalst and C. Stahl, Modeling business processes : a petri net-oriented approach. Cam-bridge, Mass.: MIT Press, 2011.

[53] K. Jensen and L. M. Kristensen, Coloured petri nets. Springer, 2009.

[55] S. T. March and G. F. Smith, "Design and natural science research on information technology," Decision support systems, vol. 15, no. 4, pp. 251–266, 1995.

[56] C. Riege, J. Saat, and T. Bucher, "Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik," Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik, pp. 69–86, 2009.

[57] J. Venable, J. Pries-Heje, and R. Baskerville, "A comprehensive framework for evaluation in design science research," Design Science Research in Information Systems. Advances in Theory and Practice, pp. 423–438, 2012.

[58] yWorks GmbH, "yEd - Graph Editor," 2013.: http://www.yworks.com/de/products_yed_about.html.

[59] University of Hamburg, "Renew - The Reference Net Workshop," 2013. http://www.renew.de/.

[60] M. Werner, N. Gehrke, and M. Nüttgens, "Towards Automated Analysis of Business Processes for Financial Audits," in Proceedings of the 11th International Conference on Wirtschaftsinformatik, Leipzig, 2013.

[61] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quarterly, vol. 37, no. 2, pp. 337–355, 2013.

[62] M. Werner, N. Gehrke, and M. Nüttgens, "Business Process Mining

and Reconstruction for Financial Audits," in Proceedings of the 45th Hawaii International Conference on System Sciences, Maui, 2012, pp. 5350–5359.

11 Appendix B: Publication List from Literature Review

Title	Authors	Publication Outlet
a-Algorithm: Structured Workflow Process Mining Through Amalgamating Temporal Workcases	Kwanghoon Kim,Clarence A. Ellis	Advances in Knowledge Discovery and Data Mining
A business process mining application for internal transaction fraud mitigation	Mieke J. Jans,Jan Martijn van der Werf,Nadine Lybaert,Koen Vanhoof	Expert Systems with Applications
A Case Study on the Suitability of Process Mining to Produce Current-State RBAC Models	Maria Leitner,Anne Baumgrass,Sigrid Schefer-Wenzl,Stefanie Rinderle-Ma,Mark Strembeck	Business Process Management Workshops
A comprehensive investigation of the applicability of process mining techniques for enterprise risk management	Filip Caron,Jan Vanthienen,Bart Baesens	Computers in Industry
A family of case studies on business process mining using MARBLE	Ricardo Perez-Castillo,Jose A Cruz-Lemus,Ignacio Garcia-Rodriguez de Guzman,Mario Piattini	The Journal of Systems and Software
A genetic programming approach to business process mining	C. J. Turner,Ashutosh Tiwari,Jorn Mehnen	Genetic and evolutionary computation
A GP Process Mining Approach from a Structural Perspective	Anhua Wang,Weidong Zhao,Chongchen Chen,Haifeng Wu	Artificial Intelligence and Computational Intelligence
A Grid-Based Multi-relational Approach to Process Mining	Antonio Turi,Annalisa Appice,Michelangelo Ceci,Donato Malerba	Database and Expert Systems Applications
A Heuristic Genetic Process Mining Algorithm	Jiafei Li,Jihong Ouyang,Mingyong Feng	Computational Intelligence and Security
A Hybrid Approach for Dynamic Business Process Mining Based On Reconfigurable Nets and Event Types	Ning Li,Jianchu Kang,Weifeng Lv	E-Business Engineering
A Hybrid Approach for Process Mining: Using From-to Chart Arranged by Genetic Algorithms	Eren Esgin,Pinar Senkul,Cem Cimenbicer	Hybrid Artificial Intelligence Systems
A Hybrid Approach to Process Mining: Finding Immediate Successors of a Process by Using From-To Chart	Eren Esgin,Pinar Senkul	International Conference on Machine Learning and Applications
A Method of Adaptive Process Mining Based on Time-Varying Sliding Window and Relation of Adjacent Event Dependency	Mei-Hong Shi,Shou-Shan Jang,Yong-Gang Guo,Liang Chen,Kai-Duan Cao	Intelligent System Design and Engineering Application
A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs	Jochen De Weerdt,Manu De Backer,Jan Vanthienen,Bart Baesens	Information Systems

Title	Authors	Publication Outlet
A New Method for Business Process Mining Based on State Equation	Hua Hu,Jianen Xie,Hai-yang Hu	Web Information Systems Engineering – WISE 2010 Workshops
A New Process Mining Algorithm Based on Event Type	Dongyi Wang,Jidong Ge,Hao Hu,Bin Luo	Dependable, Autonomic and Secure Computing
A New Process Mining Algorithm of Workflow	Shenghui He,Tao Lv,Binggui Huang	Industrial and Information Systems
A novel approach for process mining based on event types	Lijie Wen,Jianmin Wang,Wil M. P. van der Aalst,Biqing Huang,Jiaguang Sun	Journal of Intelligent Information Systems
A Novel Approach of Process Mining with Event Graph	Hui Zhang,Ying Liu,Chunping Li,Roger Jiao	Knowledge-Based and Intelligent Information and Engineering Systems
A policy-based process mining framework: mining business policy texts for discovering process models	Jiexun Li,Harry Jiannan Wang,Zhu Zhang,J. Leon Zhao	Information Systems and eBusiness Management
A Principled Approach to the Analysis of Process Mining Algorithms	Phil Weber,Behzad Bordbar,Peter Tino	Intelligent Data Engineering and Automated Learning - IDEAL 2011
A Process Mining Approach to Redesign Business Processes - A Case Study in Gas Industry	N. R. T. P. van Beest,Laura Maruster	Symposium on Symbolic and Numeric Algorithms for Scientific Computing
A process mining based approach to knowledge maintenance	Ming Li,Lu Liu,Lu Yin,Yanqiu Zhu	Information Systems Frontiers
A process-mining framework for the detection of healthcare fraud and abuse	Wan-Shiou Yang,San-Yih Hwang	Expert Systems with Applications
A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare	Ronny S. Mans,Hajo Reijers,Daniel Wismeijer,Michiel van Genuchten	Information Systems
A review of business process mining: state-of-the-art and future trends	Ashutosh Tiwari,C. J. Turner,Basim Majeed	Business Process Management Journal
A Study of Quality and Accuracy Trade-offs in Process Mining	Zan Huang,Akhil Kumar	INFORMS Journal on Computing
A Workflow Process Mining Algorithm Based on Synchro-Net	Xing-Qi Huang,Li-Fu Wang,Wen Zhao,Shi-Kun Zhang,Chong-Yi Yuan	Journal of Computer Science and Technology
Abstractions in Process Mining: A Taxonomy of Patterns	R. P. Jagadeesh Chandra Bose,Wil M. P. van der Aalst	Business Process Management
Agent-Based Analysis and Detection of Functional Faults of Vehicle Industry Processes: A Process Mining Approach	Agnes Werner-Stark,Tibor Dulai	Agent and Multi-Agent Systems. Technologies and Applications
Algorithms for anomaly detection of traces in logs of process aware information systems	Fabio Bezerra,Jacques Wainer	Information Systems

Title	Authors	Publication Outlet
An Effective Algorithm for Business Process Mining Based on Modified FP-Tree Algorithm	Gun-Woo Kim,Seung Hoon Lee,Jae Hyung Kim,Jin Hyun Son	International Conference on Communication Software and Networks
An empirical comparison of static and dynamic business process mining	Ricardo Pérez-Castillo,Ignacio Garcia-Rodriguez de Guzman,Mario Piattini,Barbara Weber,Angeles S. Places	Symposium on Applied Computing
An empirical evaluation of process mining algorithms based on structural and behavioral similarities	Jianmin Wang,Shijie Tan,Lijie Wen,Raymond K. Wong,Qinlong Guo	Symposium on Applied Computing
An experimental evaluation of genetic process mining	C. J. Turner,Ashutosh Tiwari	Genetic and evolutionary computation
An Exploration of Genetic Process Mining	C. J. Turner,Ashutosh Tiwari	Applications of Soft Computing
An Extended CompDepend-Algorithm to Discovery Duplicate Tasks in Workflow Process Mining System	Wang Xiaohui,Zhao Wenqing,Guo Fengjuan	International Conference on Research Challenges in Computer Science
An improved simulated annealing algorithm for process mining	Dianfang Gao,Qiang Liu	Conference on Computer Supported Cooperative Work in Design
An Incremental Process Mining Approach to Extract Knowledge from Legacy Systems	Andre Cristiano Kalsing,Gleison Samuel do Nascimento,Cirano Lochpe,Lucineia Heloisa Thom	Enterprise Distributed Object Computing Conference
An iterative approach to synthesize business process templates from compliance rules	Ahmed Awad,Rajeev Gore,Zhe Hou,James Thomson,Matthias Weidlich	Information Systems
An Outlook on Semantic Business Process Mining and Monitoring	A. K. A. de Medeiros,C. Pedrinaci,Wil M. P. van der Aalst,J. Domingue,Minseok Song,Anne Rozinat,B. Norton,L. Cabral	On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops
Analysis of Multi-Agent Interactions with Process Mining Techniques	Lawrence Cabac,Nicolas Knaak,Daniel Moldt,Heiko Rölke	Multiagent System Technologies
Analyzing Multi-agent Activity Logs Using Process Mining Techniques	Anne Rozinat,Stefan Zickler,Manuela Veloso,Wil M. P. van der Aalst,Colin McMillen	Distributed Autonomous Robotic Systems 8
Analyzing Resource Behavior Using Process Mining	Joyce Nakatumba,Wil M. P. van der Aalst	Business Process Management Workshops
Analyzing Vessel Behavior Using Process Mining	Fabrizio M. Maggi,Arjan J. Mooij,Wil M. P. van der Aalst	Situation Awareness with Systems of Systems

Title	Authors	Publication Outlet
Anomaly Detection Using Process Mining	Fabio Bezerra,Jacques Wainer,Wil M. P. van der Aalst	Enterprise, Business-Process and Information Systems Modeling
Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital	Ronny S. Mans,M. H. Schonenberg,Minseok Song,Wil M. P. van der Aalst,P. J. M. Bakker	Biomedical Engineering Systems and Technologies
Applying Clustering in Process Mining to Find Different Versions of a Business Process That Changes over Time	Daniela Luengo,Marcos Sepulveda	Business Process Management Workshops
Applying Inductive Logic Programming to Process Mining	Evelina Lamma,Paola Mello,Fabrizio Riguzzi,Sergio Storari	Inductive Logic Programming
Applying Petri Net to Analyze a Multi-Agent System Feasibility - a Process Mining Approach	C. Ou-Yang,Yeh-Chun Juan,C. S. Li	Global Perspective for Competitive Enterprise, Economy and Ecology
Applying process mining approach to support the verification of a multi-agent system	C. Ou-Yang,Yeh-Chun Juan	Journal of Systems Science and Systems Engineering
Applying Process Mining in SOA Environments	Ateeq Khan,Azeem Lodhi,Veit Köppen,Gamal Kassem,Gunter Saake	Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops
Approaching Process Mining with Sequence Clustering: Experiments and Findings	Diogo R. Ferreira,Marielba Zacarias,Miguel Malheiros,Pedro Ferreira	Business Process Management
Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor	Wil M. P. van der Aalst,Kees M. van Hee,Jan Martijn van der Werf,Marc Verdonk	IEEE Computer Society Press
Beyond Process Mining: From the Past to Present and Future	Wil M. P. van der Aalst,Maja Pesic,Minseok Song	Advanced Information Systems Engineering
Book Review: Process Mining: Discovery, Conformance and Enhancement of Business Processes	Vojtech Huser	Journal of Biomedical Informatics
Bridging Abstraction Layers in Process Mining by Automated Matching of Events and Activities	Thomas Baier,Jan Mendling	Business Process Management
Bridging Abstraction Layers in Process Mining: Event to Activity Mapping	Thomas Baier,Jan Mendling	Enterprise, Business-Process and Information Systems Modeling
Business alignment: using process mining as a tool for Delta analysis and conformance testing	Wil M. P. van der Aalst	Requirements Engineering
Business process analysis in healthcare environments: A methodology based on process mining	Alvaro Rebuge,Diogo R. Ferreira	Information Systems

Title	Authors	Publication Outlet
Business Process Mining and Reconstruction for Financial Audits	Michael Werner,Nick Gehrke,Markus Nüttgens	Hawaii International Conference on System Sciences
Business Process Mining and Rules Detection for Unstructured Information	Dafne A. Rosso-Pelayo,Raul A. Trejo-Ramirez,Miguel Gonzalez-Mendoza,Neil Hernandez-Gress	Mexican International Conference on Artificial Intelligence
Business Process Mining Based on Simulated Annealing	Wei Song,ShaoZhuo Liu,Qiang Liu	International Conference for Young Computer
Business Process Mining by Means of Statistical Languages Model	Dafne A. Rosso-Pelayo,Raul A. Trejo-Ramirez	Mexican International Conference on Artificial Intelligence
Business Process Mining from E-Commerce Web Logs	Nicolas Poggi,Vinod Muthusamy,David Carrera,Rania Khalaf	Business Process Management
Business process mining from group stories	Joao Carlos de A. R. Goncalves,Flavia Maria Santoro,Fernanda Araujo Baiao	Conference on Computer Supported Cooperative Work in Design
Business process mining: An industrial application	Wil M. P. van der Aalst,Hajo Reijers,A. J. M. M. Weijters,B. F. van Dongen,A. K. A. de Medeiros,Minseok Song,H. M. W. Verbeek	Information Systems
Business Process Workarounds: What Can and Cannot Be Detected by Process Mining	Nesi Outmazgin,Pnina Soffer	Enterprise, Business-Process and Information Systems Modeling
Case of Process Mining from Business Execution Log Data	Joonsoo Bae,Young Ki Kang	Intelligent Decision Technologies
Case Study in Process Mining in a Multinational Enterprise	Paul Taylor,Marcello Leida,Basim Majeed	Data-Driven Process Discovery and Analysis
Classification and evaluation of timed running schemas for workflow based on process mining	Hua Duan,Qingtian Zeng,Huaiqing Wang,Sherry X. Sun,Dongming Xu	The Journal of Systems and Software
Clustering and Operation Analysis for Assembly Blocks Using Process Mining in Shipbuilding Industry	Dongha Lee,Jaehun Park,Iq Reviessay Pulshashi,Hyerim Bae	Asia Pacific Business Process Management
Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach	Santiago Aguirre,Carlos Parra,Jorge Alvarado	Data-Driven Process Discovery and Analysis
Combining Process Mining and Statistical Methods to Evaluate Customer Integration in Service Processes	Michael Leyer,Jürgen Moormann	Business Process Management Workshops

Title	Authors	Publication Outlet
Comprehensive rule-based compliance checking and risk management with process mining	Filip Caron,Jan Vanthienen,Bart Baesens	Decision Support Systems
Configurable Services in the Cloud: Supporting Variability While Enabling Cross-Organizational Process Mining	Wil M. P. van der Aalst	On the Move to Meaningful Internet Systems: OTM 2010
Conformance checking of processes based on monitoring real behavior	Anne Rozinat,Wil M. P. van der Aalst	Information Systems
Consistent Process Mining over Big Data Triple Stores	Antonia Azzini,Paolo Ceravolo	Congress on Big Data
Context-aware process mining framework for business process flexibility	Zerari Mounira,Boufaida Mahmoud	Information Integration and Web-based Applications & Services
Continuous Quality Improvement of IT Processes based on Reference Models and Process Mining	Kerstin Gerke,Gerrit Tamm	AMCIS 2009 Proceedings
Coupling Case Based Reasoning and Process Mining for a Web Based Crisis Management Decision Support System	Sameh Triki,Narjes Bellamine Ben Saoud,Julie Dugdale,Chihab Hanachi	Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises
Data and Process Mining: Minitrack Introduction	Selwyn Piramuthu,H. Michael Chung	Hawaii International Conference on System Sciences
Data Flow-Oriented Process Mining to Support Security Audits	Thomas Stocker	Service-Oriented Computing - ICSOC 2011 Workshops
Data Improvement to Enable Process Mining on Integrated Non-log Data Sources	Reinhold Dunkl	Computer Aided Systems Theory - EUROCAST 2013
Data Mining and Process Mining: Business Impact and Application Challenges	H. Michael Chung,Selwyn Piramuthu	Hawaii International Conference on System Sciences
Data Transformation and Semantic Log Purging for Process Mining	Linh Thao Ly,Conrad Indiono,Jürgen Mangler,Stefanie Rinderle-Ma	Advanced Information Systems Engineering
Data-aware process mining: discovering decisions in processes using alignments	Massimiliano de Leoni,Wil M. P. van der Aalst	Symposium on Applied Computing
Decision Support Based on Process Mining	Wil M. P. van der Aalst	Handbook on Decision Support Systems 1
Decision-Process Mining: A New Framework for Knowledge Acquisition of Business Decision-making Processes	Razvan Petrusel	Studia Universitatis Babes-Bolyai
Declarative Process Mining	Marco Montali	Specification and Verification of Declarative Open Interaction Models
Decomposing Petri nets for process mining: A generic approach	Wil M. P. van der Aalst	Distributed and Parallel Databases

Title	Authors	Publication Outlet
Decomposing Process Mining Problems Using Passages	Wil M. P. van der Aalst	Application and Theory of Petri Nets
Definition and Validation of Process Mining Use Cases	Irina Ailenei,Anne Rozinat,Albert Eckert,Wil M. P. van der Aalst	Business Process Management Workshops
Deriving Event Graphs through Process Mining for Runtime Change Management	Hui Zhang,Ying Liu,Chunping Li,Roger Jiao	Modelling and Management of Engineering Processes
Design and Implementation of Process Mining System Based on a-Algorithm	Yong Ling,Liqun Zhang,Meng Gao	International Conference on Innovative Computing Information and Control
Development of a co-operative distributed process mining system for quality assurance	G. T. S. Ho,H. C. W. Lau,S. K. Kwok,C. K. M. Lee,W. Ho	International Journal of Production Research
Development of a process mining system for supporting knowledge discovery in a supply chain network	H. C. W. Lau,G. T. S. Ho,Y. Zhao,N. S. H. Chung	International Journal of Production Economics
Development of an OLAP-Fuzzy Based Process Mining System for Quality Improvement	G. T. S. Ho,H. C. W. Lau	Intelligent Information Processing III
Development of Distance Measures for Process Mining, Discovery, and Integration	Joonsoo Bae,Ling Liu,James Caverlee,Liang-Jie Zhang,Hyerim Bae	International Journal of Web Services Research
Discovering Business Rules through Process Mining	Raphael Crerie,Fernanda Araujo Baiao,Flavia Maria Santoro	Enterprise, Business-Process and Information Systems Modeling
Discovering Changes of the Change Control Board Process during a Software Development Project Using Process Mining	Jana Samalikova,Jos J. M. Trienekens,Rob J. Kusters,A. J. M. M. Weijters	Software Process Improvement
Discovering simulation models	Anne Rozinat,Ronny S. Mans,Minseok Song,Wil M. P. van der Aalst	Information Systems
Discovery and analysis of e-mail-driven business processes	Marco Stuit,Hans Wortmann	Information Systems
Distributed Genetic Process Mining Using Sampling	Carmen Bratosin,Natalia Sidorova,Wil M. P. van der Aalst	Parallel Computing Technologies
Divide-and-Conquer Strategies for Process Mining	Josep Carmona,Jordi Cortadella,Michael Kishinevsky	Business Process Management
Does Process Mining Add to Internal Auditing? An Experience Report	Mieke J. Jans,Benoit Depaire,Koen Vanhoof	Enterprise, Business-Process and Information Systems Modeling
DPMine/P: modeling and process mining language and ProM plug-ins	Sergey A. Shershakov	Software Engineering Conference in Russia
Dynamic context-aware Business Process flexibility: an artefact-based approach using process mining	Mounira Zerari,Mahmoud Boufaida	International Journal of Business Intelligence and Data Mining

Title	Authors	Publication Outlet
EMiT: A Process Mining Tool	B. F. van Dongen,Wil M. P. van der Aalst	Applications and Theory of Petri Nets 2004
Exploiting Inductive Logic Programming Techniques for Declarative Process Mining	Federico Chesani,Evelina Lamma,Paola Mello,Marco Montali,Fabrizio Riguzzi,Sergio Storari	Transactions on Petri Nets and Other Models of Concurrency II
Exploring regulatory processes during a computer-supported collaborative learning task using process mining	Cornelia Schoor,Maria Bannert	Computers in Human Behavior
Exploring the CSCW spectrum using process mining	Wil M. P. van der Aalst	Advanced Engineering Informatics
Finding Structure in Unstructured Processes: The Case for Process Mining	Wil M. P. van der Aalst,C. W. Günther	Application of Concurrency to System Design
From Local Patterns to Global Models: Towards Domain Driven Educational Process Mining	Nikola Trcka,Mykola Pechenizkiy	Intelligent Systems Design and Applications
Generating event logs from non-process-aware systems enabling business process mining	Ricardo Perez-Castillo,Barbara Weber,Jakob Pinggera,Stefan Zugal,Ignacio Garcia-Rodriguez de Guzman,Mario Piattini	Enterprise Information Systems
Genetic Process Mining	Wil M. P. van der Aalst,A. K. A. de Medeiros,A. J. M. M. Weijters	Applications and Theory of Petri Nets 2005
Genetic Process Mining: A Basic Approach and Its Challenges	A. K. A. de Medeiros,A. J. M. M. Weijters,Wil M. P. van der Aalst	Business Process Management Workshops
Genetic process mining: an experimental evaluation	A. K. A. de Medeiros,A. J. M. M. Weijters,Wil M. P. van der Aalst	Data Mining and Knowledge Discovery
Getting a Grasp on Clinical Pathway Data: An Approach Based on Process Mining	Jochen De Weerdt,Filip Caron,Jan Vanthienen,Bart Baesens	Emerging Trends in Knowledge Discovery and Data Mining
Goal-Heuristic Analysis Method for an Adaptive Process Mining	Su-Jin Baek,Jong-Won Ko,Gui-Jung Kim,Jung-Soo Han,Young-Jae Song	Proceedings of the International Conference on IT Convergence and Security 2011
Handling Concept Drift in Process Mining	R. P. Jagadeesh Chandra Bose,Wil M. P. van der Aalst,Indre Zliobaite,Mykola Pechenizkiy	Advanced Information Systems Engineering
Healthcare Process Mining with RFID	Wei Zhou,Selwyn Piramuthu	Business Process Management Workshops

Title	Authors	Publication Outlet
Impact of Data Aggregation on the Significance of Process Mining Results: An Experimental Evaluation	Kerstin Gerke	Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises
Improving Business Process Models with Agent-Based Simulation and Process Mining	Fernando Szimanski,Celia G. Ralha,Gerd Wagner,Diogo R. Ferreira	Enterprise, Business-Process and Information Systems Modeling
Improving process models by discovering decision points	Sharmila Subramaniam,Vana Kalogeraki,Dimitrios Gunopulos,Fabio Casati,Malu Castellanos,Umeshwar Dayal,Mehmet Sayal	Information Systems
Incremental Declarative Process Mining	Massimiliano Cattafi,Evelina Lamma,Fabrizio Riguzzi,Sergio Storari	Smart Information and Knowledge Management
Insuring Sensitive Processes through Process Mining	Jorge Munoz-Gama,Isao Echizen	Ubiquitous Intelligence and Computing
Integrating Computer Log Files for Process Mining: A Genetic Algorithm Inspired Technique	Jan Claes,Geert Poels	Advanced Information Systems Engineering Workshops
Intra- and Inter-Organizational Process Mining: Discovering Processes within and between Organizations	Wil M. P. van der Aalst	The Practice of Enterprise Modeling
Leveraging Process-Mining Techniques	Geetika T. Lakshmanan,Rania Khalaf	IT Professional Magazine
Logic-Based Incremental Process Mining in Smart Environments	Stefano Ferilli,Berardina De Carolis,Domenico Redavid	Recent Trends in Applied Artificial Intelligence
Merging Computer Log Files for Process Mining: An Artificial Immune System Technique	Jan Claes,Geert Poels	Business Process Management Workshops
Model-based Business Process Mining	Jon Espen Ingvaldsen,Jon Atle Gulla	Information Systems Management
Multidimensional process mining: a flexible analysis approach for health services research	Thomas Vogelgesang,H.-Jürgen Appelrath	EDBT/ICDT 2013 Workshops
Multi-phase Process Mining: Building Instance Graphs	B. F. van Dongen,Wil M. P. van der Aalst	Conceptual Modeling – ER 2004
Net Components for the Integration of Process Mining into Agent-Oriented Software Engineering	Lawrence Cabac,Nicolas Denz	Transactions on Petri Nets and Other Models of Concurrency I
New methodology for modeling large scale manufacturing process: Using process mining methods and experts' knowledge	Pamela Viale,Claudia Frydman,Jacques Pinaton	International Conference on Computer Systems and Applications

Title	Authors	Publication Outlet
Obtaining Thresholds for the Effectiveness of Business Process Mining	Ricardo Perez-Castillo,Laura Sanchez-Gonzalez,Mario Piattini,Felix Garcia,Ignacio Garcia-Rodriguez de Guzman	Empirical Software Engineering and Measurement
On Recommendation of Process Mining Algorithms	Jianmin Wang,Raymond K. Wong,Jianwei Ding,Qinlong Guo,Lijie Wen	International Conference on Web Services
On the exploitation of process mining for security audits: the conformance checking case	Rafael Accorsi,Thomas Stocker	Symposium on Applied Computing
On the exploitation of process mining for security audits: the process discovery case	Rafael Accorsi,Thomas Stocker,Günter Müller	Symposium on Applied Computing
On the Representational Bias in Process Mining	Wil M. P. van der Aalst	Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises
Online Techniques for Dealing with Concept Drift in Process Mining	Josep Carmona,Ricard Gavalda	Advances in Intelligent Data Analysis XI
Ontological approach to enhance results of business process mining and analysis	Wirat Jareevongpiboon,Paul Janecek	Business Process Management Journal
Outlier Detection Techniques for Process Mining Applications	Lucantonio Ghionna,Gianluigi Greco,Antonella Guzzo,Luigi Pontieri	Foundations of Intelligent Systems
Performance Analysis of GRID Middleware Using Process Mining	Anastas Misev,Emanouil Atanassov	Computational Science – ICCS 2008
Petri-net integration – An approach to support multi-agent process mining	C. Ou-Yang,H. Winarjo	Expert Systems with Applications
Preprocessing Support for Large Scale Process Mining of SAP Transactions	Jon Espen Ingvaldsen,Jon Atle Gulla	Business Process Management Workshops
Probabilistic Declarative Process Mining	Elena Bellodi,Fabrizio Riguzzi,Evelina Lamma	Knowledge Science, Engineering and Management
Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining	Wil M. P. van der Aalst	Asia Pacific Business Process Management
Process diagnostics using trace alignment: Opportunities, issues, and challenges	R. P. Jagadeesh Chandra Bose,Wil M. P. van der Aalst	Information Systems
Process Diagnostics: A Method Based on Process Mining	Melike Bozkaya,Joost Gabriels,Jan Martijn van der Werf	Information, Process, and Knowledge Management
Process Mining	Wil M. P. van der Aalst	Communications of the ACM

Title	Authors	Publication Outlet
Process Mining	Rafael Accorsi,Meike Ullrich,Wil M. P. van der Aalst	Informatik-Spektrum
Process Mining and Petri Net Synthesis	Ekkart Kindler,Vladimir Rubin,Wilhelm Schäfer	Business Process Management Workshops
Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance	Wil M. P. van der Aalst,A. K. A. de Medeiros	Electronic Notes in Theoretical Computer Science
Process Mining and Security: Visualization in Database Intrusion Detection	Viet H. Huynh,An N. T. Le	Intelligence and Security Informatics
Process Mining and Simulation	Moe Wynn,Anne Rozinat,Wil M. P. van der Aalst,Arthur ter Hofstede,Colin Fidge	Modern Business Process Automation
Process Mining and the ProM Framework: An Exploratory Survey	Jan Claes,Geert Poels	Business Process Management Workshops
Process Mining and Verification of Properties: An Approach Based on Temporal Logic	Wil M. P. van der Aalst,H. T. de Beer,B. F. van Dongen	On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE
Process Mining Applied to the BPI Challenge 2012: Divide and Conquer While Discerning Resources	R. P. Jagadeesh Chandra Bose,Wil M. P. van der Aalst	Business Process Management Workshops
Process mining applied to the test process of wafer scanners in ASML	Anne Rozinat,I. S. M. De Jong,C. W. Günther,Wil M. P. van der Aalst	IEEE Transactions on Systems
Process Mining Approach for Traffic Analysis in Wireless Mesh Networks	Kirill Krinkin,Eugene Kalishenko,S. P. Shiva Prakash	Internet of Things, Smart Spaces, and Next Generation Networking
Process Mining Approach to Promote Business Intelligence in Iranian Detectives' Police	Mehdi Ghazanfari,Mohammad Fathian,Mostafa Jafari,Saeed Rouhani	Networked Digital Technologies
Process Mining as First-Order Classification Learning on Logs with Negative Events	Stijn Goedertier,David Martens,Bart Baesens,Raf Haesen,Jan Vanthienen	Business Process Management Workshops
Process Mining Based on Clustering: A Quest for Precision	A. K. A. de Medeiros,Antonella Guzzo,Gianluigi Greco,Wil M. P. van der Aalst,A. J. M. M. Weijters, B. F. van Dongen,Domenico Sacca	Business Process Management Workshops
Process Mining Based on Regions of Languages	Robin Bergenthum,Jörg Desel,Robert Lorenz,Sebastian Mauser	Business Process Management

Title	Authors	Publication Outlet
Process Mining Based on Specification Slicing for Dynamic Reconfiguration	Su-Jin Baek,Jinhyang Kim,Young-Jae Song	International Conference on Computational Science and Its Applications
Process Mining by Measuring Process Block Similarity	Joonsoo Bae,James Caverlee,Ling Liu,Hua Yan	Business Process Management Workshops
Process Mining for Electronic Data Interchange	Robert Engel,Worarat Krathu,Marco Zapletal,Christian Pichler,Wil M. P. van der Aalst, Hannes Werthner	E-Commerce and Web Technologies
Process Mining for Job Nets in Integrated Enterprise Systems	Shinji Kikuchi,Yasuhide Matsumoto,Motomitsu Adachi,Shingo Moritomo	Enterprise Information Systems
Process Mining for the multi-faceted analysis of business processes - A case study in a financial services organization	Jochen De Weerdt,Annelies Schupp,An Vanderloock,Bart Baesens	Computers in Industry
Process Mining for Ubiquitous Mobile Systems: An Overview and a Concrete Algorithm	A. K. A. de Medeiros,B. F. van Dongen,Wil M. P. van der Aalst,A. J. M. M. Weijters	Ubiquitous Mobile Information and Collaboration Systems
Process Mining Framework for Software Processes	Vladimir Rubin,C. W. Günther,Wil M. P. van der Aalst,Ekkart Kindler,B. F. van Dongen,Wilhelm Schäfer	Software Process Dynamics and Agility
Process Mining from a Basis of State Regions	Marc Sole,Josep Carmona	Applications and Theory of Petri Nets
Process Mining in Auditing: From Current Limitations to Future Challenges	Mieke J. Jans	Business Process Management Workshops
Process Mining in Healthcare: A Contribution to Change the Culture of Blame	Silvana Quaglini	Business Process Management Workshops
Process Mining in Healthcare: Data Challenges When Answering Frequently Posed Questions	Ronny S. Mans,Wil M. P. van der Aalst,Rob J. B. Vanwersch,Arnold J. Moleman	Process Support and Knowledge Representation in Health Care
Process Mining Manifesto	Wil M. P. van der Aalst,and other	Business Process Management Workshops
Process Mining Meets Abstract Interpretation	Josep Carmona,Jordi Cortadella	Machine Learning and Knowledge Discovery in Databases
Process Mining of RFID-Based Supply Chains	Kerstin Gerke,Alexander Claus,Jan Mendling	Commerce and Enterprise Computing
Process Mining Put into Context	Wil M. P. van der Aalst,Schahram Dustdar	IEEE Internet Computing
Process Mining Software Repositories	Wouter Poncin,Alexander Serebrenik,Mark van den Brand	Conference on Software Maintenance and Reengineering

Title	Authors	Publication Outlet
Process Mining Techniques in Conformance Testing of Inventory Processes: An Industrial Application	Zbigniew Paszkiewicz	Business Information Systems Workshops
Process Mining towards Semantics	A. K. A. de Medeiros,Wil M. P. van der Aalst	Advances in Web Semantics I
Process Mining Versus Intention Mining	Ghazaleh Khodabandelou,Charlotte Hug,Rebecca Deneckere,Camille Salinesi	Enterprise, Business-Process and Information Systems Modeling
Process Mining, Discovery, and Integration using Distance Measures	Joonsoo Bae,Ling Liu,James Caverlee,William B. Rouse	Conference on Web Services
Process Mining: A Block-Structured Mining Approach	Yan-Liang Qu,Tie-Shi Zhao	Communication Systems and Information Technology
Process mining: a research agenda	Wil M. P. van der Aalst,A. J. M. M. Weijters	Computers in Industry
Process mining: a two-step approach to balance between underfitting and overfitting	Wil M. P. van der Aalst,Vladimir Rubin,H. M. W. Verbeek	Software & Systems Modeling
Process Mining: Algorithm for S-Coverable Workflow Nets	Jianchun She,Dongqing Yang	Workshop on Knowledge Discovery and Data Mining
Process Mining: Discovering Direct Successors in Process Logs	Laura Maruster,A. J. M. M. Weijters,Wil M. P. van der Aalst,Antal van den Bosch	Discovery Science
Process Mining: Extending α-Algorithm to Mine Duplicate Tasks in Process Logs	Jiafei Li,Dayou Liu,Bo Yang	Advances in Web and Network Technologies, and Information Management
Process mining: from theory to practice	C. J. Turner,Ashutosh Tiwari,Richard Olaiya,Yuchun Xu	Business Process Management Journal
Process Mining: Fuzzy Clustering and Performance Visualization	B. F. van Dongen,Arya Adriansyah	Business Process Management Workshops
Process Mining: Overview and Outlook of Petri Net Discovery Algorithms	B. F. van Dongen,A. K. A. de Medeiros,L. Wen	Transactions on Petri Nets and Other Models of Concurrency II
Process Mining-Driven Optimization of a Consumer Loan Approvals Process	Arjel D. Bautista,Lalit Wangikar,Syed M. Kumail Akbar	Business Process Management Workshops
Process-Aware Information Systems: Lessons to Be Learned from Process Mining	Wil M. P. van der Aalst	Transactions on Petri Nets and Other Models of Concurrency II
Process-Mining-Based Workflow Model Fragmentation for Distributed Execution	Sherry X. Sun,Qingtian Zeng,Huaiqing Wang	IEEE Transactions on Systems

Title	Authors	Publication Outlet
Projection approaches to process mining using region-based techniques	Josep Carmona	Data Mining and Knowledge Discovery
Reality mining via process mining	O. M. Hassan,M. S. Farag,M. M. MohieEl-Din	Applied Informatics and Communications
Redesigning business processes: a methodology based on simulation andprocess mining techniques	Laura Maruster,N. R. T. P. van Beest	Knowledge and Information Systems
Regelbasierte Steuerung von Geschäftsprozessen – Konzeption eines Ansatzes auf Basis von Process Mining	Heinz Lothar Grob,Frank Bensberg,Andre Coners	WIRTSCHAFTSINFORMATIK
Relation-Centric Task Identification for Policy-Based Process Mining	Jiexun Li,Harry Jiannan Wang,Zhu Zhang,J. Leon Zhao	ICIS 2008 Proceedings
Requirements towards Effective Process Mining	Matthias Lohrmann,Alexander Riedel	On the Move to Meaningful Internet Systems: OTM 2012 Workshops
Rule-Based Business Process Mining: Applications for Management	Filip Caron,Jan Vanthienen,Bart Baesens	Management Intelligent Systems
Sequence partitioning for process mining with unlabeled event logs	Michal Walicki,Diogo R. Ferreira	Data & Knowledge Engineering
Similarity-based behavior and process mining of medical practices	Shusaku Tsumoto,Haruko Iwata,Shoji Hirano,Yuko Tsumoto	Future Generation Computer Systems
Simplifying discovered process models in a controlled manner	Dirk Fahland,Wil M. P. van der Aalst	Information Systems
Skeletal Algorithms in Process Mining	Michal R. Przybylek	Computational Intelligence
Source Code Partitioning Using Process Mining	Koki Kato,Tsuyoshi Kanai,Sanya Uehara	Business Process Management
The case for process mining in auditing: Sources of value added and areas of application	Mieke J. Jans,Michael Alles,Miklos Vasarhelyi	International Journal of Accounting Information Systems
The Need for a Process Mining Evaluation Framework in Research and Practice	Anne Rozinat,A. K. A. de Medeiros,C. W. Günther,A. J. M. M. Weijters,Wil M. P. van der Aalst	Business Process Management Workshops
The Process Mining Manifesto—An interview with Wil van der Aalst	Gottfried Vossen	Information Systems
The ProM Framework: A New Era in Process Mining Tool Support	B. F. van Dongen,A. K. A. de Medeiros,H. M. W. Verbeek,A. J. M. M. Weijters,Wil M. P. van der Aalst	Applications and Theory of Petri Nets 2005
The Research of Process Mining Assessment Used in Business Intelligence	Zhenyu Wang,Qing Yao,Yuqing Sun	International Conference on Computer and Information Science

Title	Authors	Publication Outlet
The Research on the Usage of Business Process Mining in the Implementation of BPR	Xie Yi-Wu,Li Xiao-Wan,Chen Yan	Network and Parallel Computing
Time prediction based on process mining	Wil M. P. van der Aalst,M. H. Schonenberg,Minseok Song	Information Systems
Time-interval process model discovery and validation--a genetic process mining approach	Chieh-Yuan Tsai,Henyi Jen,Yi-Ching Chen	Applied Intelligence
Towards Cross-Organizational Process Mining in Collections of Process Models and Their Executions	J. C. A. M. Buijs,B. F. van Dongen,Wil M. P. van der Aalst	Business Process Management Workshops
Towards Improving the Representational Bias of Process Mining	Wil M. P. van der Aalst,J. C. A. M. Buijs,B. F. van Dongen	Data-Driven Process Discovery and Analysis
Trace Alignment in Process Mining: Opportunities for Process Diagnostics	R. P. Jagadeesh Chandra Bose,Wil M. P. van der Aalst	Business Process Management
Trace Clustering in Process Mining	Minseok Song,C. W. Günther,Wil M. P. van der Aalst	Business Process Management Workshops
Translating Message Sequence Charts to other Process Languages Using Process Mining	Kristian Bisgaard Lassen,B. F. van Dongen	Transactions on Petri Nets and Other Models of Concurrency I
Using classification methods to label tasks in process mining	Scott Buffett,Liqiang Geng	Journal of Software Maintenance and Evolution
Using Genetic Process Mining Technology to Construct a Time-Interval Process Model	Chieh-Yuan Tsai,I-Ching Chen	Next-Generation Applied Intelligence
Using Mapreduce to Scale Events Correlation Discovery for Business Processes Mining	Hicham Reguieg,Farouk Toumani,Hamid Reza Motahari-Nezhad,Boualem Benatallah	Business Process Management
Using minimum description length for process mining	Toon Calders,C. W. Günther,Mykola Pechenizkiy,Anne Rozinat	Symposium on Applied Computing
Using process mining metrics to measure noisy process fidelity	Chris Thomson,Marian Gheorghe	Evaluation and Assessment in Software Engineering
Using Process Mining to Bridge the Gap between BI and BPM	Wil M. P. van der Aalst	Computer
Using process mining to business process distribution	Faramarz Safi Esfahani,Masrah Azrifah Azmi Murad,Md. Nasir Sulaiman,Nur Izura Udzir	Symposium on Applied Computing

Title	Authors	Publication Outlet
Using Process Mining to Generate Accurate and Interactive Business Process Maps	Wil M. P. van der Aalst	Business Information Systems Workshops
Using process mining to identify coordination patterns in IT service management	Theresa M. Edgington,T. S. Raghu,Ajay S. Vinze	Decision Support Systems
Using process mining to identify models of group decision making in chat data	Peter Reimann,Jimmy Frerejean,Kate Thompson	Computer Supported Collaborative Learning
When Process Mining Meets Bioinformatics	R. P. Jagadeesh Chandra Bose,Wil M. P. van der Aalst	IS Olympics: Information Systems in a Diverse World
Workflow simulation for operational decision support using event graph through process mining	Ying Liu,Hui Zhang,Chunping Li,Roger Jiao	Decision Support Systems

12 Appendix C: Final Declaration

Hiermit erkläre ich,

Michael Werner, geboren am 10. Februar 1979,

an Eides statt, dass ich die Dissertation mit dem Titel:

„Business Process Analysis Automation for Financial Audits - A design science-oriented approach to support internal and external auditors in process audits by using process mining techniques “

selbständig und ohne fremde Hilfe verfasst habe.

Andere als die von mir angegebenen Quellen und Hilfsmittel habe ich nicht benutzt. Die den herangezogenen Werken wörtlich oder sinngemäß entnommenen Stellen sind als solche gekennzeichnet.

Hamburg, den 7. Oktober 2014 ______________________________

Michael Werner