Process mining project methodology

developing a general approach to apply process mining in practice

van der Heijden, T.H.C.

Award date:
2012

Eindhoven, August 2012

Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice

By
T.H.C. VAN DER HEIJDEN

BSc Industrial Engineering — TU/e 2011
Student identity number 0611037

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
dr.ir. H.A. Reijers, TU/e, IS
dr.ir. A.J.M.M. Weijters, TU/e, IS
M. Tabbernee, Rabobank Nederland
B. van den Bergh, Rabobank Nederland

TUE. School of Industrial Engineering.
Series Master Theses Operations Management and Logistics

Subject headings: process management, business process analysis, process mining

Abstract

Process mining is a form of business process analysis based on recorded process data. Process mining techniques support organizations in retrieving structured process information using the logged events to discover, monitor and improve their processes. Currently, the process mining community is lacking a methodology that describes how to accomplish process mining in practice.

This research project describes an initial step towards the development of a comprehensive process mining project methodology in which different phases and main activities of business process mining projects are described and that can be used as an efficient and effective approach in order to apply process mining in practice. The methodology is developed using System Engineering Process and is validated by a case study at Rabobank Nederland. The methodology seems to be a valuable methodology for conducting process mining projects in practice.

Preface

This document describes a Master's Thesis which was partly performed at Rabobank Nederland. The project gave me the opportunity to get to know a lot more about doing research, process mining, the process mining community, myself and working life in a big company. Although this half a year was not always easy, it was definitely highly interesting, very instructive and absolutely rewarding.

First of all I would like to thank Hajo for all his support, suggestions, critical remarks and the good conversations. I want to thank Ton for his specific feedback on process mining that improved the quality of the research project. I am also very thankful to the Rabobank for giving me the opportunity to get to know the business. Martijn for his mentoring role and support, especially in order to get appropriate data for the process mining project. Ben for the discussions to make sense out of the process data and being a great 'roommate'. Frank for all his support in combining theory about process mining and practice. Sander for making me aware of how big companies work and how to get things done. Furthermore, many thanks to Anne and Christian of Fluxicon for giving me the opportunity to participate in the beta program of the great new process mining tool Disco and providing me with support in the development of the methodology.

Finally, my gratitude also goes out to my family, friends and all students that supported me during my study. In particular, my parents for all their support in giving me this chance and making me aware of the importance of studying, my friends of Meteoor, Group24 and E.S.C. that made my time as a student such a great period and, last but not least, my girlfriend Anne. Her love and support helped me to bring this project to a good end.

Tijn van der Heijden

Utrecht, August 2012

“Life is what we make it, always has been, always will be.”

- Grandma Moses

Summary

Process mining is a form of business process analysis based on recorded process data by information systems. The logs of these information systems contain information about historic events that took place during the process. Process mining techniques support organizations in retrieving structured process information using these logged events to discover, monitor and improve their processes.

Performing process mining projects in organizations requires several extra activities next to the actual application of the process mining techniques. Currently, the process mining community is lacking a methodology that describes how to accomplish process mining in practice. Besides that an appropriate process mining methodology will give practitioners guidance in applying process mining in organizations, it will also support in sharing best practices, stimulating the adoption of process mining in the field and preventing reinventing the wheel. This Master's Thesis aimed at developing an appropriate methodology that describes what is needed to accomplish process mining projects in organizations and how to execute these projects.

The methodology is developed using System Engineering Process, a framework for designing and managing complex engineering projects. The development consisted of four stages: 1. identifying the requirements, 2. identifying the main activities of a process mining project, 3. synthesizing all information and designing the methodology, 4. an evaluation of the proposed methodology.

During the identification of the requirements of the methodology, the scope of the methodology was created in terms of the conditions that must be applied and what the methodology must be able to deliver. Several sources were used to formulate the requirements of the methodology: knowledge gathered from scientists in the process mining field, process mining professionals, managers that facilitated process mining projects, scientific literature, summaries of process mining projects and hands on experience. The requirements were described using eight different tasks: 1. customer expectations, 2. project constraints, 3. external constraints 4. operational scenarios, 4. measure of effectiveness, 5. methodology boundaries, 6. life-cycle, 7. functional requirements, 8.performance requirements.

In the second stage of the project all functional requirements were further decomposed to low-level requirements which described the main activities that needed to be executed to accomplish a process mining project. This resulted in eighteen different activities divided up in six different phases.

All requirements were combined during the third stage, the design synthesis, which resulted in the Process Mining Project Life-cycle (PMPL), and the Process Mining Project Methodology (PMPM). PMPL, visualized in *figure 1*, presents an overview of the relationships between the different process mining project phases. The arrow description gives the output or input of the phases. In *figure 2*, an overview of all six phases and the activities that should be performed during each phase, is mentioned.

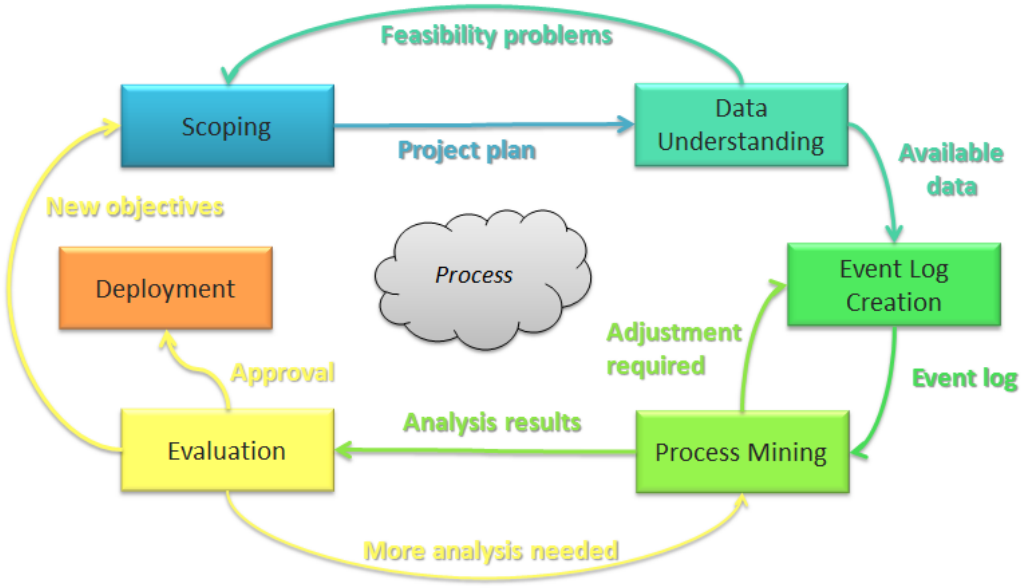


Figure 1, Life-cycle of process mining projects (PMPL)

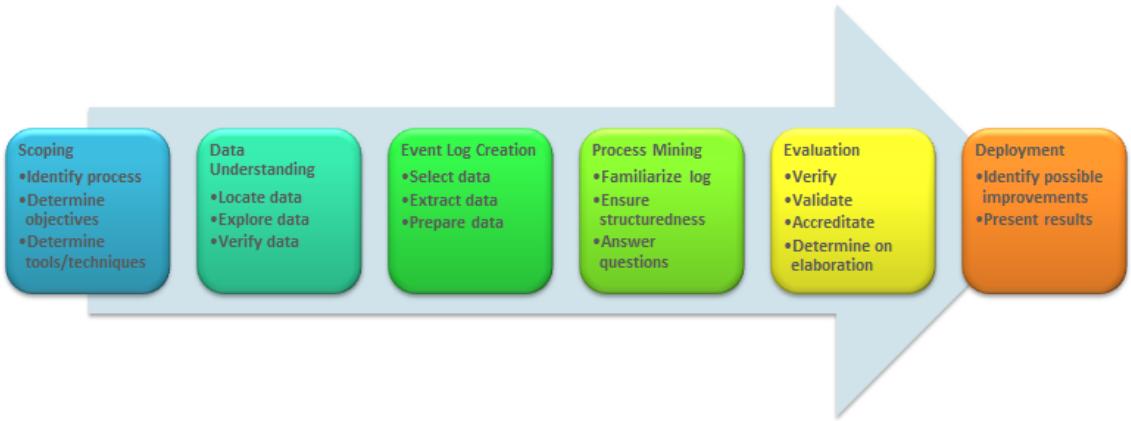


Figure 2, summary of Process Mining Project Methodology (PMPM)

The proposed methodology has been evaluated during the last part of the research project in a case study at the Financial Services department of Rabobank Nederland. The methodology gave useful support during the project in proposing activities that were needed. Furthermore, no main activities were skipped or missed during this case study. PMPM was experienced as especially useful in guiding this project since it made sure that all important activities were performed and the methodology prevented redundant work.

This research project is an initial step to a comprehensive process mining project methodology in which all phases and main activities of business process mining projects are described and that can be used as an efficient and effective approach for applying process mining in practice. Nevertheless, the methodology needs more empirical evidence to be presented as a valuable methodology for business process mining projects. Therefore, the main priority is to evaluate this methodology more extensively.

Table of Contents

1. Introduction

Organizations spend a lot of effort in analysing and improving their processes. Traditionally, analysing processes is time consuming, involves many people and is expensive. Process mining is an emerging discipline that allows for the analysis of business processes based on automatically logged events, which can often be done quicker, cheaper and in a more reliable way than traditional analysis. The promising research area of process mining provides techniques to discover, monitor and improve processes in a variety of application domains. Extracted information from an IT system, could for example, discover process models, detect deviations from the blueprint or investigate the interaction of resources in a process. In the last decades, a comprehensive set of different process mining techniques has been developed.

1.1 Problem Statement and Relevance

Based on the development of commercial process mining software tools (Perceptive Reflect[2], Fujitsu Interstage Process Analytics[3], QPR ProcessAnalyzer[4], Disco[5]), and attention of large software companies[6] and business technology watchers[7], process mining is emerging in practice. However, little research has been done about how process mining is applied in practice.

The application of process mining in an organizational context requires several additional activities next to the actual process mining analysis, e.g. definition of objectives, creating an appropriate dataset and evaluation of the results. Practitioners should be supported in identifying required activities and preventing problems which could occur during the process mining projects. In this context, a methodology that describes what is necessary to accomplish process mining in practice will be of great value. However, the process mining community does not have a methodology to conduct organizational projects yet. A lack of a process mining methodology for business is also pointed out recently at 'Process Mining Camp 2012'[8], a conference where process mining professionals share their experiences in applying process mining in an organizational context. At this conference, organizer C.W. Günther pointed out: "What we are still lacking in the process mining community is a certain kind of commonly agreed methodology that shares best practices and describes a way how to apply process mining."

A Literature study by [Heijden 12] analysed several methodologies (KDD process [Fayyad 96], CRISP-DM model [Chapman 00], L* life-cycle model [Aalst 11a], Process Diagnostics Method [Bozkaya 09], Methodology for BPA in Healthcare [Rebuge 12]) for conducting data- or process mining projects. This paper conclude that all these guideline systems have shortcomings in what is needed for a methodology that guides organization in applying process mining. There are several arguments why a process mining methodology adds value to the field of process mining. Besides that an appropriate process mining methodology will give practitioners

1 Fluxicon Software. Retrieved August 3rd, 2012, from http://fluxicon.com/camp/
2 Perceptive Software. Retrieved August 3rd, 2012, from http://www.perceptivesoftware.com/products/product-explorer/business-process/perceptive-reflect.psi
3 Fujitsu. Retrieved August 3rd, 2012, from http://www.fujitsu.com/global/services/software/interstage/solutions/bpmgt/bpma/
4 QPR. Retrieved August 3rd, 2012, from http://www.qpr.com/products/qpr-processanalyzer.htm
5 Fluxicon Software. Retrieved August 3rd, 2012, from http://fluxicon.com/disco/
6 PR newswire. Retrieved August 3rd, 2012, from http://www.prnewswire.com/news-releases/lexmark-acquires-pallas-athena-132040058.html
7 CIO Business Technology Leadership . Retrieved August 3rd, 2012, from http://www.cio.co.uk/article/3337087/20-companies-watch-in-2012/?pn=2
8 Fluxicon Process Mining Camp 2012. Retrieved August 3rd, 2012, from http://fluxicon.com/camp/

guidance in applying process mining in an organization, it will also assist in sharing best practices, stimulate the adoption of process mining in the field and prevent to reinvent the wheel.

This Master's Thesis aimed at developing a methodology that describes what needs to be done to apply process mining in practice. All main activities from scoping the project and developing basic understanding of the business and process to transferring the results to the organizational process must be included. The approach must be suitable to improve business processes in all kinds of sectors and functional areas and be independent of time, budget and the tools used for working with the data.

1.2 Research Structure

The remainder of this report is structured according to the logical steps in which the research has been conducted. In chapter 2 a brief overview of the literature that is related to the topic of this research is given. In the third chapter the research design is described, including research questions and the project approach. A detailed list of the requirements of the methodology is presented in Chapter 4. Subsequently, in chapter 5 a detailed analysis of the activities that are required in an organizational process mining project is described. In Chapter 6 the new methodology is composed, based on the knowledge which is described in the previous chapters. The developed methodology is evaluated by a case study conducted at Rabobank Nederland and described in chapter 7. In the remaining two chapters the discussion and conclusions of the Master's Thesis are presented, in which also the value for both science as well as practice are is mentioned.

2. Theoretical Background

This chapter establishes the background and context for this research. It starts with a general introduction of process mining. Next to that, a summary of the available literature about the application of process mining in practice is presented. Finally, several identified methodologies in the field of data- and process mining are compared and evaluated.

2.1 Basics of Process Mining

Process mining is a process management technique that can be used to support several activities of the process management spectrum [Aalst 11a]. The Business Process Management (BPM) life-cycle describes the different phases of a business process, as visualized in *figure 2.1*. In the *design* phase, a process is designed. The designed process is transformed into a running system in the *configuration/implementation* phase. When the system supports the process, the *enactment/monitoring* phase starts. Operational changes during the process can be handled in the *adjustment* phase. Insights gathered during the evaluation in the *diagnosis/requirements* phase can trigger a new iteration of the BPM-life cycle starting with the *redesign* phase. The model also lists the different ways data and models are used in the life-cycle.

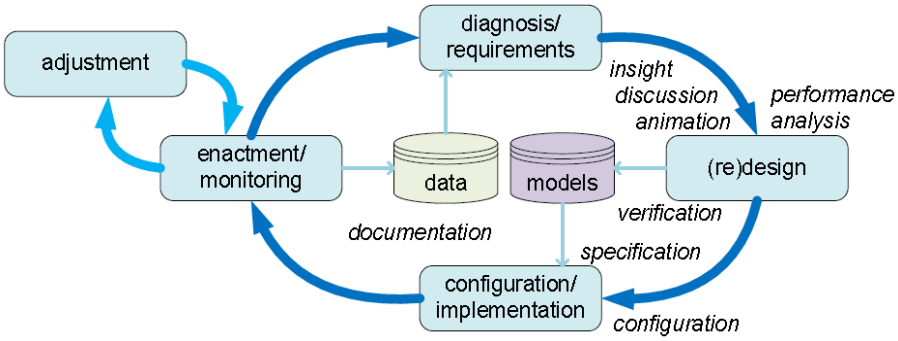


Figure 2.1, BPM life-cycle showing the different uses of process models [Aalst 11a]

In most organizations the diagnoses/requirements phase is only triggered by severe problems or major external changes. Process mining offers the possibility to truly 'close' the BPM lifecycle by using recorded process data to provide a better view on the process [Aalst 11a]. In other words, process mining automatically constructs process models that explain the observed behaviour in an event log that is recorded by an information system [Aalst 05].

Process mining is defined by [Aalst 04] as: "the method of distilling a structured process description from a set of real executions" and its aim is "to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's (information) systems" [Aalst 11b]. Event logs are typically recorded by information systems such as Enterprise Resource Planning systems, Workflow Management Systems, Customer Relationship Management systems, et cetera. [Aalst 07b]. Many of these information systems do have some kind of event log often referred to as 'history', 'audit trail' or 'transaction log' [Aalst 03].

The information in event logs relates to 'real' events and contains usually several aspects of the events. The case, or 'process instance' is the object which is being considered by an activity, e.g. invoice, insurance claim or customer order. Activities or tasks, are operations on a case, e.g. registering, checking or approving. Timestamp refers to the time of occurrence, which can be

recorded as a period containing a start and stop time, or just as a single moment. When people are involved, the resource, that for instance executes or initiates an event, can be included in the event log.

A process can be mapped in different perspectives, e.g. control-flow perspective, organizational perspective, case perspective [Aalst 07b, Aalst 11a]. The control-flow perspective focuses on the ordering of activities. The goal is to find a good characterization of all possible paths. The Organizational perspective focuses on information about resources hidden in the log, i.e. which resources are involved and how are they related. The goal of this perspective is either to structure the organization by classifying people or to show the social network. The case perspective focuses on properties of cases. For example, if a case represents a replenishment order, it may be interesting to know the supplier or the number of products ordered [Aalst 11a].

Process mining can map the described perspectives of the process. Orthogonal to the different perspectives, three main types of process mining can be identified. *Figure 2.2* positions the main types of process mining related to the 'world', software system, (process) model and event log. Process mining establishes a link between the event logs on the one hand and process models on the other hand [Aalst 11a].

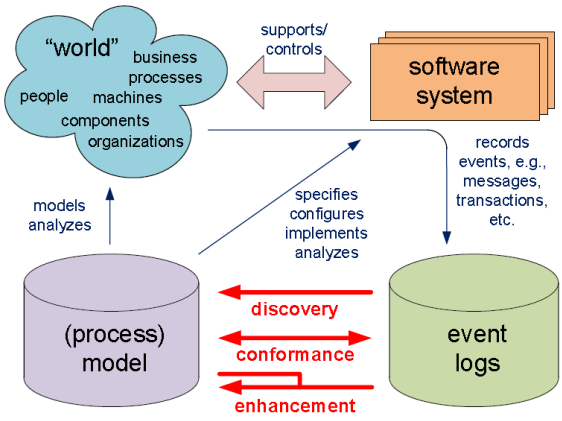


Figure 2.1, Positioning of the three main process mining types: discovery, conformance and enhancement [Aalst 11a]

The first main type of process mining is *discovery.* A discovery technique uses an event log and procures a model without using any a-priori information, e.g. a control-flow showing the flow of cases in a process or a social network showing how people work together in an organization. The second type of process mining is *conformance*. Conformance techniques compare existing process models with an event log of the same process, e.g. detect, locate and explain deviations or checking the 'four-eyes' principle, . The third main type of process mining is *enhancement.* Here, the idea is to extend or improve an existing process model using information about the actual process recorded in an event log, e.g. repair the current process model or extend the model with information about resources, decision rules, quality metrics et cetera [Aalst 11a].

2.2 Process Mining in Practice

Not much research has been done about the adoption and use of process mining in practice, which is probably because of the emerging status of process mining in the field. Nevertheless, the great number of techniques that are developed in a short time and several new commercial tools that are becoming available show the potential of this domain. [Ailenei 11] has compared

four commercial process mining software tools that are available for using process mining techniques and concluded that the potential of process mining is not yet completely exploited by the commercial process mining systems that are available.

[Prince 11] pointed out in his study several factors that are relevant to conduct a process mining project successfully [Prince 11]. The result of this study is a 'Process Mining Success Model' which shows the relationships between several success factors, moderating factors and success measures of process mining projects and is validated by a multiple case study of four process mining projects. The research of [Prince 11] supports the people involved in a process mining project by giving them insight in the factors that are important to perform the project successfully.

Furthermore, there are studies that apply process mining in a specific context. Some of these studies empirically evaluated process mining techniques [Goedertier 10], [Medeiros 07], [Wen 04]. Some studies developed a methodology for a specific purpose, for example to give a broad overview of the process(es) of the organization within a short period of time [Bozkaya 09] or to reduce fraud risk [Jans 08]. Other studies developed a methodology for a specific context, e.g. healthcare environments [Rebuge 12, Janssen 11].

2.3 Methodologies

Methodologies are conceptual structures that are used to analyse and organize data [Herrman 09] and serve as a guideline for solving a problem [Irny 05]. Developing a methodology to guide business process mining projects can help practitioners in performing such a project, especially when they have not much experience in this domain. To enhance its applicability, the methodology developed in this research must be a general approach that can be applied in all kinds of process mining projects in practice and consists of different stages next to the actual process mining stage to cover the whole project life-cycle.

Data Mining is defined as: "the analysis of (often large) datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner" [Hand 01]. Process mining uses several classical data mining techniques such as discovery and enhancement approaches focusing on data and resources. Since process mining is partly build on data mining [Aalst 11a] and also uses datasets as input for its techniques, data mining methodologies could be helpful to develop a methodology for process mining. Although the domains of data- and process mining are probably party overlapping, differences exist. The main difference is that process mining is considered with processes and thereby combines different events while data mining is usually applied on static data and aims to find unsuspected correlations.

The literature study of [Heijden 12] compared five different methodologies in the area of data- and process mining which could be helpful by providing inspiration for developing a process mining project methodology, although they all lack in being a general approach for all business process mining projects. The first two methodologies are from the data mining field and the other three methodologies are from the domain of process mining.

1. Knowledge Discovery of Databases (KDD) process is a common framework that aims to understand the variety of activities in the KDD field and how these activities are related.

[Fayyad 96] views the KDD process as a set of various activities in order to make sense of data. The core of this process is the application of data mining methods for pattern discovery.

2. CRoss-Industry Standard Process for Data Mining (CRISP-DM) is a widely used methodology developed to support the professionals that apply data mining and to demonstrate prospective customers that data mining was sufficiently mature to be adopted as a key part of their business processes [Chapman 00].
3. Process Diagnostics Method (PDM) is developed by [Bozkaya 09]. This methodology highlights three different perspectives of process mining and aims at giving a broad overview of the organization's process(es) within a short period of time.
4. Business Process Analysis in Healthcare environments (BPA-H) is built on PDM and is introduced by [Rebuge 12]. This methodology for the application of process mining techniques in a healthcare setting aims to identify regular behaviour, process variants, and exceptional medical cases.
5. L* life-cycle model for mining Lasagna processes (L*), a five-stage model that describes the life-cycle of a typical process mining project aiming to improve a structured process [Aalst 11a].

[Heijden 12] investigates the similarities and differences of these methodologies by outlining these approaches along four different stages: 1. *developing understanding*, 2. *data preparation*, 3. *performing mining* and 4. *feedback*. *Figure 2.2* gives a graphical overview of this comparison.

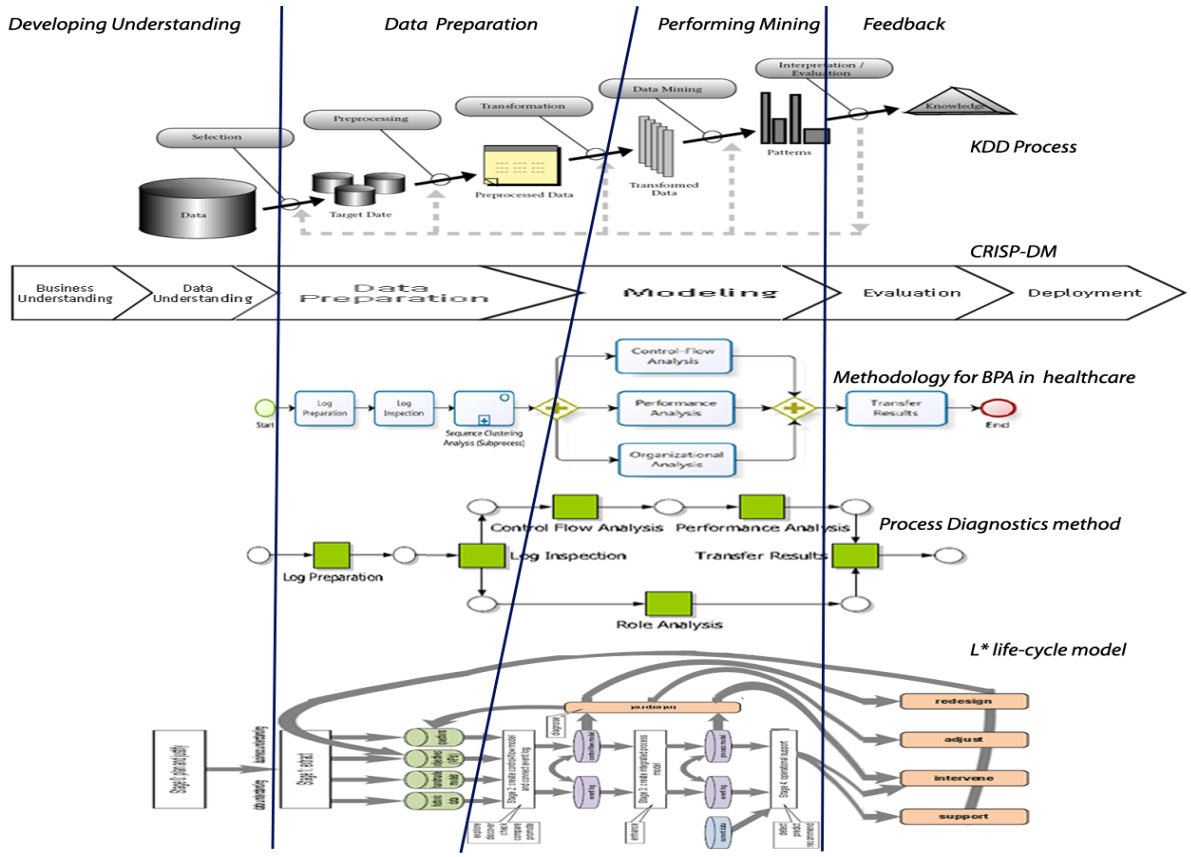


Figure 2.2, compared data- and process mining methodologies by [Heijden 12]

[Heijden 12] concludes that there exists quite some overlap between the different methodologies in their described phases as can also be found in *figure 2.2*. For business-driven mining projects (CRISP-DM, KDD, L*) it is important to start with determining the goals of the project or specific questions that need to be answered. Data-driven projects are executed to deliver valuable insights [Aalst 11a] for which PDM and BPA-H can be used. Next to the determination of the objective of the project, the available data also has to be gathered and converted so that it is suitable for use. After applying mining techniques, all methods use one or more steps to evaluate the mined information and present this information so that it can be used by the organization.

Differences between the methodologies can mainly be found in the actual mining step. KDD does not present an approach to perform these activities, CRISP-DM describes a repeating approach to filter, use and evaluate different methods and the process mining methodologies make a separation between different mining perspectives that can be used.

Methodology	Domain	Driven by	Process-specific
KDD	data mining	business	No
CRISP-DM	data mining	business	No
PDM	process mining	data	No
BPA-H	process mining	data	Yes
L*	process mining	business	Yes

Table 2.1, methodology characteristics

As described in *table 2.1*, several shortcomings for being a suitable organizational process mining project approach can be found in the methodologies.

- CRISP-DM and KDD are tailored to data mining projects. According to the definition of data mining, its aim is "to find unsuspected relationships and to summarize the data", which is different from the aim of process mining, "to discover, monitor and improve real processes". A data mining approach that guides projects with a different aim will probably not be an appropriate methodology for process mining, next to the fact that data mining data and techniques can (and usually will) differ from process mining data and techniques.
- PDM and BPA-H do not take the business in consideration. These methods aim at discovering knowledge apart from what is interesting for business. Practical process mining projects need an objective since time and money are not unlimited. Therefore a step to understand the goals of the organizational process has to be included.
- BPA-H and L* are designed for specific processes. BPA-H is designed for unstructured, healthcare processes. While L* describes the typical life-cycle for mining structured processes.

Methodology L* should be appropriate for structured processes, is tailored to process mining and shows that business understanding is important. Nevertheless, this methodology has drawbacks in the mining part. First, L* presumes that a process mining analysis always has to start with a control-flow model. This is definitely not always the case, since also other perspectives can be started with, e.g. organizational perspective or case perspective. Secondly, an integrated process model is presented as an enhancement of the control-flow model, but this could also be another model. Furthermore, operational support techniques can be used

immediately without first discovering other models in the same project if a pre-mortem event log and the required process knowledge are already available.

Since none of the five methodologies is, tailored to process mining, business driven, and appropriate for all processes, the identified methodologies cannot be presented as a general approach for process mining projects in organizations. Therefore this research aimed to develop a methodology that is appropriate for business in discovering, monitoring and improving their processes using process mining.

3. Research Design

This chapter outlines the design of the research aimed to meet the objective. First, the chapter the definitions of process mining and business process mining projects that were used are presented. Furthermore, the research question and several sub research questions are outlined that guided this research. Finally, the approach that described the different phases of the Master's Thesis is presented.

3.1 Definition

The definition of process mining and business process mining projects that was used in this research project is described to prevent misunderstanding and to provide clarity.
Process mining is: "the method of extracting process knowledge from a set of automatically (, partially) logged events."
Process mining is the main part of business process mining projects. A business process mining project is: "A Planned set of interrelated tasks to be executed over a fixed period and within certain cost and other limitations that aim to discover, monitor and improve business processes by discovering useful knowledge from logged events."

3.2 Research Questions

Drawing upon the goal of the project, the main research question of the project was defined as:

What would be an industry-, tool-, and application neutral approach for practitioners to conduct business process mining projects?

This approach describes what has to be done to apply process mining in practice. All main activities from scoping the project and developing basic understanding of the business and process to transferring the results of the project to the organizational process are included.

On a lower level, several sub research questions (SRQ) were defined in order to arrive at an answer to the main research question above. These sub research questions were derived from the main research question. The first two questions were answered using literature, expert opinions and hands on experience in applying process mining in practise. SRQ1 and SRQ2 describe what is required in a guiding approach for process mining projects.

SRQ1: What are the requirements for an industry-, tool-, and application neutral approach for practitioners to conduct business process mining projects?

Question SRQ1 identified the objectives of the process mining methodology that end-users have. Using these requirements, lower level functional requirements were derived to identify the activities that are required to conduct a process mining project. The next sub research question was formulated to identify how process mining should be applied in organizations in terms of the different activities that should be performed.

SRQ2: *What are the required activities to perform a business process mining project and what should be the order of these activities?*

SRQ2 resulted in a clearly defined ranking of required activities. For SRQ3, the answers SRQ1 and SRQ2 were synthesized to design a process mining project methodology. The shortcomings that were identified in other methodologies as described in *table 2.1*, do not apply to the proposed methodology as was implied by the third sub research question:

SRQ3: *What can be an appropriate methodology according to requirements from SRQ1 and SRQ2 and that does not have the shortcomings of the other methodologies?*

As a result of answering SRQ3, a new methodology that can be used for all organizational process mining projects is presented. To evaluate the practical use of this methodology a case study was conducted at the Financial Services department of Rabobank Nederland. This resulted in the last sub research question:

SRQ4: *How does the proposed methodology perform in a business process mining project?*

The feedback from the case study was used as an evaluation of the practical value of the methodology in organizational process mining projects.

3.3 Project Approach

This section introduces the different stages of the Master's Thesis, which is also visualized in *figure 3.3*. These different stages are aligned with the four sub research questions. In the first part, the business requirements of the process mining approach are described. Subsequently, the activities which are required to perform a process mining project are listed. The third part is considered with proposing a suitable methodology based on the requirements in the first two parts. The proposed methodology is evaluated in practice during a case study. A description of this case study and the corresponding evaluation can be found in the last part of the project.

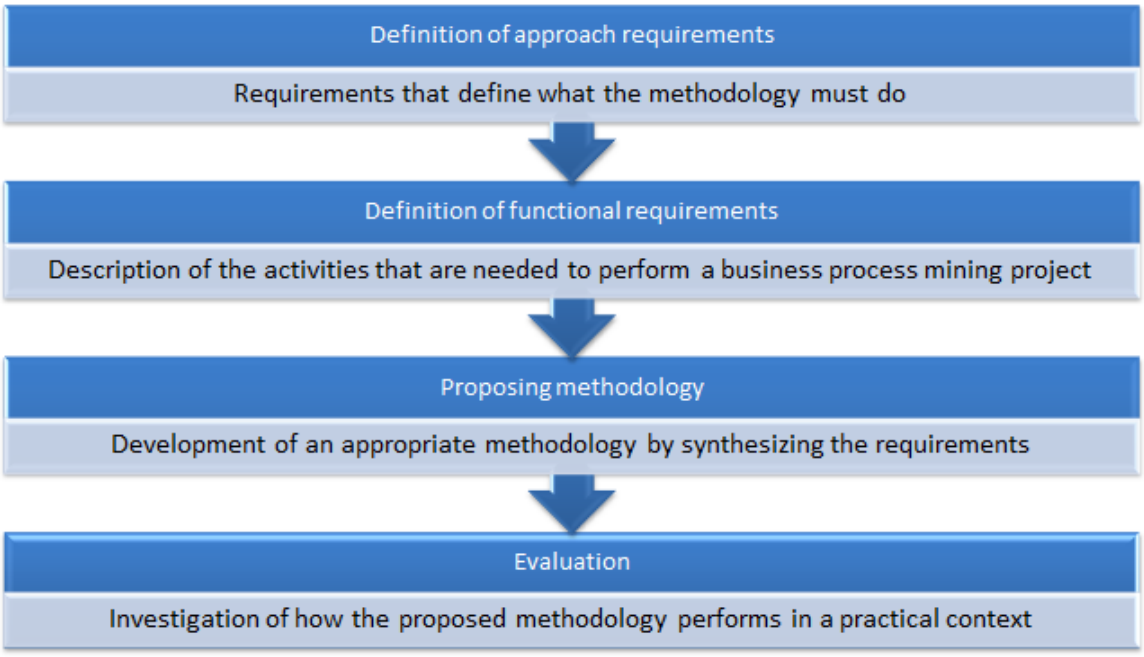


Figure 3.1, Project Approach

It is important to identify the requirements of the methodology from a practical and scientific point of view to come to a global process mining approach. 'Systems Engineering' is an interdisciplinary engineering management process that evolves and verifies an integrated, life-cycle balanced set of system solutions that satisfy customer needs [DSMC 01].
According to [DSMC 01], a system is, simply stated, an integrated composite of people, products, and processes that provide a capability to satisfy a stated need or objective. Although the process mining methodology should provide a capability to satisfy a stated need or objective, it is not a composite of people, products and processes, but who is using the methodology will definitely interact with people, products and processes.

Given the comparison between the methodology and a system, and the comprehensiveness of Systems Engineering, this framework was helpful to develop the process mining methodology. Therefore the System Engineering Process (SEP), as described in *figure 3.2*, was used during the development of the process mining methodology. First, a *requirements analysis* was made to identify the objectives of the process mining methodology. Secondly, the activity *functional analysis and allocation* transformed formulated functional requirements into a description of the low level functions that the process mining methodology can and should have. Thirdly, *design synthesis* is done, which is the process in which the design of the methodology was developed based on the outcome of *functional analysis and allocation* and verified by the requirements of the first activity.

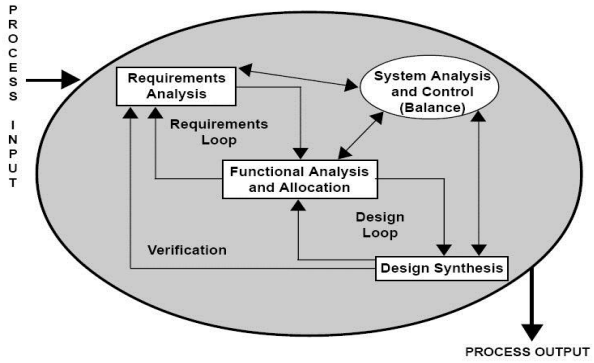


Figure 3.2, The Systems Engineering Process [DSMC 01]

Definition of approach requirements
The task that determines the requirements of the methodology is called *requirements analysis* in systems engineering. *Requirements analysis* is one of the fundamental activities of systems engineering and critical to the success of a project [DSMC 01]. The aim of this research was both descriptive as prescriptive. On the one hand the methodology should be a useful approach for performing process mining projects, otherwise the methodology would not have any value. On the other hand it should describe the activities that should be performed to guide practitioners by applying process mining in organizations, as can be concluded from the main research question. The methodology must be suitable for different sectors and industries, different functional areas and also for a different degree of process structuredness. In the first part of the research, customer requirements were translated into a set of requirements that define what the methodology must do. The analysis of the requirements was done using theoretical knowledge, involvement of the experiences of practitioners that conducted process mining projects, existing methodologies of data- and process mining and, of course, some common sense.

A helpful approach for project management is the Work Breakdown Structure (WBS), a hierarchical structure to decompose a project. WBS defines and groups a project's discrete work elements in a way that helps organize and define the total work scope of the project [Pritchard 99]. Using WBS, work packages can be divided into activities, organization and outputs. These three dimensions can be found in every project and every phase of a project. The dimension 'organization', which defines the people involved, was not taken into account for the development of the process mining methodology, because this does very much depend on the organization and the size of the project.

Definition of required activities

The second stage of SEP, *functional analysis and allocation*, describes what the methodology logically does. High level functions as described in the former phase were decomposed into lower-level functions, i.e. the activities that are required to perform a process mining project. The list of required activities includes and describes all activities that should be taken into account for any business driven process mining project. This list is not specific in the context of functional area of the process, the type of IT system that is used, sector of the company, amount of people involved, maturity of the process et cetera. Furthermore, the forming of a project team, convincing people, determining time, budget and all other organizational aspects that are 'ordinary' for undertaking a business project and do not radically change within a typical process mining project, were also out of scope.

Proposing methodology

By synthesizing the approach requirements and its required activities, the design was created. In this third stage of SEP the actual methodology is proposed. The methodology should meet all requirements as described in the former parts, including the low-level functional requirements.

While designing the methodology the following principles of [Husar 08] were taken into account to make the methodology as understandable as a possible and in order to increase the probability of success:

- Use clear names
- Avoid complex structures
- Be realistic
- Simplify rather than elaborate
- Be open to change
- Focus on the methodology, not on tools
- Describe key points, not creating an extensive document that deals with everything

Evaluation

The aim of this last part was to test the proposed process mining methodology in practice to evaluate its usefulness. This was done by means of a case study, performed at the Financial Services department of Rabobank Nederland. The Financial Services department is responsible for handling all invoices which are sent to the Rabobank. In the case study different organizational project objectives were formulated to expand the set of tested aspects of the methodology. After the case study, the support of the methodology was evaluated and discussed to identify problems or shortcomings.

4. Requirements Analysis

This chapter describes an identification of the requirements of a business process mining methodology, the first activity of SEP. The requirements create the scope of the methodology in terms of the conditions that must apply and what it must be able to do. Unconstrained and non-integrated requirements are seldom giving a sufficient solution for a problem [DSMC 01]. Therefore formulating requirements is inevitable. The requirements were formulated by knowledge gathered from scientists in the process mining field, process mining professionals, managers that facilitated process mining projects, scientific literature, summaries of process mining projects and own experiences while promoting and applying process mining at Rabobank Nederland, *appendix B*.

In systems engineering, requirements analysis should, in general, result in a clear understanding of:

- Functions: What the system has to do
- Performance: How well the functions have to be performed
- Interfaces: Environment in which the system will perform
- Other requirements and constraints [DSMC 01]

The Institute of Electrical and Electronics Engineers (IEEE) that is dedicated to advancing technological innovation and excellence produced an industry standard for requirement analysis, IEEE P1220. The requirements analysis in this research is based on IEEE P1220, which lists the 15 tasks as described in *figure 4.1.*

1. Customer expectations
2. Project and enterprise constraints
3. External constraints
4. Operational scenarios
5. Measure of effectiveness (MOEs)
6. System boundaries
7. ~~Interfaces~~
8. ~~Utilization environments~~
9. Life-cycle
10. Functional requirements
11. Performance requirements
12. ~~Modes of operation~~
13. ~~Technical performance measures~~
14. ~~Physical characteristics~~
15. ~~Human systems integration~~

Figure 4.1, IEEE P1220 Requirements analysis task areas [IEEE 94]

However, as concluded before in *section 3.3*, a methodology does not include the full concept of a system. Several tasks are considered with requirements that do not apply for the development of a methodology:

- Task 7 of IEEE P1220 describes the functional and physical interfaces.
- Task 8 describes the environmental factors that have impact on the performance.
- In task 12, the conditions that determine the modes of operations under development are defined.
- Key indicators that are tracked during the design phase (budget, time, et cetera) are formulated in task 13.
- Task 14 describes all physical characteristics of the system.
- In the last task, number 15, the human factor considerations (e.g. space limit, eye movement) which affect the system are identified.

Since the process mining project methodology is not a physical product, it cannot have performance problems because of environmental issues and is not developed in a team with a budget and deadlines, The tasks with the numbers 7, 8, 12,13, 14 and 15 are therefore excluded from the requirement analysis.

4.1 Customer Expectations

Applying process mining in practice is often difficult, because of the non-existence of an appropriate methodology for business process mining projects. An overview is missing of what is needed to accomplish a process mining project, e.g. managers often do not know how process mining can be helpful for their process and data specialists do not know which data is needed. The main questions that the process mining methodology must be able to answer to satisfy these needs are:

- What are the main phases of a process mining project?
- How can process goals be aligned with the application of process mining?
- What parts of the process are suitable for process mining?
- What kind of process data is needed?
- What must be done to use exported process data as input for process mining?
- What types of analysis can be done using process mining?
- How can the analysed results be deployed in an organization?

Or, more general:

- What are main activities in a process mining project?

The methodology aimed to answer these questions for all customers that are applying or think about applying process mining to their organizational process. This includes another requirement/expectation, that the methodology must be suitable to be used in every organization, functional area and sector.

4.2 Project Constraints

Project and enterprise constraints are the constraints that apply to the development of the process mining methodology. Traditionally, project constraints are listed as 'scope', 'time' and 'cost', i.e. the project management triangle, as a useful device for analysing the goals of a project [Bethke 03]. The costs for this research project are one student on full-time basis with some part-time support from scientists and practitioners. There are time issues that applied on the project, since the total time to perform this Master's Thesis is about half a year. Because of these time issues is decided, in close consultation with all supervisors that the practical evaluation (scope) of the methodology will be limited to one case study. The scope of the project and the methodology is further described in *section 4.6*.

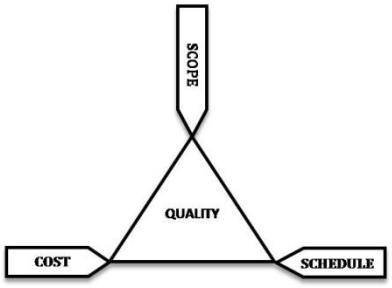


Figure 4.2, project management triangle

4.3 External Constraints

This section describes the external constrains that are impacting the use of the process mining methodology. The performance of the methodology depends on the people, products and processes using this approach. First, people that are conducting the project, e.g. project leader and process miner must be capable of doing their tasks. For example they must be able to

perform the different activities that interact with the organizational environment and know how to apply the mining techniques to the event log to retrieve requested information. Besides, the people working in the organization of the process that is mined, must not impede the project, e.g. prevent availability of data and lie about process issues. Secondly, usually several products/tools are used in a process mining project, e.g. ERP system that logs and extracts data, software to create an event log and process mining tools. The tools that are planned to use must be available for the project and be capable to fulfil its needs. Next to that, the process that is mined in the process mining project must generate data that is appropriate for mining, e.g. log several activities and different aspects, and be trustworthy. Furthermore, the process may not radically change during the project which can make mined results useless. Inappropriate people, products or processes can undermine the success of a process mining project. Therefore to make optimal use of the methodology the people, products and processes must be well-managed.

4.4 Operational Scenarios

Operational scenarios scope the anticipated use of the process mining methodology. In this section three scenarios are described that define how the methodology should guide a process mining project, which can be found in *table 4.1, 4.2* and *4.3*. All three scenarios start with an initiator that has an objective for a specific process. The objective must be identified by the project team after which several activities take place to meet the objectives of the project. Each of the three scenarios describes an example based upon one of the three main process mining types: discovery, conformance and enhancement.

Scenario 1:
The following scenario describes a complicated process of a hospital, containing many activities and different flows, to heal patients that have different types of cancer. The doctor wants to know how the process flow looks like to be better able to manage the process.

Scenario1	Discover the control-flow model of an unstructured process
Goal	Be better able to manage a process by knowing the possible flows
Actors	doctor, project team, data specialist, employees
Pre-conditions	An event log containing cases, activities and time (order)
Post-conditions	A control-flow model that describes the flow of cases in the hospital's process
Quality Requirements	1. All main activities of the process are logged 2. Cases and activities can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Develop understanding of the main process 2. Identify and gather the required data 3. Prepare the required data to apply analysis techniques 4. Apply a suitable mining technique to discover the control-flow 5. Analyse the output 6. Ensure that the control-flow is useful for the doctor, i.e. structured enough 7. Present the discovered flow

Table 4.1, Scenario 1, Discover the control-flow

Scenario 2:

The following scenario describes the procedure of a department manager of an invoice process that has a process objective to manage risk in the process and wants to know if the four-eyes principle is adhered to all cases to check if there are cases in an invoice process that are incorrectly handled.

Scenario2	Check the four eyes principle on activities in a process
Goal	Identify incorrect handled cases to manage risk
Actors	Department manager, project team, data specialist, employees
Pre-conditions	An event log containing cases and the resources that handled them
Post-conditions	All cases that are not handled according the four eyes principle
Quality Requirements	1. The required activities that need to be checked on the four eyes principle must be contained in the event log including the employee that executed that activity. 2. Cases and resources can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Identify the activities that use the four eyes principle 2. Identify and gather the required data 3. Prepare the required data to apply analysis techniques 4. Apply a suitable mining technique to check the principle 5. Analyse the output 6. Return the risky cases

Table 4.2, Scenario 2, Check the four eyes principle

Scenario 3:

The last scenario describes a manager of call center that wants to identify the average handling time of its agents. The types and amount of activities that each agent performs is known and the manager wants to extend this model with the time aspect for each type of activity to be better able to manage his work force.

Scenario3	Extend the resource activity model with the time aspect
Goal	Identify agent performance to be better able to manage the work force
Actors	Call center manager, project team, data specialist, employees
Pre-conditions	An event log containing activities, resources and time (detailed) and a current resource activity model
Post-conditions	A resource activity model extended with time
Quality Requirements	1. Time is logged detailed enough to be useful 2. Resources and activities can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project 4. The event log contains the resource activity combinations of the current resource activity model
Description (activities)	1. Identify and gather the required data

	2. Prepare the required data to apply analysis techniques 3. Apply a suitable mining technique to add time to the current resource activity combinations 4. Analyse the output 5. Present the new model and the times for each resource activity combination

Table 4.3, Scenario 3, Extend the resource activity model with the time aspect

4.5 Measure of Effectiveness

A measure of effectiveness reflects the relation between customer expectation and satisfaction about the methodology. *Section 4.1* describes questions that customers have and which the process mining methodology must be able to answer. The effectiveness can be measured by identifying if these questions can be answered by the new proposed process mining methodology.

4.6 Methodology Boundaries

This section describes what must be under control of the process mining methodology and what must be outside control. This means the creation of boundaries to define the scope of the methodology. The methodology contains a guide for process mining projects, but not in terms that are not specific for process mining projects compared to 'regular' projects. This means that the methdology does not provide support for people-, time-, budget- or team management. The methodology provides an ordered list of activities that should be executed in any business. The methodology must be abstract, that means applicable to any organizational process mining project and time independent, and thus not for a specific time span and also useful in the future. Furthermore, the methodology does not provide support for technique-, tool- or systemrelated choices or actions. All activities that are listed should be applied to all projects, specific activities that are related to specific processes may be mentioned, but not further explained in detail.

4.7 Life-cycle

The key life cycle phases in system development are: develop, produce, test, distribute, operate, support, train, dispose. Two of these phases apply for the development of the process mining methodology: development and testing.

After the development according to the adapted SEP, the value of the methdology will be tested by a case study conducted at Rabobank Nederland Financial Services. The testing will be done by conducting an extensive process mining project using and evaluating the developed methodology to indicate imperfections and improve the methodology.

4.8 Functional Requirements

Functional requirements describe what the methodology must be able to do. Using the identified methodologies of *section 2.3*, a business process mining project can be divided in six different phases. Three general project phases (scoping, evaluation, deployment) and three process mining specific phases (data understanding, creating event log, apply process mining). These different phases can be described in the following way:

- Developing understanding to identify how process mining can be applied to the process and to formulate the objectives that drive the process mining project. (Scoping)

- Understanding the data that is needed for these objectives and investigating if and how this data is available. (Data understanding)
- Describe how the data must be gathered and prepared to be appropriate as input for process mining techniques. (Event log creation)
- Apply process mining techniques to answer the business questions. (Process Mining)
- Evaluating the accuracy and value of the output of the process mining techniques. (Evaluation)
- Report the results to the organization so that it is possible to deploy the gathered knowledge in the process environment. (Deployment)

These requirements were further decomposed in the second fundamental activity of SEP, *functional analysis and allocation* which is described in chapter 5.

4.9 Performance Requirements

Performance requirements give the required effectiveness measures as described in *section 4.5*. The new process mining methodology should satify the customer by providing answers to all questions as described in *section 4.1*.

4.10 Chapter Conclusion

The aim of this chapter was to answer sub research question 1:

SRQ1: What are the requirements for an industry-, tool-, and application neutral approach for practitioners to conduct business process mining projects?

The requirements for the development and the final methodology are identified in the different sections contained in this chapter: customer expectations, project constraints, external constraints, operational scenarios, effectiveness measures, methodology boundaries, life-cycle and performance requirements. These sections described the requirements of the methodology from different perspectives, created a scope, and supported in satisfying the business and creating the requested methodology. A summary of the requirements can be found in *table 4.4*.

Task	Requirements
Customer expectations	A description of the main activities in a process mining project
Project constraints	One student on full-time basis for about half a year
External constraints	People, products and processes must be well-managed
Operational scenarios	Three scenario examples are described that define how the methodology should guide a process mining project
Measure of effectiveness	Identification to what degree the customer expectations are realized
Methodology boundaries	Appropriate methodology for any business process, but no support for people-, time-, budget- or team management
Life-cycle	Development and testing
Functional requirements	Scoping, data understanding, event log creation, process mining, evaluation, deployment
Performance requirements	Satisfy customer expectations

Table 4.4, An overview of the methodology requirements

5. Functional Analysis and Allocation

This chapter describes the second fundamental activity of SEP (*figure 3.2*) which is *functional analysis and allocation*. *Functional analysis and allocation* decomposes the high-level functional requirements as described in *section 4.8*. For all functional requirements that are described in the requirements analysis, a detailed description of specific activities is given.

The required activities in a business process mining project are identified and described, just as in the former chapter, by knowledge gathered from scientists in the process mining field, process mining professionals, managers that facilitated process mining projects, scientific literature, summaries of process mining projects and own experiences while promoting and applying process mining at Rabobank Nederland, *appendix B*. *Table 5.1*. gives an overview of the high level functional requirements and their corresponding main activities that are described in this chapter.

Nr	Functional requirement	Required activities
1	Scoping	Identify the process and gather basic knowledge
2	Scoping	Determine the objectives of the project mining project
3	Scoping	Determine the required tools and techniques
4	Data Understanding	Locate the required data in the system's logs
5	Data Understanding	Explore the data in the system's logs
6	Data Understanding	Verify the data in the system's logs
7	Event log creation	Select the dataset in terms of event context, timeframe and aspects
8	Event log creation	Extract the set of required data
9	Event log creation	Prepare the extracted dataset, by cleaning, constructing, merging and formatting the data
10	Process Mining	Get familiar with the log by gathering statistics
11	Process Mining	Make sure that the process contained in the event log is structured enough to apply the required process mining techniques
12	Process Mining	Apply process mining techniques to answer business questions
13	Evaluation	Verify the modelled work
14	Evaluation	Validate the modelled work
15	Evaluation	Accreditate the modelled work
16	Evaluation	Decide on an elaboration of the process mining project
17	Deployment	Identify if and how the process can be improved by improvement actions
18	Deployment	Present the project results to the organization

Table 5.1, An overview of the main activities in a business process mining project

To determine how process mining can support in discovering, monitoring and improving processes, it can be helpful to identify the possible input that can be used to apply process mining in organizations. According to its definition (*section 3.1*), process mining uses recorded information based on process events as input to extract process knowledge. These events can contain a range of different types of information about processes. An increase of logged

information about the real activities that took place has a positive influence on the amount of knowledge that can be retrieved from the event log using process mining techniques. Aiming to describe the full set of possible information that could be involved in process mining, it is interesting to map all possible aspects of an event.

Research in journalism describes the concept of the five Ws (and one H) that are regarded as basics in information-gathering and which should be able to getting the complete story on a subject [MAN 96]. This concept describes the following six questions: What? Who? When? Where? Why? How? i.e. the primitive interrogatives, that were memorialized in a poem that opens with:

"*I keep six honest serving-men*
(They taught me all I knew);
*Their names are **What** and **Why** and **When***
*And **How** and **Where** and **Who**.*"
From 'The Elephant's Child' by Rudyard Kipling (1902)

The principle is that each question gives a factual answer and none of them can be answered with a simple 'yes' or 'no'. These questions are used by [Ha 06] as a method to analyse user behaviour. The 'Zachman framework' [Zachman 87] also uses the concept to describe the abstractions that a product can have. In addition to describing the complete set of information of a story, behaviour or a product, these questions can also be used to describe the aspects of an event. The following descriptions can be given to all primitive interrogatives:

5W and 1H	Description	Example
What?	Case	Invoice ID01251
Who?	Resource	Henk de Vries
When?	Time	10-06-2012 11:20 ; 10-06-2012 11:29
Where?	Location	Den Dolech 2, 5612 AZ Eindhoven
Why?	Motivation	To get salary
How?	Activity	Registering

Table 5.2, An example of an event description using the five W's (and one H) concept

The idea is that all possible information that could be involved in an event could be divided to one (or a combination of more) of these aspects. Not all the information of events will (and probably can) be recorded by an information system. Nevertheless, if a piece of information is recorded; other, more detailed information can often be derived. For example 'Employee C653' does also have a name, age, gender, mother et cetera which are all characteristics of the specific resource. Moreover, sometimes information is described as a combination of more than one aspect. If, as an example, cost regards the duration of the work that is spent by an employee, this information is a combination of time and resource for example, which can be recorded as a combination in the event log, but also derived from these aspects in an event if the timely rate of the resource is known and the duration of the event. A structured example of event information structured by the five Ws (and one H) is given in *appendix C*.

5.1 Scoping

A1: Identify the process and gather basic knowledge

Knowing the type of information that could be involved in events, one can start with understanding the organizational process. This is usually far from simple, because business processes are usually performed by a number of different people that oversee only a part of the process and often work at different departments across the organization. It is not necessary to know all details, but having general knowledge about the process e.g. main flows, type and amount of cases, lead times and resources will be helpful for the next steps in the process mining project. Furthermore, it is important to identify the parts of the process that are probably logged and those parts which are certainly not logged. Process objectives that are related to activities or other information of the process that is not logged, are not able to meet with the help of process mining. In the most optimistic way, information of activities that are not logged should be derived from logged events. Events can be logged by humans, but logging is done more reliably using devices. Usually, process mining is applied on the event information logged by information systems that is recorded digitally and automatically.

A2: Determine the objectives of the project mining project

When the process mining project initiator knows what information could be retrieved from the process using process mining, objectives can be formulated. According to [Aalst 11b], process mining can support businesses in discovering, monitoring and improving their processes as long as the required data is recorded in (information) systems. This implies that process mining can support a wide range of business objectives in mature processes as long as the objective is supported by historical information in the event log. There are basically three types of process mining projects according to [Aalst 11a]:

- Data-driven: no concrete question or goal, but curiosity driven. This is type of project has an explorative character and its goal is to deliver valuable insights.
- Goal-driven: projects that aspire to improve a process with respect to particular KPI's, e.g. cost reduction or improve response time.
- Question-driven: projects that aim to answer specific questions.

The first of these three types, *data-driven* is the most difficult to apply, because of its explorative character [Aalst 11a]. Existing process mining techniques can deliver a wide range of insights in different dimensions of the process, e.g. discovering control-flow, social network and case performance. In combination with a massive event log (which also can be filtered over and over again), it will be impractical to apply the full process mining functionality. Moreover, for business it is often not possible in terms of time and budget. This implies that it is normally not the most sensible method to use the complete process mining spectrum for delivering valuable business insights. If a few types of analysis are selected on forehand then the type of project is basically different, because this implies that it is driven by some kind of goal or question.

Another type of project is *goal-driven*. It can be difficult to determine how to use process mining in goal-driven projects. For example cost reduction means that there are some costs which can be calculated or derived from the process and that the objective is to decrease these costs. If costs are calculated as the working time that it takes to process a case, then the spent time of resources is needed. This implies that a derived objective can be: decrease the handling time of cases.

Question-driven projects have questions that drive a process mining project. These questions are part of process goals. E.g. using a KPI that aims at costs reduction, several questions can be formulated which give a clear plan for process mining, e.g. *What is the average time spent on cases? Which resources have the highest impact on the process in terms of spent time?* Formulating these kinds of questions from KPIs will support in clarifying the business objectives. Questions can give a clear direction in selecting data that is needed to apply process mining techniques. Besides, questions support in deriving subquestions for selecting the next step in the process mining project to answer the main question. As an example:
What is the average spent time for cases? → What activity has the highest impact on this time? → What kind of resources are working on this activity? → Do some resources take clearly more time than others?
By formulating an answer to the last question, managers can improve the process by e.g. replacing or training certain resources in order to reduce costs, which stimulates the main objective in this example. Each of the formulated questions includes a metric that can be measured, which makes it easy to support in meeting the objective.

A method that works according the principle of specifying metrics to target goals, is the Goal Question Metric (GQM) approach. GQM is based upon the assumption that for an organization to measure in a useful way it must first specify the goals for the organization and its projects, then it must trace those goals to the data that are intended to define those goals operationally, and finally provide a framework for interpreting the data with respect to the stated goals [Basili 94, Basili 92]. The resulting measurement model includes three levels: 1. Conceptual level (goal), 2. Operational level (question), 3. Quantitative level (metric). *Figure 5.1* gives an overview of this approach:

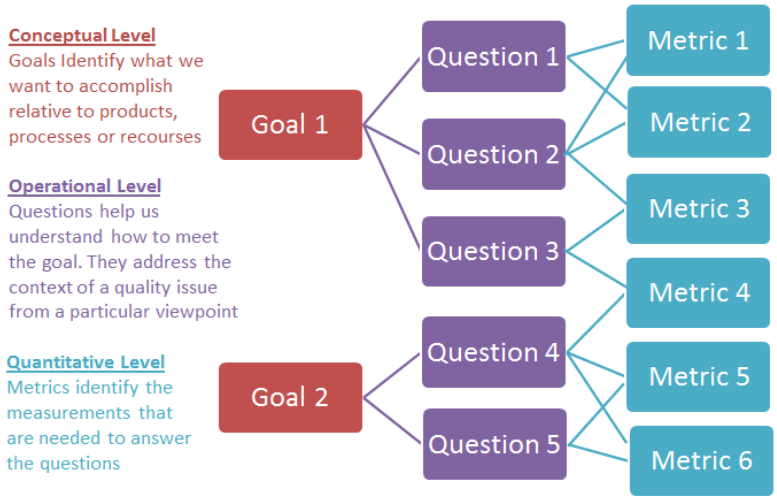


Figure 5.1, GQM definition according to its three levels

In order to meet the objective of applying process mining to a process, the GQM can be a helpful approach to derive clear metrics that are supported by defined goals and questions. These metrics are helpful in order to select the data that is needed for process mining. Formulating objectives start by selecting clear goals, creating questions and deriving metrics to clarify the according process mining related activities and as a description to select input. By knowing the as-is situation of the process and having an idea of the process events that are logged, it is possible to determine the objectives.

A3: Determine the required tools and techniques
In the scoping phase of the process mining project it is sensible to think about the tools and techniques that are required in the next phases, e.g. Which system can extract the data? What software tools can be used to create the event log? What are the process mining techniques that must be applied? What software will be used to apply the process mining techniques? This will help to get a global idea of what is needed and to have the required knowledge and tools available when they are needed during the project.

5.2 Data Understanding

A4: Locate the required data in the system's logs
The next phase after *business understanding*, is *data understanding*. For a process mining project this phase starts by identifying if the required input is digitally available in the organizational (information) systems. Process mining is only possible if process events are automatically recorded and that there is a sort of guarantee that this recorded information matches reality [Aalst 11b]. As described earlier in this chapter, the possible event information can be described or derived from the answers to the six primitive interrogatives.

Systems that support collaborative activities and their coordination are called Computer Supported Cooperative Work (CSCW) systems [Carstensen 99]. Not all these CSCW systems consider a structured process. [Aalst 07a] argues that process mining can also be applied to less structured processes supported by CSCW systems. In [Aalst 07a] a classification of CSCW systems based on two dimensions is proposed. On the one hand d*ata centric* (i.e., the focus is on the sharing and exchange of data) and *process centric* (i.e., the focus is on the ordering of activities) approaches/systems are distinguished. On the other hand a distinction is made between *structured* (predefined way of dealing with things) and *unstructured* (things are handled in an ad-hoc manner) approaches/systems.

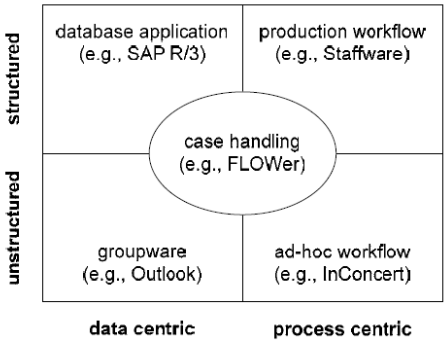


Figure 5.2, Spectrum of Computer supported cooperative work [Aalst 07a]

Identifying the process by considering event logs will usually be easiest when systems are structured and process centric, unfortunately it is often a challenge in unstructured and data centric systems to identify the case and the task for each event that is recorded. The cause of this challenge rises due to the fact that data centric systems (e.g. ERP systems) are not built to keep event logs in mind. However, this does not mean process mining is not possible. Furthermore, in some systems one must turn on the logging functionality before data is record. Another hurdle that sometimes should be overcome, is that data may be distributed over a variety of sources [Aalst 11b]. It could be problematic to identify the same type of information in different sources when different identifiers are used, which is often the case for ERP systems.

A5: Explore the data in the system's logs
Localized data needs to be explored to develop knowledge about the data that is contained in the logs and the way this data is organized. This can be done using manual techniques to find one's way through a dataset and bring important aspects into focus for further analysis. The aspects which are present in the data and how the data relates to one another can also be checked. If the data is distributed over different sources one should also analyse how this data is related and find a possible solution to combine this information. [Nooijen 12] developed a process analysis approach that provides a way to combine information from several parts of a process in ERP systems. This approach describes several steps, constructing a database schema, describing and extracting the log of the different parts, and mapping the different logs to create a complete process log.

A6: Verify the data in the system's logs
After the identification and exploration of the available data, the quality of the log data has to be evaluated. According to [Aalst 11b] event data can be judged along four criteria:

- Trustworthy: it should be safe to assume that the recorded events actually happened and that the attributes of events are correct.
- Completeness: no events may be missing.
- Semantics: any recorded event should have a well-defined meaning
- Safeness: privacy and security concerns are addressed when recording the events

In order to benefit from process mining, organizations should aim at event logs at the highest possible quality [Aalst 11b].

5.3 Event Log Creation

A7: Select the dataset in terms of event context, timeframe and aspects
The requirements that are described up to now do not directly imply a selection of the data that is required. Before a selection could be made of the data that is needed, a decision on several dimensions hast to be made. First of all, data can be divided into "pre mortem" and "post mortem" event data [Aalst 11a]. "Post mortem" event data refers to information about cases that have been completed, historic data. This type of data is most relevant for off-line process mining, e.g. discovering the control-flow of a process based on 3 months of event data. "Pre mortem" event data considers cases that have not yet completed, i.e. are still 'alive'. This is a required type of data for online process mining, which means that running cases are considered and potentially, still be influenced e.g. to predict completion time or to detect deviations. Depending on the process mining objective only historic, only current or both types of cases can be select. Although this decision is required, it depends on the type of IT system and available information if it is possible to select the type of cases on forehand. Therefore a selection depending on this requirement is occasionally done in a later stage.

A second dimension which must be considered is the timeframe of the dataset. Selecting the complete dataset is normally not desirable. A main reason is that processes are changing and therefore 'old' data will not be representative to show the current process. Another reason is that a complete dataset can consider a huge amount of data, which can result in performance and extraction problems, and often it does not significantly improve the dataset. Normally the time duration of the dataset is determined by the average duration of a case, but gives room to catch unusual long-running cases as well. Process mining company Fluxicon[9] introduced a

9 Fluxicon Software. How Much Data Do You Need?, Retrieved August 3rd, 2012, from http://fluxicon.com/blog/page/2/

formula based on the expected throughput time for the process:

$$timeframe = expected\ case\ completion\ time * 4 * 5$$

The factor four in this formula ensures that there is as much data to see four cases that were started and completed after each other. To ensure that occasional long-running cases are also accounted, the product of five is used, which is based on the 20/80 rule. This formula can be considered as a rule of thumb and will be functional most of the time. However, the more is known about a process, the better one will be able to judge about the amount of data that should be attracted. Furthermore, the time frame could be chosen in different ways. The possible ways in process mining software Disco[5] to filter events are:

1. Events extracted in a fixed timeframe
2. Events extracted based on cases started in a fixed timeframe
3. Events extracted based on cases completed in a fixed timeframe.
4. Events extracted based on cases completely contained in a fixed timeframe.
5. Events extracted based on cases intersecting a fixed timeframe.

Option 3 does always result in a selection of historic cases. However, it could be that these options are not possible in the event log creation phase, because of the type of available information and IT system. This selection could also be applied in a later phase using process mining software.

A third dimension that is of importance for selection an event log is the information that is contained about the events. The formulated objectives give assistance to identify the required event information that is needed. Without concrete questions it is very difficult to extract meaningful event data [Aalst 11b]. Process Mining can be applied to different perspectives. A main perspective of process mining is the control-flow perspective. This perspective shows a model of all possible flows of a case in the process. The control-flow model can be generated by discovering the flow of activities of each case and merging all flows together. To generate a control-flow model, aspects: activity, case and time (to order the flow of activities of a case) are necessary. Another perspective of process mining is the organizational perspective, for example to show 'handover work'. A handover work model shows the connections between resources that hand over cases to each other. To discover a handover work model, aspects activity, case, time and resource are required as input. For each objective, the required aspects should be derived to that are needed to meet the objective of a process mining project.

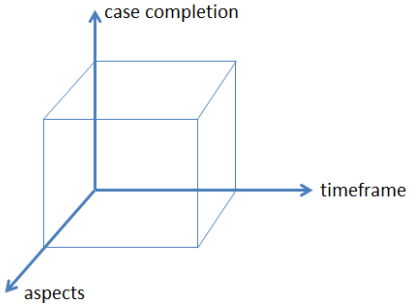


Figure 5.3, three dimensions of an event log

A8: Extract the set of required data
If the objective is clear, the available data is identified and the required dataset is available, then the next step is to extract the data. The challenges at this stage do very much depend on the amount of data and the software system/tool that is used. The dataset could often be exported to different kinds of data files.

A9: Prepare the extracted dataset, by cleaning, constructing, merging and formatting the data
Normally, an extracted dataset is not directly appropriate as input for a process mining technique. An extracted data has to be prepared, which can include several tasks. First of all, data aspects and records that are not necessary but included in the dataset can be removed in order to get a better overview. If the dataset contains outliers or missing values this has to be solved. Explaining how to deal with outliers and missing values will be out of the scope for this research and not process mining specific, but keep in mind that this can be necessary. Another preparation step is creating extra/new aspects, i.e. derived or combined attributes (for example deriving age from birth date, a salary group from a certain wage or combining day and time to one timestamp). When process information is logged in more than one source, an extra step may be required, which is merging the different datasets to one dataset. These sets of data should have a shared identifier so that combining them is possible. The final preparation step is structuring the values in terms of a required format, renaming values and inserting default variables. This action could require an investigation of the input that is required to use a process mining tool/technique.

5.4 Process Mining

A10: Get familiar with the log by gathering statistics
Up to now all requirements are associated with the preparation of process mining. At this stage the specific input for applying process mining should be prepared. Before applying a specific process mining technique it is sensible to get familiar with the event log and the process information that is contained in this event log [Bozkaya 09]. In this inspection step several statistics about the log are gathered. Examples could be the amount and diversity of process instances, activities and resources, but also the average number of events per case and the different start and end activities. These statistics give a first impression about how the events in the event log are correlated and can give an indication in how a process mining technique will perform on this data.

A11: Make sure that the process contained in the event log is structured enough to apply the required process mining techniques
The log inspection can indicate that the data is less structured, which is also referred to as a 'Spaghetti' process. The business processes of healthcare organizations are often 'Spaghetti' processes, because they are usually highly dynamic, highly complex, multi-disciplinary and ad hoc [Rebuge 12]. For less structured processes only a subset of process mining techniques is applicable. Given that not all process mining techniques perform well in capturing complex and ad hoc natures of processes it is sometimes essential to first make the 'Spaghetti-like' processes more 'Lasagna-like' [Aalst 11a]. There are several ways to simplify event logs to make them applicable for process mining techniques. Event logs can be filtered by activities based on their characteristics, e.g. absolute or relative frequency. Abstracting from infrequent activities can make models as simple as desired [Aalst 11a]. Besides activity-based filtering, there are more advanced types of filtering that transform low-level patterns into activities [Bose 2009]. Moreover, cases in a log can be clustered in homogeneous groups that show similar types of

behaviour to generated simpler models for groups of cases [Rebuge 09]. Furthermore, [Janssen 11] developed a set of best practices for process mining research in healthcare. The often unstructured processes in healthcare have already been studied by several researchers, but these studies suffer from disappointing results, which may be partly blame to the low accessibility and reproducibility of the process mining research methodologies. In the research project of [Janssen 11] 11 out of 22 identified patterns covering process mining research steps were selected as best practices as they showed the most promising results and highest accessibility.

A12: Apply process mining techniques to answer business questions

Applying process mining requires a decision for a specific process mining activity. As described in *section 2.1*, there are three main types of process mining: discovery, conformance and enhancement. This classification of process mining types can be adapted to a more comprehensive overview. [Aalst 11a] describes a 'refined process mining framework', *figure 5.4*, which reflects that process mining can be done online or off-line and that there are two types of models, 'de jure' models and 'de facto' models. Online process mining on the one hand is considered with pre-mortem data and off-line process mining on the other hand is considered with historic data. A 'de jure' model specifies how things should be done or handled and a 'de facto' model aims to capture reality.

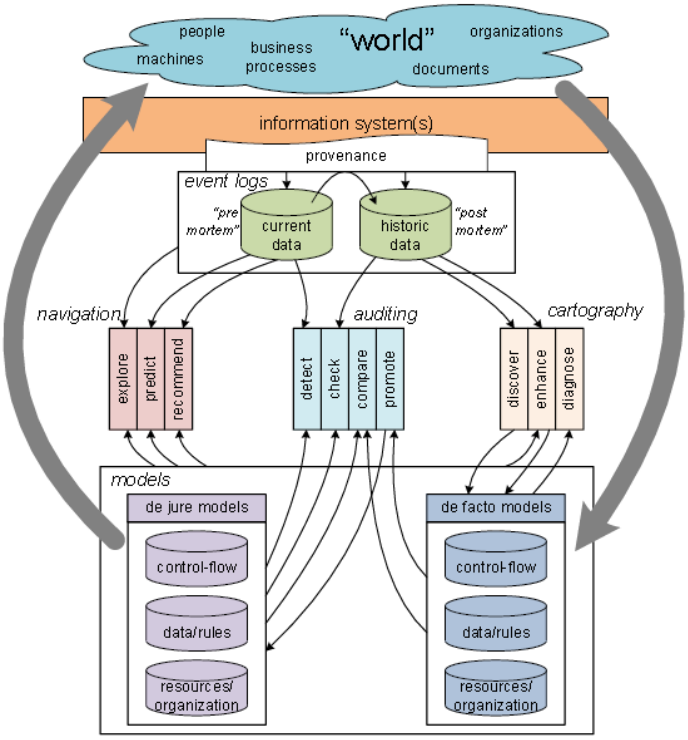


Figure 5.4, Refined process mining framework [Aalst 11a]

In the refined process mining framework, ten process mining related activities can be identified, which are grouped into three categories: cartography, auditing, and navigation. Cartography can be seen as the 'maps' describing the operational process of organizations and lists three activities:

- Discover: extraction of process models
- Enhance: extend or repair existing models

- Diagnose: focuses on classical model-based process analysis

Auditing activities are used to check whether business processes are executed within certain boundaries and lists four activities:

- Detect: comparing 'de jure' models with current data to detect deviations at runtime
- Check: crosscheck historic data with 'de jure' models to pinpoint deviations and quantify the level of compliance
- Compare: compare 'de facto' models with 'de jure' models to identify differences
- Promote: promote parts of a descriptive model to a new prescriptive model

Navigation activities are unlike cartography and auditing activities forward looking and list three activities:

- Explore: combining event data and models to explore business processes at run-time
- Predict: combining information about running cases with models to make predictions about the future of cases
- Recommend: use predictions to recommend suitable actions

[Aalst 11a]

The first and probably most easy way to divide process mining related activities is to check what kind of input they need and what kind of output they produce. Do they only need the prepared 'new' event log or are also models required as input? If a process mining related activity only uses the 'new' event log to extract knowledge, then it is definitely a discovery technique. The discovery activity is one of the three activities that generates a model. The other process mining related activities that generate models are *enhance* and *promote* which can be used to improve models. Activities that do not generate models and use the event log and models as input are *detect, recommend, predict* and *explore*. The remaining activities do not generate models and do not use the event log as input directly: *diagnose, compare*. Therefore this last group of process mining related activities cannot be used as the first activity in a process mining project. An overview of this classification is given in *figure 5.5*.

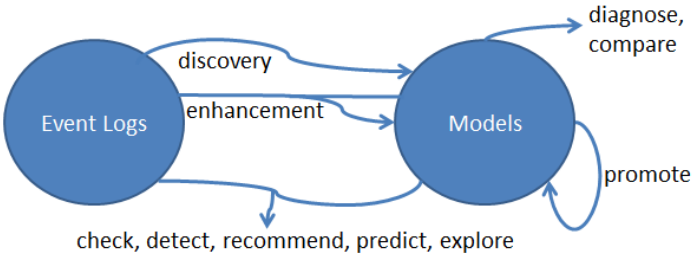


Figure 5.5, classification of process mining related activities in relation to event logs and models

It is important to choose appropriate mining techniques that are able to answer the questions and to meet the objective of the project. Indicating the specific mining technique that is needed for a project is not included in the process mining methodology, though. Suggesting mining techniques is excluded because of several reasons. First, available techniques are divided over several process mining software tools and the amount of techniques is very large, e.g. ProM 5.2 features over 280 plug-ins for process mining, analysis, monitoring and conversion [Verbeek 10]. Secondly, techniques change over time, techniques will be elaborated, improved and new ones will be created. Thirdly, the appropriateness of a mining technique does very much depend on the available data, knowledge, process maturity and the goals of the project. However, choosing an appropriate technique can be rather difficult for practitioners. Therefore, specific

support for choosing an applying an applying techniques will be an interesting topic for further research.

As already has been described in *section 2.1*, orthogonal to process mining activities, several perspectives of a process could be mapped. A perspective consists of, usually, a combination of different aspects of an event. To get a requested model with the help of a process mining technique, the technique must, besides doing the right activity, also generated an appropriate perspective. The control-flow perspective that is considered with a control-flow model consists of case, activity and time (at minimum the order of events). An organizational network showing resources that interact with each other in cases, consists of the case and resource aspect. When selecting a technique that generates a model, it is important to select a technique that is able to generate output in a requested perspective.

Process mining techniques are applied to retrieve process knowledge from an event log which can be used to meet the objectives of the project initiator and finally improve the process. Project objectives are usually supported by several questions which need to be answered. Normally, the process of answering the questions that are of interest for a business requires applying several techniques. Moreover, answering these questions is also a process of filtering, adjusting and digging in the event log to retrieve the most valuable information.

5.5 Evaluation

There are three ways to evaluate modelled results: verification, validation and accreditation. *Verification* deals with transformational accuracy. *Validation* deals with behavioural or representational accuracy. *Accreditation* is defined as "the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose" [DoDI 96]. During the verification process the model is checked on its correctness according to the technical description and specifications. Secondly, validation is the process of determining the degree to which the model is accurate in representing the real organizational process. Thirdly, accreditation is the assessment of the degree to which the process mining results meet the business objectives of the initiator of the process mining project. *Figure 5.6* describes the different types of evaluation and the role of the person that is usually necessary to do this evaluation in a process mining project.

Figure 5.6 Verification, Validation and Accreditation

A13: Verify the modelled work

Verification is a technical assessment based on the outcome of the applied techniques. Not every process model represents a correct process. A correctness criterion that can be checked using verification is soundness. Soundness guarantees the absence of deadlocks, activities that

can never become active, livelocks and other anomalies that can be detected without domain knowledge [Aalst 11c]. This type of evaluation is related to the process mining related activity 'diagnose' as described by the 'Refined Process Mining Framework', *figure 5.4*.

A14: Validate the modelled work

The validation of the modelled results is an evaluation of the degree to which the model represents the real process. The mined results can be checked to determine if there is some business reason why this model is inappropriate. Models could for example be difficult to work with for the business because of its language or complexity. Another reason could be that the model represents only a part of the process instances or allows for more process flows than exist in the process. [Rozinat 07] have outlined an evaluation framework intended to enable end-users to evaluate the validity of their process mining results, *figure 5.7*. Furthermore they introduced evaluation dimensions for discovered process models.

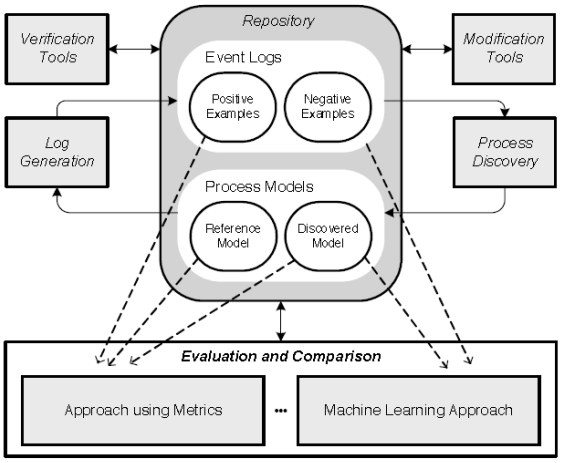


Figure 5.7, Process Mining evaluation framework [Rozinat 07]

There may be several approaches to compare and evaluate process models. Given similarities between process mining and the Machine Learning domain it appears to be beneficial to consider existing validation mechanisms used in the Machine Learning community. The quality of a process model can also be direct evaluated using metrics, which can take place in different, orthogonal dimensions [Rozinat 07]:

- Fitness: indicates the degree of observed behaviour that is captured by the process model
- Precision: addresses overly general models, by evaluating the allowance of traces.
- Generalization: addresses overly precise models, by evaluating the allowance of traces.
- Structure: is determined by the language of the model, that should express the model in a suitable presentation.

An illustration of these evaluation dimensions on different models is shown in *appendix D*.

In more recent work, [Huang 08] presents a systematic approach for developing quality metrics for block structured process models. A 'badness score' is calculated for each model to determine the quality of the model. Models that have more self-loops, optional tasks and blocks are "more bad" than models with fewer of them. [Huang 08]. Axioms that apply to calculate the 'badness score' are related to quality dimensions as described in [Rozinat 07].

A15: Accreditate the modelled work
Accreditation, the assessment of the degree to which the process mining results meet the objectives, is another way of evaluating that is required. The initiator of the process mining project, which will be often the process owner, should evaluate if the models that are generated using process mining techniques are of interest for the goals of the process. Next to answering questions that are proposed upfront, also other insights that are delivered by the model could be important for the initiator of the project.

A16: Decide on an elaboration of the process mining project
In this evaluation phase, one has also to think about a possible elaboration of the process mining applicability. Given the dynamic nature of processes, it is not advisable to see process mining as a one-time activity [Aalst 11b]. Therefore to keep improving the process, projects should initiate new projects or process mining should be applied on long-term basis. Based on the results of the process mining activities more specific types of analysis, i.e. drill down, that add value to the objectives can be done or new questions could raise which require another event log. Furthermore, process mining can also be applied for a longer term, i.e. monitoring the process or supporting the operational process. Monitoring means that the state of the process is analysed at several moments in time. This can be interesting to indicate the influence of decisions/adjustments made during the process. Operational support is considered with pre-mortem cases and these cases can, potentially, still be influenced. Operational support corresponds to three activities described in the 'Refined Process Mining Framework', *figure 5.4*, namely 'Detect', 'Predict' and 'Recommend'. The output of operational support does not need to be interpreted by process mining analysts, because results can be deployed directly (and eventually automatically) in the process if actions are declared beforehand [Aalst 11a].

5.6 Deployment

A17: Identify if and how the process can be improved by improvement actions
A process mining project aims to add value to the process that is analysed. This can be done by evaluating e.g. if what is expected is true, or if the results deliver new insights by presenting unknown information about the process. However, the project results can also directly influence the process by deploying improvement actions. This can be done independently of the IT system, e.g. by adding more people to the process or train people to work more efficient or effective, but improving the process can also be done by influencing the way cases are handled by the information system, e.g. redesigning the workflow or applying rules to certain activities.

Process mining can result in one or more of the following improvement actions [Aalst 11a]:

- Redesign: permanent change to the software or model, e.g. execute sequential activities in parallel or introduce a four-eyes principle.
- Adjust: realizing adjustments to the process without changing the software or model, e.g. allocate more resources to the process, lower the threshold for delegation.
- Intervene: interventions in the process for particular cases or resources, e.g. aborting long queuing cases, discipline workers that violated compliance regulations.
- Support: supporting the operational process, e.g. predict the remaining flow time of cases, recommend the actions with the lowest expected costs.

Improvement actions can be divided along two orthogonal dimensions, decision level and (non)technical actions. r*edesign* and a*djust* have to do with strategic or tactical decisions. These actions which stand apart from the specific cases that are currently in process, e.g. execute sequential activities in parallel, allocate more resources to the process. *intervene* and *support* improve the operational process and can be done quick. These actions determine how a specific activity is actually done, e.g. abort cases that have been queuing for a long time, suspend an employee, predict the remaining flow time of a case and recommend the action with the lowest costs. The second dimension is (non)technical, which groups on the one hand *redesign* and *support*, which have technical implications (software or model). Actions *adjust* and *intervene* on the other hand, improve the process by changing the process in 'soft', nontechnical terms. An overview of the improvement actions is given in *figure 5.8*.

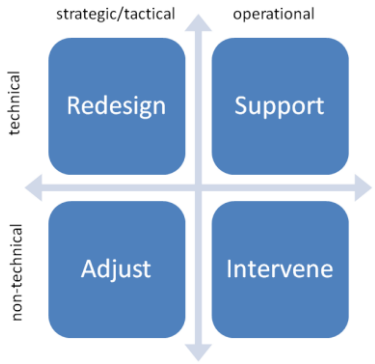


Figure 5.8, Process improvement actions

Using the evaluated results, one or more of the introduced improvement actions can be identified that can be helpful for the business to meet their objectives and improve the process.

A18: Present the project results to the organization
Having executed all actions in a process mining project and generated an output that is valid according to the objectives of the project, there is one last action remaining: presenting the results. A summary of the project, the process insights gathered during the project and the proposed actions to improve the process have to be presented to the business, so that this information can be used to improve the process.

5.7 Chapter Conclusion

The aim of this chapter was to answer sub research question 2.

SRQ2: *What are the required activities to perform a business-driven process mining project and what should be the order of these activities?*

In the different sections of this chapter, driven by six high-level functional requirements described in the former chapter, the complete set of actions that are required to perform a business-driven process mining project are identified. The activities are described in the usual order they are executed in a general process mining project. However, this does not mean that a project should always apply the activities in this order. An overview of all activities and their corresponding phase can be found in *table 5.1*.

6. Design Synthesis

The Design synthesis is a creative activity that develops the actual guideline approach. In this last activity of SEP the methodology will be created that supports the defined scope and combines the functional requirements as described in *requirements analysis* and *functional analysis and allocation.*

6.1 Life-cycle

Section 4.8 describes the main functional requirements. These requirements refer to the six main phases of the methodology:

1. Scoping
2. Data understanding
3. Event log creation
4. Process mining
5. Evaluation
6. Deployment

Normally, these phases are executed in ascending order. However, iterations are possible between these phases. Disappointing results during the *data understanding* phase can require a return to the scoping phase for example because of a lack of- or unreliable data. Applying extra process mining techniques can require an adaption of the data e.g. creating a derived attribute. The evaluation of the process mining results may give a need for other or more profound types of analysis, for instance models that are not representative for the process or not meet the objectives of the initiator. Furthermore, the evaluation may show the need for an elaboration of the project, for what the objectives can be formulated during the business understanding phase.

Taking these iterations into account, it is possible to design the life-cycle of process mining projects, which can be found in *figure 6.1*:

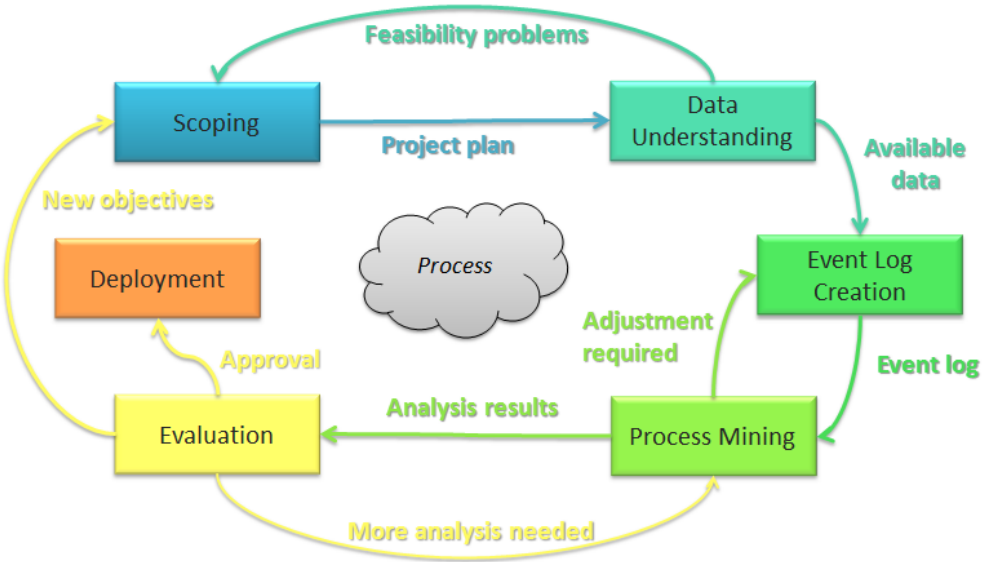


Figure 6.1, Life-cycle of process mining projects

In the presented process mining project life-cycle (PMPL) an overview of the relationships between the different process mining project phases is given. The arrow description gives the output/input of the phases. The product of scoping is a plan of the project, which should be checked on feasibility with the available data. The feasibility check of the data may give problems that result in an inability to meet the project objectives. Using the available process data, an event log must be created that is appropriate for applying the required process mining techniques. During the process mining phase adjustments of the event log can be necessary to do all required types of analysis. After the mining phase, the results of the data analysis are evaluated and may require extra analysis, an elaboration of the project or the results can be approved so that deployment of the results is possible.

It is important to note that PMPL is a high level overview of the life-cycle of a process mining project and does not identify all relationships. Depending on the goals, the organization, the process and the data, more relationships could be possible in a specific project, including relationships between the tasks of different phases. However, the arrows between the process mining phases in as described in PMPL are the most important and frequent dependencies.

6.2 Methodology

The previous chapter described 18 main activities that are required in a process mining project. These activities are classified in six different phases. *Figure 6.2* outlines a summary of the methodology containing the phases and the main activities in key words.

Figure 6.2, summary of Process Mining Project Methodology

A more detailed description of the Process Mining Project Methodology (PMPM) can be found in *appendix E*.

6.3 Verification

In this section the proposed methodology, PMPM, is compared with the requirements. All the requirements that are formulated in chapter 4 should be met in PMPM. Questions that the methodology must be able to answers as described in *section 4.1* are the following:

- What are the main phases of a process mining project?
- How can process goals be aligned with the application of process mining?
- What parts of the process are suitable for process mining?
- What kind of process data is needed?
- What must be done to use exported process data as input for process mining?
- What types of analysis can be done using process mining?
- How can the analysed results be deployed in an organization?

- What are main activities in a process mining project?

Answers to several of these questions can be found in *figure 6.2*. The answers of the other questions are descriptions of specific activities of the methodology. A description of the specific activities can be found in *appendix E*.

Next to customer requirements, also several operational scenarios as described in *section 4.4* should be guided by the methodology: 1. Discover the control-flow model of an unstructured process, 2. Check the four eyes principle on activities in a process, and 3. Extend the resource activity model with the time aspect. For each of the scenarios a list of activities is described that should be supported by the methodology. In the following table these activities are summarized and aligned with the corresponding activity of PMPM:

Activities operational scenario's	PMPM activities
Develop understanding of the main process	Identify process, Determine objectives
Identify and gather the required data	Locate data, Explore data, Select data, Extract data
Prepare the required data to apply analysis techniques	Prepare data
Apply suitable mining techniques	Answer questions
Analyse the output	Verify, Validate
Ensure that the output is useful	Accreditate
Present the results	Present results

Table 6.1, aligning scenario activities with activities of PMPM

As can be seen in *table 6.1*, PMPM supports each activity of the operational scenario's with one or more activities that are included in the methodology.

The functional requirements that are described can be directly found in *figure 6.2* and the methodology is proposed within the boundaries set in *section 4.6*. The life-cycle of the development requires one more phase: 'testing' of the methodology, which will be described in the next chapter.

6.4 Methodology Comparison

As described in *section 2.3*, other methodologies as KDD, CRISP-DM, PDM, BPA-H, and L* have several shortcomings in guiding process mining projects [Heijden 11]. To emphasize the added value of PMPM, these identified shortcomings should not apply to the new developed methodology. Below is *table 2.1* extended with the new methodology.

Methodology	Domain	Driven by	Process-specific
KDD	data mining	business	No
CRISP-DM	data mining	business	No
PDM	process mining	data	No
BPA-H	process mining	data	Yes
L*	process mining	business	Yes
PMPM	process mining	business	No

Table 6.2, methodology characteristics (table 2.1 extended with PMPM)

- PMPM is business driven. The first phase, *business understanding* provides support to identify the objectives of the organizational process and the last phase *deployment* supports in transferring the results of the project to the organization to improve the process.
- PMPM is tailored to process mining. The methodology describes specific activities that are needed in a process mining project, e.g. how to select and extract data, create an event log and guidance in process analysis.
- PMPM is appropriate for all organizational process mining projects since all phases and activities should/can always be performed in a business process mining project and the methodology is not process specific.
- Unlike L*, this new methodology is not concrete in the order of specific process mining related activities that should be used, because this does very much depend on the available data, knowledge, process maturity and the goals of the project.

6.5 Chapter Conclusion

The aim of this chapter was to answer sub research question 3.

SRQ3: *What can be an appropriate methodology according to requirements from SRQ1 and SRQ2 and that does not have the shortcomings of the other methodologies?*

PMPM is developed with the help of the Systems Engineering Approach. Which included three main activities: 1. Identifying main requirements and constraints, 2. Identifying the low functional requirements, and 3. Designing the methodology by synthesizing the requirements. Subsequently, the developed methodology is verified by the formulated requirements and evaluated on the shortcomings of other identified methodologies. Therefore PMPM (*figure 6.2*), including its life-cycle (*figure 6.1*) and its detailed description in *appendix E*, is an answer to the third research question.

7. Practical Evaluation

This chapter describes a case study that was conducted at the department of Financial Services (FS) of Rabobank Nederland using the PMPM. The purpose of the case study was to validate the new methodology as being a valuable approach to apply process mining to an organizational process. During the case study the methodology was evaluated on the activities of PMPM that were executed, if phases/activities were missing, the order of the phases/activities, and if the iterations that took place between the phases during the project are described in the life-cycle of the methodology. The results of this evaluation are described in the last section of this chapter to identify the value of the methodology. Furthermore, the case study as described below can also be seen as a demonstration how to use the methodology in a business process mining project. This case study was guided by sub research question 4 of the research design.

7.1 Organizational Introduction

The Rabobank Group is a Dutch financial services provider with offices in 44 countries and employs worldwide about 60.000 FTE[10]. The organization is among the top 30 largest financial institutions in the world, and a global leader in Food & Agri financing and in sustainability-oriented banking. Rabobank Group comprises 141 independent local Dutch Rabobank affiliates, central organization 'Rabobank Nederland', 'Rabobank International' and subsidiaries. Rabobank reported net profit of EUR 2.6 billion for the 2011 and total assets amount to EUR 665 billion[11]. The organization serves globally about 10 million customers and has a top rating for creditworthiness[10].

Rabobank Nederland is a cooperation of the 141 local banks, which are the members of Rabobank Nederland and have equal voting rights. Rabobank Nederland employs about 10 percent of the FTE of Rabobank Group[12]. The FS department, part of Control Rabobank Group is a division of Rabobank Nederland, concerns about credit collection and debt payments. Next to Rabobank Nederland, credit and debt management is also done for Rabobank International and most (135) of the Rabobank affiliates.

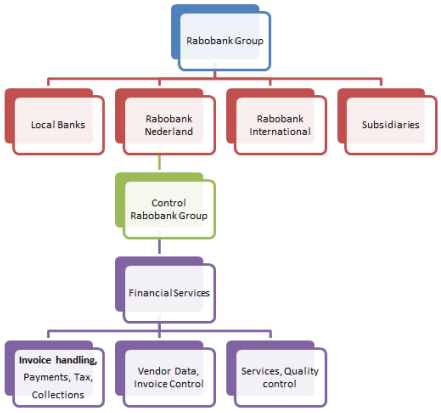


Figure 7.1, Organizational structure Rabobank

7.2 Scoping

Identify Process

The concerning process in this case study was the invoice handling process of the organization. The invoice process starts by a customer who sends an invoice, either physically or digitally. Digital invoices, which can be either an e-invoice or a PDF invoice, are managed in London. A paper invoice is normally directly sent to Rabobank Nederland where invoices are combined and forwarded per flight to Riga (Latvia). In Riga, all physical invoices are digitalized by scanning. A company in Malmö (Sweden) analyses the scanned invoices to recognize the information on the document (OCR). The next activity for all invoices including the digital ones is to register the invoice, which is done at FS by a 'central administrator' (CA). Registered invoices need to be checked by an 'invoice inspector' (FC, in Dutch: 'Factuurcontroleur') who checks the details. For

10 Rabobank. Retrieved August 3rd, 2012, from http://www.rabobank.nl/particulieren/servicemenu/over_rabobank/
11 Wikipedia Rabobank. Retrieved August 3rd, 2012, from ttp://en.wikipedia.org/wiki/Rabobank
12 Wikipedia Rabobank Nederland. Retrieved August 3rd, 2012, from http://nl.wikipedia.org/wiki/Rabobank_Nederland

local banks the registered invoice may go to a 'goods inspector' (MC, in Dutch: 'Materieelcontroleur'). When the invoice details, and if necessary the goods are checked, the invoice can be approved by a 'budget holder' (BH). The last activity in the process is the payment of the invoice. Not all activities are executed at FS, the three checks (details, goods, approve)

Not all activities are executed at FS, the three checks (details, goods, approve) are done by employees working for the department that is concerned with the specific invoice. One of the targets of the invoice process is to process all invoices, ready for payment, within 14 days after the invoice data. A summary of the main statistics of the invoice process can be found in *table 7.1*.

Amount of invoices	550.000 invoices/year
Suppliers	107.000
FTE	40 (central administration)
Payments	2.5 billion/year
Productivity	13.500 invoices/year/FTE
Outsourcing	Scanning, Optical Character Recognition (OCR), PDF, e-Invoicing

Table 7.1, statistics of invoice process FS

Process mining requires (partially) logged events as input in order to retrieve process information. The activities 'register', 'check details', 'goods check' and 'approve' are executed in an ERP system, SAP. Probably all transitions of the invoices that are done using SAP are logged. A graphical overview of the invoice process is given in *figure 7.2*. Other activities that may be logged are not feasible (e-Invoice/PDF, Scan, OCR) or not interesting (Pay) for this process mining project.

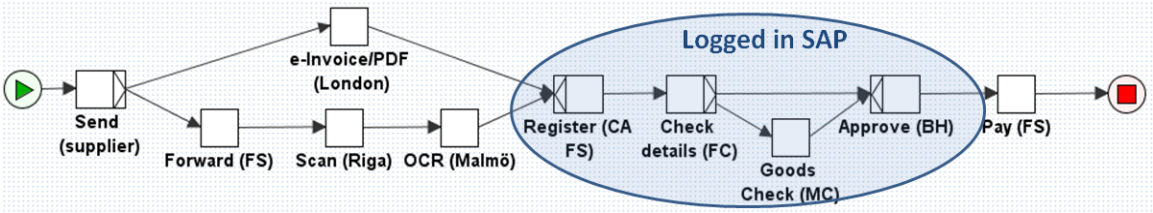


Figure 7.2, overview invoices process (happy flow) and logged parts

Determine Objectives

In dialogue with project initiators of Rabobank Nederland, several process mining project goals were formulated using the strategy and tactics of the department. An overview of the goals and derived questions and metrics can be found in *appendix F*. These questions and metrics assisted in clarifying the objectives and guide the project. For each of the goals the required event aspects were derived using the questions and metrics. An overview of these results is given in *table 7.2*. 'first time right' means that an invoice is not sent back to a department (role) where the invoice has been before, e.g. when mistakes are made in the registering activity, a CA has to confirm changes of a FC.

Objective	Goals	Aspects
Process control	discover the flow of invoices in the process	case (id), activity (id), time (order)
Process efficiency	improve productivity of employees	case (type), activity (id), resource (id), time (start + end)
Risk control	role segregation	activity (id), resource (id), resource (role), activity-role (activity (id) and resource (role) combined),
Process quality	increase 'first time right'	case (id), activity (id), resource (id), resource (role), time (order), motivation (id)

Table 7.2, Objectives of the project and its required event aspects

Using the information in *table 7.2* and the basic knowledge that is gathered about the organizational process, the operational scenarios were formulated, *appendix F*.

Determine tools/techniques
The following table gives an overview of the tools and techniques that were planned to be used during the project:

Tools/techniques	Name	Motivation/Techniques
ERP system	SAP	Extract data to a spread sheet file
Software	MS Excel	Event log creation and basic types of analysis
Software	Disco	Process discovery, log statistics, filtering
Software	ProM 5.2	Role hierarchy miner

Table 7.3, Overview of expected tools and techniques to be used

7.3 Data Understanding

Locate data
After clarifying the scope of the project, the next step was to identify if the required data was available in SAP. A positive coincidence in this case was a few years ago a project identified the SAP table that was considered with the invoice process. A small data dump (20 minutes) showed that the data indeed contained event information of the process, *appendix G*. Besides, the dump also indicated that it was not feasible to export a dataset of more than one day. The huge amount of the dump would result in performance problems of the system and probably a time out. Another option was to check if the event data was available in the business warehouse, which was indeed the case, although in a different format. However, without building work the same performance problems would occur for exporting the huge amount of data. Because of several issues, the building work was not possible within the project time. The last possibility to retrieve data was using data located on a test platform, which contained the same type of data as the business warehouse.

Explore data
Exploring the data on the test platform indicated that all event data on this platform had two big problems: First of all, the data was outdated, the most recent events were more than one year old, and secondly it was not possible to export more than a certain amount of data, about 300.000 events. Outdated data has a negative impact on the reliability of the data, because the process could have been changed in the last year, which is indeed the case. Furthermore, a 'small' dataset has a negative impact on the reliability too. Therefore the value of this project for the organization can be questioned. However, since the aim of this case study was to evaluate PMPM and to demonstrate its use, the problem that the data was outdated and that the set of data was small were therefore considered to be acceptable so that it still was possible to perform the case study.

Two interesting tables were identified on the test platform, one containing event information and one containing case details. An impression of the type of information contained in these tables is given in *appendix G*. The required aspects can be found in the tables as following:

- Event table: case (id), activity (id), resource (id), time (start + end)
- Case table: case (id), case (type)

Aspect ‘case (id)’ exists in both tables and could therefore be used as a connection between event and case information.
Aspects ‘resource (role)’, ‘activity-role (activity (id) and resource (role))’ and ‘motivation (id)’ cannot be found in the data. However, ‘resource (role)’ and ‘activity-role (activity (id) and resource (role))’ could be derived using available data and knowledge of employees. ‘Motivation (id)’ is not present and could not be derived. Because of the lack of a motivation aspect for identified iterations, the project plan was not feasible. Together with the initiator of the project, the GQM of goal ‘Process Quality’ was adapted, see *figure E5*, so that aspect ‘Motivation (id)’ was no longer required.

The event data was organized differently than expected. Not the activities are recorded, but the time between the events. This is the time that an invoice is waiting in an inbox to be handled. It was possible to derive the time and resources of the activities using this information, though.

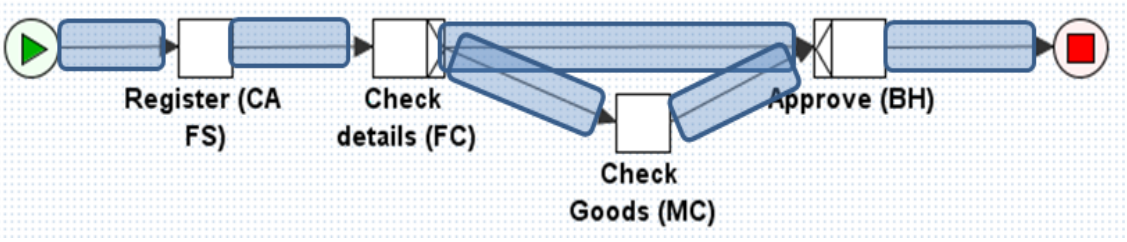


Figure 7.3, Logged events in the process

Verify data
A few of the invoices in this process were analysed more profoundly in the ERP systems, which gave the idea that the data was trustworthy. However, it could be that there were events missing since data of a test platform was considered and maybe not all data is copied to this location. Therefore it could not be verified if the data was complete. The semantics of the data looked well-defined, no reason to think differently. Finally, the data appeared to be safe in terms of privacy, since names of resources and senders of invoices are given in numbers, not in names. Nevertheless, since it could be possible for employees of the Rabobank to find out the real names of employees at the process, the employee numbers are converted to a new number.

7.4 Event Log Creation

Select data
Process mining analysis needs to be done with appropriate data.
Considering that the data was outdated and that there was a maximum amount of events for this selection, events were chosen by selecting the most recently finished cases. Besides, to meet the objectives, only historic cases were selected. All required available aspects of this data were selected. A summary of this selection can be found in *figure 7.4*.

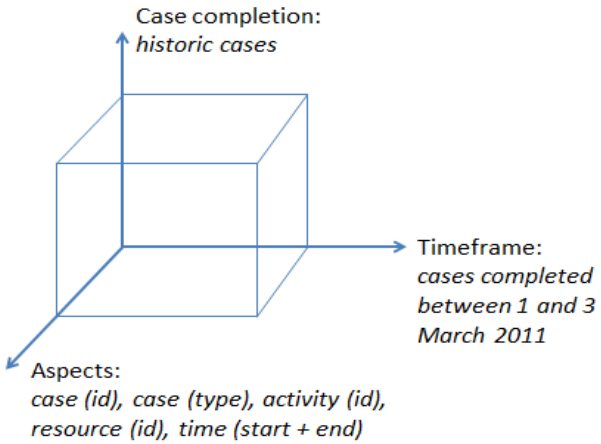


Figure 7.4, Selection of data

Extract data
The selected data was extracted to an MS Excel file directly by the SAP system in a file containing event information and a file containing case information. An impression of the extracted files is given in *appendix H*.

Prepare data
'Resource (role)' and 'activity-role (activity (id) and resource (role))' were derived from this data since these aspects were not directly available in the extracted data. Business indicated that there were five main roles: three already given (CA, FC, BH) and two other ones, Project Leader (PL), Supplier Creator (SC). For convenience the not directly available aspects were derived later in the project by using a process mining technique. Since objectives 'Process Discovery' and 'Invoice Efficiency' did not need aspects 'resource (role)' and 'activity-role (activity (id) and resource (role))', meeting those objectives was possible without the missing aspects.

To create an event log that is suitable as input for process mining analysis, the two tables were merged by 'Case (id)', date and time are combined to one timestamp and the inbox tasks were transformed to activities that were executed by resources. An impression of the extracted files is given in *appendix H*.

7.5 Process Mining

Familiarize log
The process mining phase starts with gathering basic statistics of the information contained in the event log, *figure 7.5*. Almost 5500 cases were contained in the log with an average of 5 events per case. Since the amount of average events per case was quite low, the control-flow model was probably 'Lasagna-like'. However, the amount of activities, 36, was rather high compared to this average of 5. The timeframe of the data was about half a year, 20th of September 2010 to the 3rd of March 2011 and the amount of resources handling the invoices was very large, more than 2100 people. Furthermore, the overview of the data also showed that 90 percent of the activities had a duration of less than 15 minutes. Business indicated that activities that last much longer did not sound plausible, certainly not if it took more than one hour to complete. Therefore, all cases containing activities that lasted longer than one hour were filtered, *appendix I*, which resulted in a subset of 95% of the cases.

Events	27.141
Cases	5.420
Activities	36
Resources	2.114
Attributes	3
Start	20.09.2010 13:12:04
End	03.03.2011 22:57:03

Figure 7.5, Summary log statistics

Ensure structuredness
Since the log inspection indicated that the process was probably quite structured, it was probably not necessary to apply extra structuring techniques. However, the first control-flow model resulted in a 'Spaghetti-like' model, *appendix I*. Since it is difficult to derive information from such an unstructured model, the model needed to be made more 'Lasagna-like'. Creating more structure was done by filtering all sequences of event that were not shared by at least five invoices, which resulted in a subset of 79% of the original invoices.

Answer questions

Insight

The control-flow model of the subset, containing 79% of the invoices, was interpretable and contained the frequencies of all activities and relations. However, also more structured control-flow models were added that showed the most frequent activities and relationships, which was done for the convenience of business, *appendix J*.

Process Efficiency

Using the same 79% subset, both the average handling time and the impact (total) of the handling time per activity could be generated in a control-flow model containing time, appendix K. Activity 'New' and 'New CA' had the most impact, which was respectively 40% and 37% of the total handling time.

There were twelve types of different invoices of which four types contained 96% of all cases. A summary of the percentage of cases, average events per case, average handling and impact on handling time is given in *appendix K*. The main duration of different resources for the high impacting activities 'New' and 'New CA' is available in *appendix J*.

Risk Control

All employees on the invoice process of Rabobank Nederland have a dedicated role. An employee may not execute activities of more than one role. To analyse if there exist overlaps in these roles by certain employees, the whole event log was used. With the help of business the resource role of each activity was identified, which may be a combination of roles, see appendix L. Appling the Role Hierarchy Miner indicated that there were several resources with conflicting roles (indicated by circles), see *appendix L*, especially between BH and MC.

Process Quality

Not all invoices are directly approved. Invoices can be sent back to a department (role) it has been before if invoice details should be changed or the invoice is wrongly allocated. It is interesting to identify which invoices iterate, because this can often be considered as a mistake of a resource. Therefore iterations that go back to the CA role and their frequencies were identified, just like the resources that initiate the iterations and the resources that executed the activities before the invoice was sent back. In this process returning invoices get another inbox name, therefore iterations should not be identified by arrows but by activity name. These activities are 'Critical Data' and 'Return from FC' the control-flow model containing these cases and the resources handling these iterated cases can be found in, *appendix M*. The whole event log was used for this objective.

7.6 Evaluation

Verify

The results were evaluated to measure its accuracy. This evaluation started with transformational accuracy, assessing the model quality in technical terms. The first technical evaluation was done for the control-flow model containing 79% of the cases, *figure J1*. The generated model represented an sequential process, which made it easier to assess. The following checks were executed:

1. Each activity has a next activity or is connected to the end place,
2. Each activity has a former activity or is connected to the start place
3. All cases that start the process and activities also finish them, activities have the same frequencies for in- and outgoing arrows

All three checks were verified for the model of *figure J1*. The next two models, *figure J2* and *J3*, were not correct according to the last criteria, but these models are adaptions (activities and relationships were removed for the convenience of the FS manager) of the model of *figure J1*. The models of *figure K1, K2, K3* and *K4* do not have frequencies, therefore check 3 was not necessary and both check 1 and 2 were verified. *Figure K1* shows resources and their accompanying roles. A resource may not exist more than once in the model, which cannot be verified since not all resources are shown. Nevertheless, the total amount of the resources sounded with the log inspection results and none of the shown resources is shown twice in the model. The last model*, M1*, does not show all relations for convenience, but check 1 and 2 are verified.

Validate

The second evaluation step is validation, checking if the results are representative for the process. The subset fitted 79% of the cases, 21% of the cases were excluded from the log. 5% of these cases were filtered because they contained activities that took more time than 1 hour to complete. A more thorough investigation of these cases showed that this would probably have two main causes: First, an activity of the invoice was indeed open in SAP for the whole time, but nobody was working on it. Secondly, not all activities were properly logged, as a result the end timestamp of the activity was the end timestamp of a next activity. Therefore excluding these cases from the process gave a better view of the performance of the process than including them.

18% of the cases were excluded because their flows have a low (less than 5 cases) occurrence (2% were already excluded because of longer activity duration). By excluding cases with low occurring flows, 15 of the 36 activities were excluded, which did not stimulate the mapping of all possible paths in the flow. However, the project initiator wants to have a general overview of the process flows and including low frequent activities does not stimulate the simplicity of the model. For the goal 'Process Efficiency' the same subset was used, which is valid because of the same reason. The 79% subset was also used for 'Process Quality', since it excludes low occurring behaviour. The goal was to identify people that make mistakes while keeping in mind that not every iteration is a mistake. The chance that this is not a mistake is bigger in cases with a 'strange' flow. For 'Risk Control' the whole log was used. This goal aimed at detecting behaviour in the form of resources that were able to execute a certain combination of activities, which requires the whole log. Using the results, the profiles of these resources are checked.

Accreditate

Accreditation is the third form of validation that is performed. Accreditation is checking if the process mining results meet the objectives. The metrics as given in *appendix F* are all given by the results. Besides checking the formulated objectives, the results were also propounded to the initiator of the project. Generally the results were giving astonishment about the possibilities of the process mining analysis. Even though it was a pity that the motivation for the iterations of goal 'Process Quality' was missing in the data, the results were however very valuable.

Determine on elaboration

The last evaluation step is determining on elaboration of the process mining project. The results were raising new questions and the results proved the possibilities of the process mining techniques for the organization. Therefore an elaboration was desirable. Although, since the logged data of the process was not available for recent events and huge amounts, the priority

was first to solve this problem before a decision is made on an elaboration the process mining project.

7.7 Deployment

Identify possible improvements

The last phase in the process mining project is deployment, presenting the results to the organization to let them improve their process. This phase starts with analysing how the objectives can be met using the results. Since the first objective, 'Insight', was only concerned with developing knowledge about how invoices are handled in the process, no advice was given on this topic.

The second objective 'Process Efficiency' targeted to increase the average productivity of resources by spending less time in handling cases by the FS department. A high productivity is possible by increasing the amount of cases that is handled by the main flow or decreasing the average duration of activities, especially activity 'NEW' has huge influence on the total handling time of employees of FS. Decreasing the handling time for invoice type 'ZLB_FAC' has the most impact of all invoice types. Furthermore there are a few resources that do need much more time to register an invoice than the average, these resources should further be analysed to investigate what the root causes are that are causing this low productivity and eventually to act on the result.

Objective 'Risk Control' was considered with identifying the resources that have more than one role in the process. An advice to lower the risk considered with the process, is to further analyse the resources with more than one role and give them rights to execute only one of the roles.

The last objective 'Process Quality' identified the relationship between resources and their occurrence with iterations of cases in the process. There is one resource that is related to a huge percentage of iterated cases in relation to the registered invoices, *appendix N*. Probably this resource makes many mistakes in registering invoices or registers a difficult type of invoice. The organization should verify the root cause and investigate how the iterations can be prevented.

Present results

In a session with the initiator of the project, the project was summarized including the results and the improvement recommendations as described above. The initiator recognized the control-flow model and agreed that it indeed represented the core flow of invoices. Frequencies for the flows as given in *figure J2* and *J3* were not surprising. The handling times of activities and employees were very interesting for the organization, since they never measured them before, although they had some ideas. These handling times of the work force will be used to evaluate the employees. Furthermore, Rabobank decided to check all resources that probably had too much authorization according to *figure L1* and they confirmed that these resources indeed had too many rights. The resource that did relatively often register invoices that were sent back to the registration department (EMP77984) as can be seen in *figure N1*, was indeed someone who made much mistakes. The many mistakes that this employee made were the main reason that this person was fired just one month before this cases study was performed. Currently, the organization is investigating how the logged data of the process can be made up-to-date and in huge amounts available to do more process mining analysis in future.

7.8 Chapter Conclusion

The aim of this chapter was to validate PMPM, which was the aim of the fourth sub research question.

SRQ4: ***How does the proposed methodology perform in a business process mining project?***

The business process mining project was started from scratch and precisely executed according to the description of PMPM, and if possible in the chronologically given order. All the activities of PMPM were performed and none of these activities was experienced as a redundant activity. No extra activities were executed during the project that are not described in the methodology and that should be added to the methodology. The phases and activities were performed in the same order as described in the methodology and no problems were experienced with this order, although some iterations took place. Besides the chronologic life-cycle arrows, also the 'feasibility problems', 'adjustment required' and 'more analysis needed' arrows of PMPL were used. The lack of a motivation for iterated invoices (which was actually given by employees in the system) caused 'Feasibility problems' whereupon the project plan was slightly adapted. For each existing activity in the log, the corresponding role was added to the event log after having done some analysis already. The validation of the results required the creation of some extra models (*J2, J3, K3* and *K4*) to assist the understandability for the initiator.

During all phases of the process mining project extra process knowledge was required or desirable. Of course gathering process knowledge was done in the first phase to identify the main process and to determine objectives. In the second phase, 'Data understanding', it was needed to identify the meaning of the available datasets. The third phase required extra process knowledge to transform the inbox names to executed activities. During the analysis phase more process knowledge was needed to use appropriate filter values and to identify the activities that were related to iterations. In the fifth phase, 'Evaluation', verification was done by discussing with employees if the models were understandable. For the last phase, formulating recommendations raised some extra questions that required some clarification, e.g. further explanation of invoice type and iterations. This 'extra' process knowledge gathering during the project cannot be prevented, but it is good to realize that keeping in touch with people that have a lot of specific process knowledge is often desirable. The methodology will not be adapted by adding an activity 'gathering extra process knowledge' in every phase. Nevertheless, one has to keep in mind that that having enough possibilities to gather more specific process knowledge from the organization during the project is required. The importance of the connection with the organizational process during all phases of the project is also showed in PMPL where 'Process' is the central entity and should not be ignored.

In summary, the methodology did give much support during the project in proposing the activities that were needed and no main activities were missed during this case study. PMPM was experienced as especially useful in guiding the project since it made sure that all important activities were performed and the methodology prevented from a deviation of the project plan. PMPM did also assist in making the objectives more concrete, which supported not to get lost in a lot of opportunities during the analysis phases. Nevertheless, the other opportunities should not be ignored and can be proposed during the elaboration activity. Furthermore, thinking about the improvements forced a connection between the analysed results and added value for the organization.

8. Discussion

A challenge for research projects is to keep it valuable for science and practice, or in other words finding the balance between rigour and relevance. On the one hand, to keep scientific value, the methodology must be well grounded and supported by empirical evidence. On the other hand, practice was lacking a methodology that supported practitioners in applying process mining in their organizations and that shared the best practices of process mining specialists in the field.

The 'rigour-relevance' debate is long-lasting issue in management science, that produced several arguments for both sides. Some statements in favour of rigour are: "*a respectable objective for academic research is the development of knowledge for knowledge's sake*" [Huff 00], "*nothing is so practical as a good theory*" [Lewin 45], and "*the key quality criterion for knowledge is validity, i.e. it is deemed valid by an informed audience – the relevant scientific community – on the basis of the arguments and empirical proof presented*" [Peirce, 60]. Statements for relevance are: "*if academic research is irrelevant, practitioners will look elsewhere for solutions*" [Aken 07], "*The sheer complexity of organizations frustrates scientific research of the usual type*" [Daft 90]. And, "*a discipline aimed at actively changing reality is incomplete without prescriptive statements, especially if this change involves other people*" [Aken 07]. A methodology that describes how people should perform process mining projects in organizations cannot leave out the organizations and the people that should use the methodology to make it successful.

[Shrivastava 87] formulated eight criteria to assess the rigor and practical usefulness of research.
Practical usefulness variables:

1. Meaningfulness: The research is meaningful, understandable and adequately describes problems faced by decision-makers.
2. Goal relevance: It contains performance indicators which are relevant to managers' goals.
3. Operational validity: It has clear action implications which can be implemented
4. Innovativeness: It transcends 'commonsense' solutions and provides non-obvious insights into practical problems.
5. Cost of implementation: The solutions suggested by the research are feasible in terms of their costs or timeliness.

Rigor variables:

6. Conceptual adequacy: Well grounded, it uses a conceptual framework consistent with existing theories in the field
7. Methodological rigor: The program uses analytical methods and objective quantifiable data
8. Accumulated empirical evidence: The research program has generated a substantial amount of accumulated empirical evidence supporting it

Concerning the practical usefulness, the methodology seems to meet the different criteria of [Shrivastava 87]:

1. First, the methodology does certainly describe a problem of practitioners as also was described at Process Mining Camp 2012, and the description of the methodology is

made as understandable as possible for non-experts, e.g. excluding formulas and technical terms.

2. Secondly, the methodology starts by identifying the business objectives of processes in organizations to formulate project goals, which show the relevance for managers' goals.
3. Thirdly, the methodology contains a clear description of all activities that are needed to accomplish a process mining project and how these activities can be accomplished, except 'commonsense' activities that are 'normal' for performing organizational projects. Chapter 7 describes an organizational process and how the methodology was used to perform a process mining project.
4. Furthermore, the methodology that is described is not 'commonsense' as process mining practitioners are conducting projects differently, *appendix B.4*. Besides, inexperienced practitioners of process mining do often not know the activities that are required for a process mining project. Having an overview of the life cycle of process mining projects and to use a list of main activities as a guide during the project was experienced as very useful during the case study. The methodology prevents from a deviation of the project plan and makes sure that all important activities are performed.
5. Finally, the methodology is realistic as a guiding framework for projects in the sense that it can be used for all projects independent of the available time and budget.

According to the rigor criteria, the methodology should need more value:

6. The methodology is developed using the 'Systems Engineering Process' a framework that has shown its usefulness in designing and managing complex engineering projects. Since the methodology is validated by only one case study, no framework is used for combining empirical evidence.
7. Several existing methodologies in the data- and process mining domain that were used as inspiration during the development. Besides, the activities that are required for a process mining project are grounded by scientific literature.
8. Unfortunately, the value of the proposed methodology is not (yet) supported by a lot of empirical evidence. A single case study showed that the methodology seemed to be valuable, however, a broader investigation of the value of the methodology is required.

The methodology shows which activities must be completed to perform a process mining project, but it cannot be concluded (yet) that PMPM will be valuable to perform process mining projects more effectively and efficiently. [Rescher 77] describes two types of human knowledge, 'knowledge how', (the way of doing things) and 'knowledge that' (statements or assertions about the world). For this research project 'knowledge how' has been the primary focus. 'Knowledge that', establishing the truth that the methodology is valuable, is only validated by conversations with practitioners and a single case study. Although, the methodology is built on scientific literature and practical experiences.

[Moody 03] proposes a theoretical model and associated measurement instrument for evaluating IS design methods, the 'Method Evaluation Model (MEM)', *figure 8.1*. The model is based on the 'Technology Acceptance Model (TAM)' from the IS success literature and 'Methodological Pragmatism' from the philosophy of science.

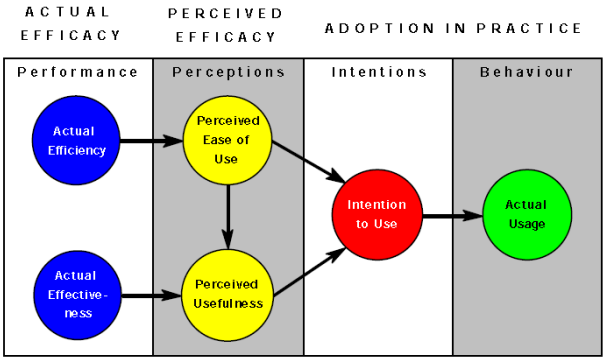


Figure 8.1, Method Evaluation Model [Moody 03]

The constructs of the model are:

- Actual Efficiency: the effort required to apply a method.
- Actual Effectiveness: the degree to which a method achieves its objectives
- Perceived Ease of Use: the degree to which a person believes that using a particular method would be free of effort
- Perceived Usefulness: the degree to which a person believes that a particular method will be effective in achieving its intended objectives
- Intention to Use: the extent to which a person intends to use a particular method
- Actual Usage: the extent to which a method is used in practice

To generate empirical evidence that the proposed process mining methodology is efficient, effective and has chance to be adopted in practice, MEM can be used. A possibility to test the methodology can be a laboratory experiment by giving one half of the participants the developed methodology and the other half no or another data- or process mining methodology. Results can be evaluated to what degree they meet the business objectives and the amount of effort that is spent to complete the experiment. Furthermore a survey can be used to evaluated the perceived ease of use and usefulness.

Another possibility is to evaluate the methodology in a field experiment. Practitioners can apply PMPM in a 'real' process mining project. At the end of the project they can answer questions about their perceptions of the methodology, for example understandability, completeness, usefulness. However, this second way to test the methodology is not measuring the actual efficacy, and the laboratory experiment does not consider 'real' projects. Therefore, of course, a combination of both types of experiments will be more reliable, e.g. laboratory for pre-testing and field for post-testing the methodology.

9. Conclusions

In response to an existing gap in the use of process mining, this research aimed at developing a methodology that described what is necessary to perform a process mining project and which supported practitioners in applying process mining in organizations. Several arguments were given in *section 1.1* that showed the need a business process mining methodology. The professional process mining community was lacking a methodology that could guide projects, that shared best practices, and that prevented from reinventing the wheel. Scientific literature did not have a comprehensive description of all required activities for completing a business-driven process mining project.

Developing the basics of this methodology meant that an extensive literature research of all required activities and solutions to perform a project had to be done. Gained experience of practitioners was used to include best practices from process mining professionals and to get insight in the requirements of the methodology. This practical knowledge was gained by a lot of conversations with practitioners, involving in discussions, attending sessions and events, and last but not least performing a business process mining project at Rabobank Nederland.

9.1 Research Questions

The research objective was to answer the main research question:

What would be an industry-, tool-, and application neutral approach for practitioners to conduct business process mining projects?

As a result, guided by four sub research questions and using Systems Engineering Approach, a framework in designing and managing complex engineering projects, the 'Process Mining Project Methodology' was developed. The development consisted of four phases: identifying the requirements, identifying the main activities of a process mining project, synthesizing all information and designing the methodology, and a practical evaluation of the proposed methodology.

The resulting PMPM is a methodology consisting of six main phases that together contain eighteen different activities that have to be performed in any business-driven process mining project. PMPM starts by identifying the business of the considered process and scoping the project, and completes by presenting the results of the project that meet the objectives and suggestions for improvement of the process. The proposed methodology does not have the shortcomings that were identified in data- or other process mining methodologies. Furthermore, the methodology was tested in a real life business process mining project and showed value as being an appropriate and useful approach. PMPL is the corresponding life-cycle of PMPM, which shows the main and most important relationships between the phases of the project. The methodology is not specific in terms of process types or industries that must be analysed, and it is also not specific in tools or applications, which makes PMPM a general approach for process mining projects.

9.2 Research Contributions

This Master's Thesis contributes to research in several ways. First, PMPM is unlike KDD, CRISP-DM, PDM, BPA-H and L* a methodology that is tailored to process mining, driven by business and not specific for a certain type of process, because:

- PMPM includes certain activities that are necessary for process mining, but not included in the data mining methodologies,
- PMPM contains a phase that identifies the organizational process and formulates business objectives is included,
- All activities as described in PMPM can and should be performed in any business process mining project.

Secondly, PMPM is an initial step to a comprehensive process mining project methodology in which all phases and main activities of business process mining projects are described and that can be used as an efficient and effective approach for applying process mining in practice. Thirdly, this research can contribute to the adoption of process mining in practice and generate attention for process mining research.

9.3 Limitations & Suggestions for Further Research

Research bears limitations, and this research project is no exception. First of all, process mining is an emerging field and therefore continuously changing. The proposed methodology was designed to cope with the way how different process mining projects are conducted. However, new developments in the field of process mining can change the way process mining projects can and should be performed.

Secondly, the development and validation of the methodology are (partly) built on the personal experiences and knowledge of the researcher. For example, at the start of the development of this methodology the invoice process at the Financial Services department of Rabobank Nederland was already known. Although the methodology is certainly not tailored to this process, the gathered knowledge could have had an effect on the development and validation of the methodology. Besides, the methodology is tested during a case study which is performed by the researcher, therefore characteristics like understandability of the methodology could not be evaluated.

Thirdly, PMPM does not provide specific support in choosing and applying the actual process mining techniques to answer the project questions. This was out of scope for this research project because of several reasons. First, available techniques are divided over several process mining software tools and the amount of techniques is very large. Secondly, techniques change over time, techniques will be elaborated, improved and new ones will be created. Thirdly, the appropriateness of a mining technique does very much depend on the available data, knowledge, process maturity and the goals of the project. However, choosing an applying an appropriate technique can be rather difficult for practitioners. Therefore, specific support for choosing an applying techniques will be an interesting topic for further research.

The fourth and main limitation of this research is that it is an initial step to come to a process mining project methodology that can be used to apply process mining in practice. Up to now there is no methodology that is able to guide all business process mining project from identifying the process and formulating objectives to the deployment of the results of the project in the organizational process. PMPM can be a methodology that is able to be an efficient and effective approach to perform business process mining project, but as also is described in chapter 8, on this moment the gathered empirical evidence is limited. The proposed methodology is evaluated on the identified shortcomings of other data- and process mining techniques, by people that have experience by applying process mining in organizations, and by

performing a case study using this methodology. Based on this feedback PMPM seems to be a valuable methodology for conducting process mining projects in organizations. Nevertheless, because of the lack of empirical evidence PMPM cannot be presented as an efficient and effective methodology. The next step is to gather empirical evidence to be able to draw conclusions on the added value of PMPM. This implies an investigation of the actual effectiveness and efficiency of the methodology. As described in chapter 8, this can for example be done in a field- and/or laboratory experiment, but also a survey can support in generating more evidence.

9.4 Chapter Conclusion

Concluding the chapter and thereby concluding the project, PMPM is the first methodology that describes the main activities that must be performed to apply process mining to an organizational process. The methodology is based on experiences of practitioners and scientific literature, and should be appropriate for all processes. PMPM showed to be valuable in a case study conducted at Rabobank Nederland. Nevertheless, the methodology has not (yet) gained enough empirical evidence to be presented as a successful way to apply process mining in organizations. Therefore, the main priority is to evaluate this methodology more extensively, and if necessary, to make adjustments so that everybody is able to experience the 'magic' of process mining.

10. Bibliography

[Aalst 11a] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

[Aalst 11b] W.M.P. van der Aalst et al. *Process Mining Manifesto*. Business Process Management Workshops, LNBIP 99, F. Daniel, K. Barkaoui and S. Dustdar, eds., Springer-Verlag, 2011

[Aalst 11c] W.M.P. van der Aalst, K.M. van Hee, A.H.M. ter Hofstede, N. Sidorova, H.M.W. Verbeek, M. Voorhoeve, M.T. Wynn. *Soundness of Workflow Nets: Classification, Decidability, and Analysis.* Formal Aspects of Computing, 2011.

[Aalst 07a] W.M.P. van der Aalst. *Exploring the CSCW spectrum using process Mining*. Advanced Engineering Informatics, 21(2), 2007, pp. 191–199.

[Aalst 07b] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, H.M.W. Verbeek. *Business process mining: An industrial application*. Information Systems 32, 2007, pp. 713-732.

[Aalst 05] W.M.P. van der Aalst, A.J.M.M. Weijters. *Process Mining*. In M. Dumas, W.M.P. van der Aalst, A.H.M. ter Hofstede, (Eds.) Process-Aware Information Systems: Bridging People and Software through Process Technology Wiley & Sons, Hoboken, New Jersey, USA. 2005, pp. 235-255.

[Aalst 04] W.M.P. van der Aalst, A.J.M.M. Weijters. *Process mining: a research agenda*. Computers in industry : an international journal, 53(3), 2004, pp. 231-244.

[Aalst 03] W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, A.J.M.M. Weijters. *Workflow mining: a survey of issues and approaches*. Data Knowledge, Eng. 47 (2), 2003, pp. 237-267.

[Ailenei 11] I.M. Ailenei. *Process Mining Tools: A Comparative Analysis*. Master's Thesis, Eindhoven University of Technology, Department of Mathematics and Computer Sciences, Eindhoven, The Netherlands, 2011.

[Aken 07] J.E. van Aken, H. Berends, H. van der Bij. *Problem Solving in Organizations: A Methodological Handbook for Business Students*. Cambridge University Press, 2007.

[Basili 94] V.R. Basili, G. Caldiera, .H.D. Rombach. *The Goal Question Metric Approach*. in Encyclopedia of Software Engineering, Wiley, 1994.

[Basili 92] V.R. Basili. *Software modeling and measurement: The Goal/Question/Metric paradigm*. Technical Report, CS-TR-2956, Department of Computer Science, University of Maryland, College Park, MD 20742, 1992.

[Bethke 03] E. Bethke. *Game development and production*. Plano, TX: Wordware Publishing, 2003.

[Bose 99] R.P.J.C. Bose, W.M.P. van der Aalst. *Abstractions in Process Mining: A taxonomy of Patterns.* In U. Dayal, J. Eder, J. Koehler, H. Reijers, editor, Business Process Management (BPM 2009), volume 5701 of Lecture Notes in Computer Science, Springer, Berlin, 2009, pp. 159-175.

[Bozkaya 09] M. Bozkaya, J. Gabriel, J.M.E.M. van der Werf. *Process Diagnostics: A Method based on Process Mining*. International Conference on Information, Process and Knowledge Management, 2009.

[Carstensen 99] P.H. Carstensen, K. Schmidt. *Computer Supported Cooperative Work: New Challenges to Systems Design.* Handbook of Human Factors, 1999, pp. 619-636.

[Chapman 00] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth. *CRISP-DM 1.0, Step-by-step data mining guide*. SPSS inc, 2000.

[Daft 90] R. L. Daft, A.Y. Lewin. *Can organization studies begin to break out of the normal science straitjacket*? An editorial essay. Organization Science, 1(1), 1990, pp. 1–9.

DoDI 96] Department of Defense Instruction. *DoD modeling and simulation verification, validation, and accreditation*. Department of Defense Instruction 5000.61, 1996.

[Fayyad 96] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth. *Knowledge discovery and data mining: toward a unifying framework*. In Proceeding of The Second Int. Conference on Knowledge Discovery and Data Mining, 1996, pp. 82–88.

[Goedertier 10] S. Goedertier, J. de Weerdt, D. Martens, J. Vanthienen, B. Baesens. *Process discovery in event logs: An application in the telecom industry*. Applied Soft Computing, 2010.

[Ha 06] T.S. Ha, J.H. Jung, S.Y. Oh. *Method to Analyze User Behavior in Home Environment*. Graduate School of Techno Design, Interaction Design Lab, Personal and Ubiquitous Computing, 10 (2,3), Springer-Verlag, London, 2006.

[Hammer 93] M. Hammer, J. Champy. *Reengineering the Corporation: A Manifesto for Business Revolution*. Harper Business, New York, 1993.

[Hand 01] D. Hand, H. Mannila, P. Smyth. *Principles of Data Mining*. MIT Press, Cambridge, MA, 2001.

[Heijden 12] T.H.C. van der Heijden. *Supporting end-users to answer different types of analysis questions by means of process mining*. Literature Study 1ML05, Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences, Eindhoven, The Netherlands, 2012.

[Herrman 09] C.S. Herrman. *Fundamentals of Methodology.* a series of papers On the *Social Sciences Research Network* (SSRN), 2009.

[Huff 00] A.S. Huff. *Changes in Organizational Knowledge Production: 1999 Presidential Address*. Academy of Management Review, 25, 2000 pp. 288–293

[Huang 08] Z. Huang, A. Kumar. *New quality metrics for evaluating process models. Proceedings*. BPM 2008 International Workshops, Springer, Milan, 2008, pp. 164-170.

[Husar 08] L. Husar. *9 tips for creating your methodology.* Retrieved August 3rd, 2012, from http://blogs.ittoolbox.com/print.asp?i=21945#.

[Irny 05] S.I. Irny, A.A. Rose, *Designing a Strategic Information Systems Planning Methodology for Malaysian Institutes of Higher Learning (isp- ipta)*, Issues in Information System, 6(1), 2005.

[Janssen 11] R. Janssen. *Increasing accessibility and reproducibility of process mining research in healthcare*. Master's Thesis, Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences, Eindhoven, The Netherlands, 2011.

[Jans 08] M. Jans, N. Lybaert, K. Vanhoof. *Business Process Mining for Internal Fraud Risk Reduction: Results of a Case Study*. 9th International Research Symposium on Accounting Information Systems. Paris. 2008.

[IEEE 94] IEEE P1220. *Standard for Systems Engineering*. IEEE Standards Dept*., NY, 1994.*

[Lewin 45] K. Lewin. *The Research Center for Group Dynamics at Massachusetts Institute of Technology*. Sociometry, 8, 1945, pp. 126-135.

[MAN 96] Media Awareness Network. *Knowing What's What and What's Not: The Five Ws (and 1 "H") of Cyberspace.* Retrieved August 3rd, 2012, from http://www.glenforestlibrary.com/pdf-files/knowyourwayaround.doc

[Moody 03] D.L. Moody. *The method evaluation model: a theoretical model for validating information systems design methods*. In Proceedings of the 11th European Conference on Information Systems (ECIS 2003), Naples, Italy, 2003.

[Nooijen 12] E.H.J. Nooijen. *Artifact-Centric Process Analysis: process discovery in ERP systems*. Master's Thesis, Eindhoven University of Technology, Department of Mathematics and Computing Science, Eindhoven, The Netherlands, 2012.

[Pierce 60] C.S. Peirce. *The Rules of Philosophy*. In M. Konvitz and G. Kennedy (eds), The American Pragmatists, New American Library, New York, 1960.

[Prince 11] R. Prince. *Business process mining success: Developing knowledge on how to successfully conduct a process mining project*. Master's Thesis, Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences, Eindhoven, The Netherlands, 2011.

[Medeiros 07] A.K.A. De Medeiros, A.J.M.M. Weijters, W.M.P. van der Aalst. *Genetic process mining: an experimental evaluation.* Data Mining and Knowledge Discovery, *14*(2), 2007, pp. 245-304.

[Pritchard 99] C.L. Pritchard. *Nuts and Bolts Series*: *How to Build a Work Breakdown Structure*. ESI International, 1999.

[Rebuge 12] Á. Rebuge, D.R. Ferreira. *Business process analysis in healthcare environments: A methodology based on process mining*, Information Systems, vol.37, no.2, 2012, pp.99-116.

[Rescher 77] N. Rescher. *Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge*. Basil Blackwell, Oxford, 1977.

[Rozinat 07] A. Rozinat, A.K.A. de Medeiros, C.W. Günther, A.J.M.M. Weijters, W.M.P. van der Aalst. *Towards an Evaluation Framework for Process Mining Algorithms*. BPM Center Report BPM-07-06, BPMcenter.org, 2007.

[Russel 05] N. Russell, A.H.M. ter Hofstede, D. Edmond, W.M.P. van der Aalst. *Workflow Data Patterns*. In Proc. of 24th Int. Conf. on Conceptual Modeling (ER05), volume 3716 of LNCS, Springer Verlag, 2005, pp. 353-368.

[Shrivastava 87] P. Shrivastava. *Rigor and practical usefulness of research in strategic management.* Strategic Management Journal 8, 1987, pp. 77–92.

[DSMC 01] Defense Systems Management College, *Systems Engineering Fundamentals*. Defense Acquisition University Press, Fort Belvoir, VA, 2001.

[Wen 04] L. Wen, J. Wang, W.M.P. van der Aalst, Z. Wang, J. Sun. *A Novel Approach for Process Mining Based on Event Types*. BETA Working Paper Series, WP 118, Eindhoven University of Technology, Eindhoven, 2004.

[Verbeek 10] H.M.W. Verbeek, J.C. Buijs, B.F. van Dongen, W.M.P. van der Aalst. *ProM 6: The process mining toolkit*. In Proceedings of the Business Process Management 2010 Demonstration Track, Hoboken, USA, September 14-16, 2010, volume 615 of CEUR Workshop Proceedings. CEUR-WS.org, 2010.

[Zachman 87] J.A. Zachman. *A framework for information systems architecture*. IBM Systems Journal, 26(3), 1987, pp. 276–292.

Appendix A: List of Abbreviations

BH	Resource role ‘Budget Holder’ of Financial Services
BPA-H	Methodology for BPA in healthcare
BPM	Business Process Management
CA	Resource role ‘Central Administrator’ of Financial Services
CRISP-DM	Cross Industry Standard Process for Data Mining
FC	Resource role ‘Invoice Inspector’ of Financial Services
FS	Financial Services department of Rabobank Nederland
GQM	Goal Question Metric
IEEE	Institute of Electrical and Electronics Engineers
KDD	Knowledge Discovery in Databases process
KPI	Key Performance Indicator
L*	L* life-cycle model for mining Lasagna processes
MC	Resource role ‘Goods Inspector’ of Financial Services
MEM	Method Evaluation Model
OCR	Optical Character Recognition
PDM	Process Diagnostics Method
PL	Resource role ‘Project Leader’ of Financial Services
PMPL	Process Mining Project Life-cycle
PMPM	Process Mining Project Methodology
SC	Resource role ‘Supplier Creator’ of Financial Services
SEP	System Engineering Process
TAM	Technology Acceptance Model

Appendix B: Sources of Methodology Requirements

B.1 Scientists

- Wil van der Aalst
- Hajo Reijers
- Ton Weijters
- Other scientists that contributed are also mentioned with their scientific work in the bibliography, chapter 10

B.2 Professionals

- Frank van Geffen
- Anne Rozinat
- People at Process Mining Camp 2012
- People that commented on the topic of this research in the process mining community of LinkedIn
- Process analysts of Rabobank Nederland

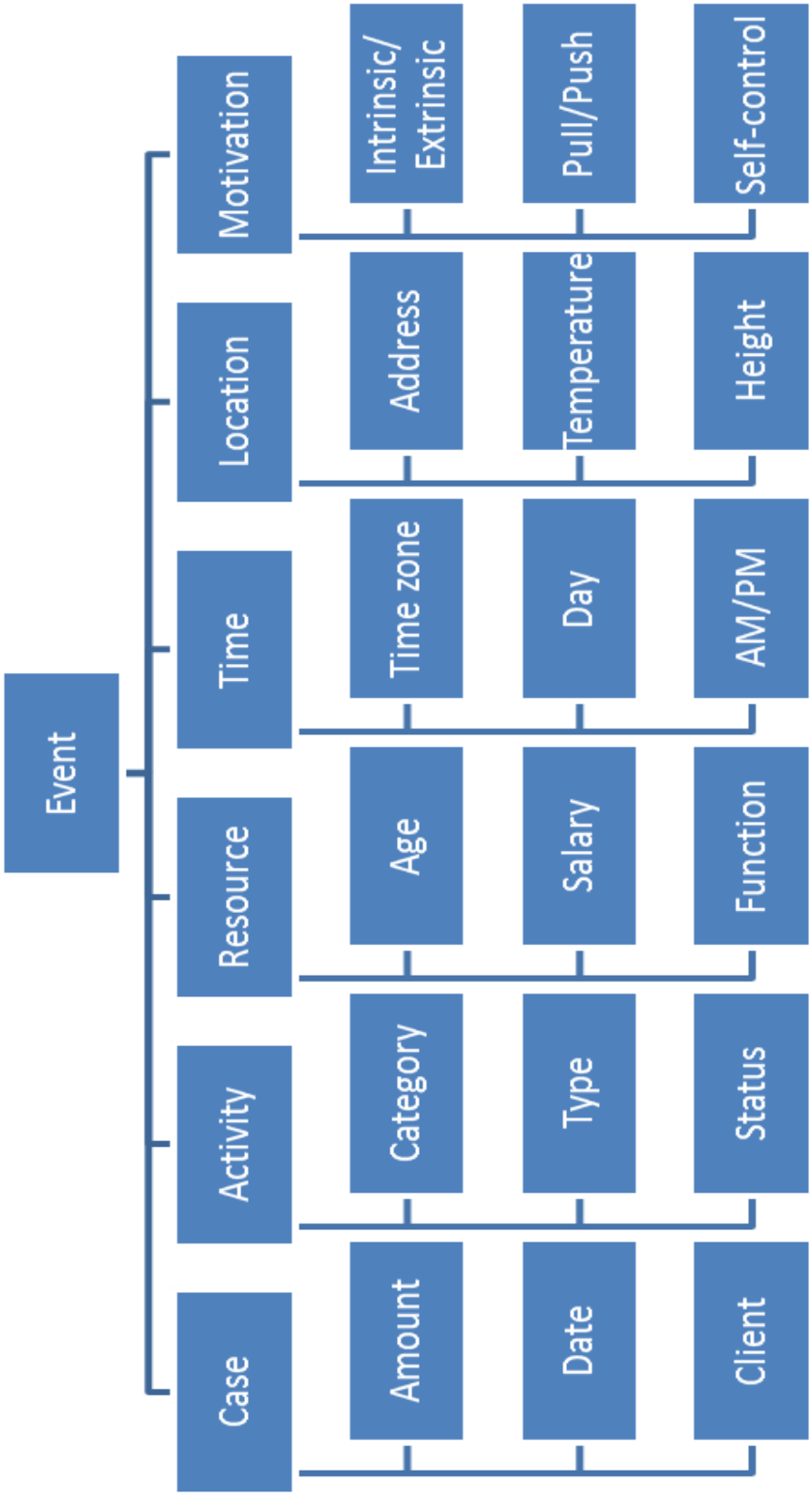
B.3 Facilitators

- Sander van den Berg
- Benedetto Paijmans
- Martijn Tabbernee
- Other managers of Rabobank Nederland

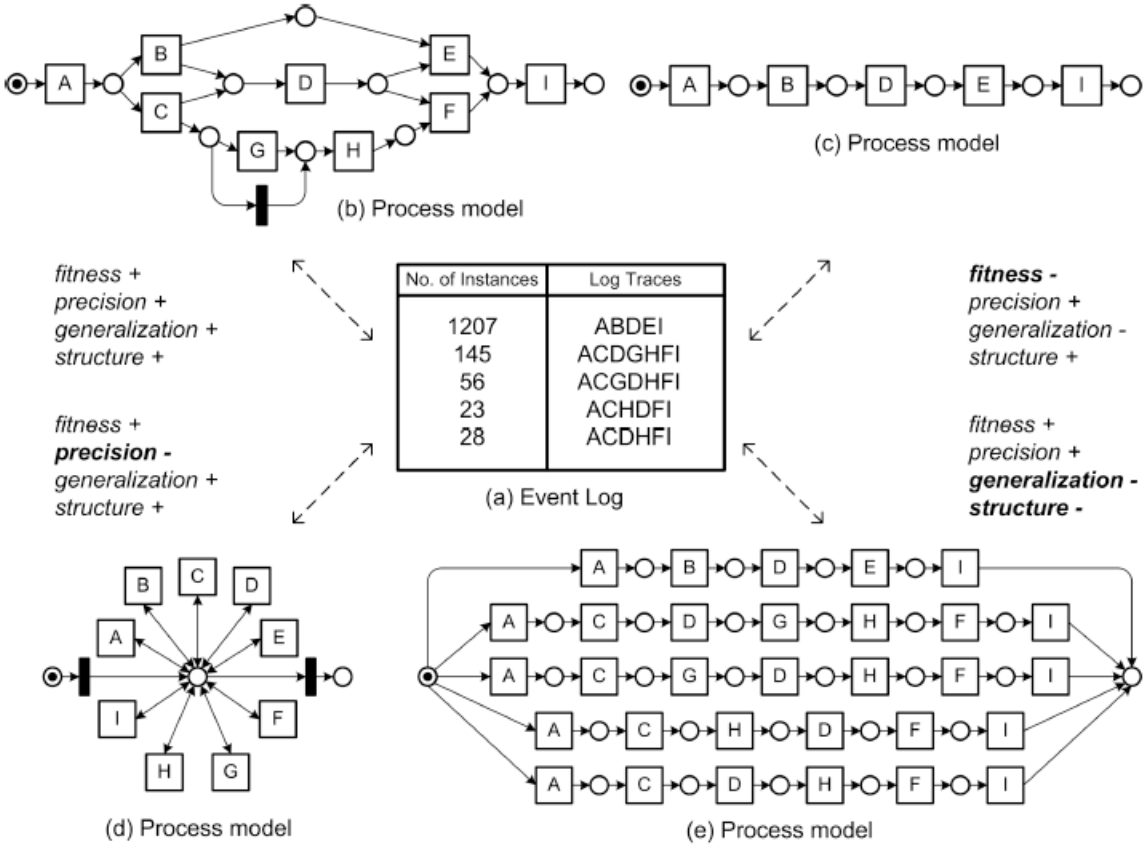
B.4 Process Mining Project Summaries and Approaches

- azM hospital, (colo)rectal cancer process, August 3rd, 2012, from http://is.ieis.tue.nl/research/bpmroundtable/slides/2012-06-11/RoundTableRonnyMansV3.pptx
- City of Lausanne, setting up an internal control system, August 3rd, 2012, from http://fluxicon.com/camp/2012/slides/camp2012-leonard.pdf
- City of Utrecht, August 3rd, 2012, from http://is.ieis.tue.nl/research/bpmroundtable/slides/2011-11-07/01%20-%20Gemeentelijke%20Subsidieprocessen%20in%20Utrecht.pdf
- Ministry of Finance, trace audit trails, August 3rd, 2012, from http://fluxicon.com/camp/2012/slides/camp2012-wim.pdf
- Municipality (unkown), process of a new social support act, process to estimate the value of houses and apartments, August 3rd, 2012, from http://is.ieis.tue.nl/research/bpmroundtable/slides/2011-11-07/02%20-%20Gemeentelijke%20en%20andere%20Processen%20in%20de%20Wolken.pdf
- Process Mining Project tutorial, August 3rd, 2012, from http://is.ieis.tue.nl/research/bpmroundtable/slides/2010-04-26/02%20Rozinat%20en%20Gunther%20-%20Fluxicon.pdf
- Rabobank, different projects, August 3rd, 2012, from http://fluxicon.com/camp/2012/slides/camp2012-frank.pdf
- Seven steps of process mining in an audit context, August 3rd, 2012, from http://fluxicon.com/camp/2012/slides/camp2012-mieke.pdf
- AE, Payroll Accounting, Public Administration, Postal Services, August 3rd, 2012, from http://fluxicon.com/camp/2012/slides/camp2012-bram.pdf

Appendix C: Structured Example of Event Information



Appendix D: Quality of a Process Model [Rozinat 07]



Appendix E: Detailed Description of PMPM

E.1 Scoping

Develop understanding of the process in order to develop a project plan containing the objectives for the project and thoughts about the tools and techniques that are required to accomplish the project.

- Identify the process
  Asses the process environment in terms of activities, resources, goals, (information) systems, constraints et cetera. This step must give an idea what the customer wants to accomplish and what the process mining value can be in this situation, which includes a global identification of the parts of the process that are logged.
- Determine Objectives
  Uncover the objectives of the organizational process (e.g. quality, efficiency, reliability, effectiveness) and formulate specific project goals in cooperation with the initiator. Derive questions and metrics from these formulated goals to clarify the goals and in order to determine the analysis techniques that must be used. Furthermore, identify the basic aspects of the events (case, resource, time, location, motivation, activity) that must be present in the available data to meet the objectives.
- Determine tools/techniques
  Make an initial selection of the tools and techniques that will be used in the project to create the event log and to perform the process mining activities. This will help to get a global idea of what is needed and to have the required knowledge and tools available.

E.2 Data Understanding

Locate where the logged data of the process can be found, explore if this contains the required elements and verify if it is reliable.

- Locate data
  Identify the location of the organizational process data. A data specialist of the organization could be very valuable to find this data. Note that this data can be scattered over different locations in the system.
- Explore data
  Develop knowledge about the identified data to uncover its meaning and to find the way the data is organized. This can usually be done by using manual techniques and bring important aspects into focus for further analysis. The main idea is to check what aspects are present in the data and how the data relates to one another.
- Verify data
  The quality of the process data should be verified on different criteria. *Trustworthy*, it should be safe to assume that the recorded events actually happened and that the attributes of events are correct. *Completeness*, no events may be missing. Any recorded event should have well-defined *semantics* and *safeness*, privacy and security concerns are addressed when recording the events. To maximize the benefit from process mining, the quality should be as high as possible.

E.3 Event Log Creation

Create an event log that is an appropriate input for the required process mining techniques to meet the objectives by selecting, extracting and preparing data from the available identified set.

- Select data

An appropriate selection of the data must be made to use as input for the analysis. This selection must be evaluated on different dimensions. *Case completion*, only historic cases or also current cases? It is not always possible to make this selection upfront. *Timeframe*, the period for which the events are extracted, which can be done on the one hand by selecting all events in a selected timeframe, or on the other hand also by e.g. a selection that is based on cases started/contained/ended/intersected in a fixed timeframe. *Aspects,* what are the aspects (case, resource, time, location, motivation, activity or a derivative of these aspects) of the recorded events that are required?

- Extract data
  Exporting the selected data from the system to a file which can be used for further analysis. The challenges at this stage do very much depend on the amount of data and the software system/tool that is used. The dataset could often be exported to different types of data files. Motivations to choose a data type can relate to its ease of conversion, adaptation or appropriateness as input for the process mining software that is used.
- Prepare data
  Usually, the extracted set of data is not appropriate as input to apply process mining. Therefore the data has to be prepared in different manners. *Cleaning* the data which may involve removing subsets or estimating missing data. *Constructing* the data such as creating derived attributes or transforming data to other values. *Merging* the data when multiple sets are concerned. *Format* the data, e.g. structuring the values in terms of a required format or inserting default variables.

E.4 Process Mining

Developing knowledge from the event log that is interesting to improve the organizational process. The main activities in this phase are familiarizing with the event log, ensuring that the event log is structured enough to apply the required process mining techniques and actual analysis of the data by applying the process mining techniques.

- Familiarize log
  Get familiar with the event log and process information that is contained in the event log. The inspection gathers several statistics about the log, e.g. amount and diversity of process instances, activities and resources, but also the average number of events per case and the different start and end activities. These statistics give a first impression in how the events in the event log are correlated and can give an indication in how a process mining technique will perform on this data.
- Ensure structuredness
  The log inspection can indicate that the data is less structured, a 'Spaghetti' process. For less structured processes only a subset of process mining techniques are applicable. Therefore it is sometimes essential to first make the 'Spaghetti-like' processes more 'Lasagna-like'. There are several ways to simplify event logs to make them more suitable for process mining techniques. Event logs can be simplified in several manners, e.g. filtering activities based on their characteristics, abstracting from infrequent activities, transforming low-level patterns into activities or clustering cases in homogeneous groups that show similar types of behaviour.
- Answer questions
  In this step, the actual mining work is done. There are ten different types of process mining related activities possible. The models considered with the process analysis can show different perspectives usually containing several aspects of an event. Moreover,

answering these questions contains more work than just applying techniques. It is also a process of filtering, adjusting and digging in the event log to retrieve the most valuable information.

E.5 Evaluation

In this stage, the models that are built using the process mining activities are evaluated. This evaluation is done by a verification, validation and accreditation of the analysed results. Furthermore, it can be decided to elaborate the process mining project.

- Verify
  Verification is a technical assessment based on the outcome of the applied techniques. Not every workflow-net represents a correct process. Several correctness criteria can be checked without domain knowledge, e.g. absence of deadlocks, activities that can never become active, livelocks.
- Validate
  The validation of the modelling results can be done by measuring how representable the modelled work is. The mined results can be checked to determine if there is some business reason why this model is inappropriate. Model could for example be not understandable by business because of its language or complexity. Another reason could be that the model represents only a part of the process instances or allows for more process flows than exist in the process.
- Accreditate
  Accreditation is the assessment of the degree to which the process mining results meet the objectives. The initiator of the process mining project should evaluate if results generated by the process mining techniques are interesting for the goals of the process. Next to answering questions that are proposed upfront, also other insights that are delivered by the modelled work could be important for the initiator of the project.
- Determine elaboration
  The process mining results can raise new questions or ideas to think about a possible elaboration of the process mining project. More specific types of analysis that add value to the objectives can be done or new questions could raise which require another event log. Furthermore, process mining can also be applied for a longer term, i.e. monitoring the process or supporting the operational process. Operational support is considered with unfinished cases and these cases can, potentially, still be influenced. However, this contains the most ambitious parts of process mining.

E.6 Deployment

At the end of a process mining project the gained knowledge must be transferred to the organization. This may include an advice for several improvement actions that can be carried out and definitely includes presenting a summary of the project results by giving a presentation and/or writing a report.

- Identify possible improvements
  A process mining project aims to add value to the process that is analysed. This can be done by checking if what is expected is true, or if results deliver new insights by presenting unknown information about the process. However, the project results can also directly influence the process by deploying improvement actions. Improvements can be done independent or dependently of the IT system, e.g. redesign, adjust, intervene, support.
- Present results

The last action remaining in the project is presenting the results. The process insights gathered during the project and the proposed actions to improve the process have to be presented to the business, so that this information can be used in improving the process.

Appendix F: Case study – Business Understanding

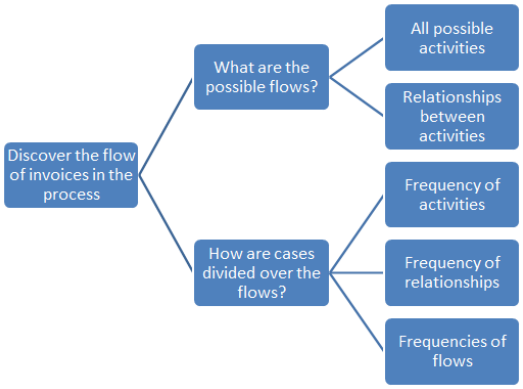


Figure F1, GQM of Insight

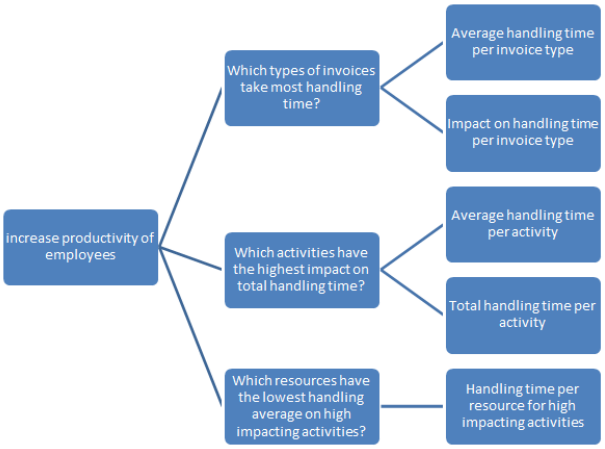


Figure F2, GQM of Process Efficiency

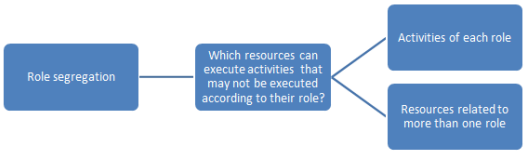


Figure F3, GQM of Risk Control

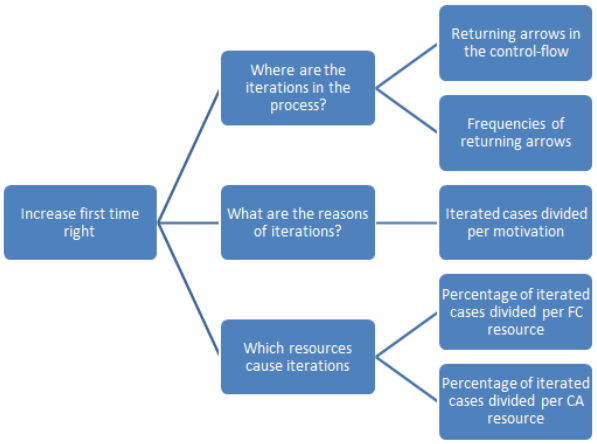


Figure F4, GQM of Process Quality

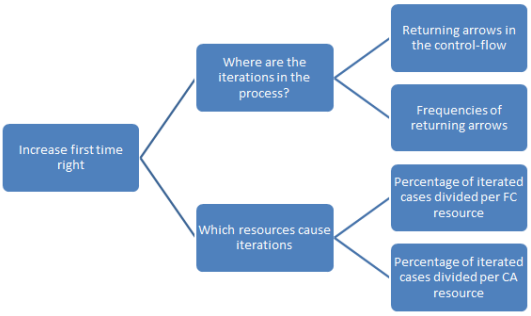


Figure F5, Adapted GQM of Process Quality

Objective Process control	Discover the flow of invoices in the process
Goal	Be better able to manage a process by knowing the possible flows and frequencies
Actors	manager, researcher, data specialists, employees
Pre-conditions	An event log containing cases, activities and time (order)
Post-conditions	A control-flow model that describes the flow of invoices in the process
Quality Requirements	1. All main activities of the process are logged 2. Cases and activities can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Develop understanding of the main process 2. Identify and gather the required data 3. Prepare the required data to apply analysis techniques 4. Apply a suitable mining technique to discover the control-flow 5. Analyse the output 6. Ensure that the control-flow is useful, i.e. structured enough 7. Present the discovered flow

Table F1, Operational scenario objective process control

Objective Process efficiency	improve productivity of employees
Goal	Save resource costs by improving the productivity of employees
Actors	manager, researcher, data specialists, employees
Pre-conditions	An event log containing cases, activities, resources and time
Post-conditions	The average handling times of employees, activities and invoice types
Quality Requirements	1. Time is logged detailed enough to be useful 2. Cases, activities and employees can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Identify and gather the required data 2. Prepare the required data to apply analysis techniques 3. Apply a suitable mining technique to discover the average handling times 4. Analyse the output 5. Present the discovered flow

Table F2, Operational scenario objective process efficiency

Objective Risk control	Role segregation
Goal	Decrease risk by adapting authorizations so that an employee cannot execute activities of more than one role
Actors	manager, researcher, data specialists, employees
Pre-conditions	An event log containing cases, activities, resources and roles (authorization profiles)
Post-conditions	All resources that have overlapping roles, too much authorization
Quality Requirements	1. For each activity the corresponding role is known 2. Cases, resources and activities can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Identify the authorized role of each activity 2. Identify and gather the required data 3. Prepare the required data to apply analysis techniques 4. Apply a suitable mining technique to identify the relationship between resources and activities 5. Analyse the output 6. Present the resources that have risky profiles

Table F3, Operational scenario objective risk control

Objective Process quality	increase 'first time right'
Goal	Improve the quality of executed activities by identifying which invoices were sent back
Actors	manager, researcher, data specialists, employees
Pre-conditions	An event log containing cases, activities, resources, time and motivation
Post-conditions	Return where, why and by who invoices iterate
Quality Requirements	1. Returned invoices can be identified 2. Cases, activities and employees can be identified by their id's in the event log 3. The motivation of returned invoices is known 4. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Identify and gather the required data 2. Prepare the required data to apply analysis techniques 3. Apply a suitable mining technique to identify the iterated invoices, their motivation and the resources that worked with those invoices 4. Analyse the output 5. Present the results of the returned invoices

Table F4, Operational scenario objective process quality

Objective Process quality	increase 'first time right'
Goal	Improve the quality of executed activities by identifying which invoices were sent back
Actors	manager, researcher, data specialists, employees
Pre-conditions	An event log containing cases, activities, resources, and time
Post-conditions	Return where and by who invoices iterate
Quality Requirements	1. Returned invoices can be identified 2. Cases, activities and employees can be identified by their id's in the event log 3. The people, products and processes that are interacting with the project are appropriate to perform the project
Description (activities)	1. Identify and gather the required data 2. Prepare the required data to apply analysis techniques 3. Apply a suitable mining technique to identify the iterated invoices and the resources that worked with those invoices 4. Analyse the output 5. Present the results of the returned invoices

Table F5, Adapted operational scenario objective process quality

Appendix G: Case study – Data Understanding

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Workflow	ID	Type	Taal	Workitemtekst	Taaktekst	Status	Aanm.dat.	Aanmtijd	Afgehand.	Bewerker	Key adresobject	Auteur	Key adresobject	Workflow
2	265130052	265345646	W	NL	Bent u echt akkoord ?	UD: Bevestig akkoord geregistreerd doc	COMPLETED	13.11.2011	11:04:55	13.11.2011	P00159769	USR35000000006676	P00159769	USR35000000006676	265130099
3	265130052	265345647	W	NL	Log actie historie: GOEDGEKEURD VOOR KOSTENPLAATS	SD: Update historie	COMPLETED	13.11.2011	11:04:58	13.11.2011	P00159769	USR35000000006676	P00159769	USR35000000006676	265130099
4	262945521	265345656	B	NL	Log actie historie: DOC VOLLEDIG GOEDGEKEURD	SB: Update historie achtergrond	COMPLETED	13.11.2011	11:05:03	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	262945521
5	262945521	265345655	B	NL	Bepaal volgend kosten element uit hoofdfow	Bepaal volgend kosten element uit hoofdf	COMPLETED	13.11.2011	11:05:02	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	262945521
6	263749862	265345744	B	NL	Archiveren van de log.	SB: Archiveren van de log	COMPLETED	13.11.2011	11:22:04	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	263749862
7	263749862	265345731	E	NL	Wachten op event 'POSTED' van objecttype 'FIPP'		COMPLETED	13.11.2011	11:21:57	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	263749862
8	263749862	265345727	B	NL	Log actie historie: DOC VOLLEDIG GOEDGEKEURD	SB: Update historie achtergrond	COMPLETED	13.11.2011	11:21:57	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	263749862
9	263749862	265345743	B	NL	Log actie historie: REGISTRATIE GEBOEKT	SB: Update historie achtergrond	COMPLETED	13.11.2011	11:22:04	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	263749862
10	263749862	265345726	B	NL	Bepaal volgend kosten element uit hoofdfow	Bepaal volgend kosten element uit hoofdf	COMPLETED	13.11.2011	11:21:57	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	263749862
11	264646119	265345691	B	NL	Log actie historie: DOC VOLLEDIG GOEDGEKEURD	SB: Update historie achtergrond	COMPLETED	13.11.2011	11:06:50	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	264646119
12	264646119	265345690	B	NL	Bepaal volgend kosten element uit hoofdfow	Bepaal volgend kosten element uit hoofdf	COMPLETED	13.11.2011	11:06:49	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	264646119
13	262945521	265345657	E	NL	Wachten op event 'APPROVED' van objecttype 'FIPP'		CANCELLED	13.11.2011	11:05:03	14.11.2011			WF-BATCH	USR28000000000233	262945521
14	262945521	265345658	E	NL	Wachten op event 'RESETAPPROVALS' van objecttype 'FIPP'		CANCELLED	13.11.2011	11:05:03	14.11.2011			WF-BATCH	USR28000000000233	262945521
15	262945521	265345659	E	NL	Wachten op event 'DELETED' van objecttype 'FIPP'		CANCELLED	13.11.2011	11:05:03	14.11.2011			WF-BATCH	USR28000000000233	262945521
16	262945521	265345660	E	NL	Wachten op event 'POSTED' van objecttype 'FIPP'		COMPLETED	13.11.2011	11:05:03	14.11.2011	P00182485	USR35000000006566	WF-BATCH	USR28000000000233	262945521
17	262945521	265345661	E	NL	Wachten op event 'CANCELLED' van objecttype 'FIPP'		CANCELLED	13.11.2011	11:05:03	14.11.2011			WF-BATCH	USR28000000000233	262945521
18	264695428	265345724	B	NL	Controleer of goedkeurder ook geautoriseerd is	SB: Controle goedkeurder geautoriseerd	COMPLETED	13.11.2011	11:21:54	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	264695428
19	264695428	265345725	B	NL	Controleer of goedkeurder ook geautoriseerd is	SB: Controle goedkeurder geautoriseerd	COMPLETED	13.11.2011	11:21:55	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	264695428
20	265345662	265345664	E	NL	Wachten op event 'STOPPROCES' van objecttype 'FIPP'		CANCELLED	13.11.2011	11:05:04	13.11.2011			WF-BATCH	USR28000000000233	265345662
21	264695428	265345719	B	NL	Update attachment van actuele flow naar hoofdflow	Update attachment van actuele flow naar	COMPLETED	13.11.2011	11:21:53	13.11.2011	WF-BATCH	USR28000000000233	P00010829	USR35000000002536	264695428
22	264695428	265345720	B	NL	Ophalen kritische data NA GOEDK	SB: Ophalen kritische data registratie	COMPLETED	13.11.2011	11:21:54	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	264695428
23	264695428	265345721	B	NL	Controleer kritische data	SB: Vergelijk kritische data	COMPLETED	13.11.2011	11:21:54	13.11.2011	WF-BATCH	USR28000000000233	WF-BATCH	USR28000000000233	264695428

Figure G1, data dump of SAP table containing process events

Workitem	Source Sy	Source Sy	Top Level	Top level	Workflow	ExecTime	Workflow R	Node ID	Parent Wo	Parent Wf	Parent Wf	Top Level	Work iten	Workflow	Process Ca	Lan
1.54E+08	FAACLNT4	CP	WS950000	40		0		237	1.53E+08	WS950000	40	1.53E+08	E		WF	N
1.54E+08	FAACLNT4	CP	WS950000	22		0		28	1.54E+08	WS950000	22	1.54E+08	E		WF	N
1.72E+08	FAACLNT4	CP	WS950000	23	TS9600003	3		3249	1.72E+08	WS960000	61	1.72E+08	W		WF	N
1.72E+08	FAACLNT4	CP	WS950000	23	TS950000E	7		60	1.72E+08	WS950000	4	1.72E+08	W		WF	N
1.72E+08	FAACLNT4	CP	WS950000	23	TS9710008	0		3784	1.72E+08	WS960000	61	1.72E+08	W		WF	N

Language	Workitem	Parent WI	Top WI Ex	Work iten	Creation c	Creation t	End date c	End time c	Work iten	Work iten	Work iten	Work iten	No. of att	Number o	Number o	Wkflw ex	/BIC/MRC	/BIC/MRC	Top Worki	UTC Tim
N		E	E	COMPLETI	03.05.201(	13:28:23	03.03.2011	13:30:20	WF-BATCH	P00174932		5	0	0	1	304	Wachten op event 'D		1.53E+08	2.01E+13
N		E	E	COMPLETI	03.05.201(	13:28:23	03.03.2011	13:30:21	WF-BATCH	PIAPPLUSER		5	0	0	1	304	Wachten op event 'S		1.53E+08	2.01E+13
N		E	E	COMPLETI	03.09.201(	12:22:30	01.01.2011	4:17:49	WF-BATCH	P00010136		5	0	0	1	120	NEW CA - LB factuur		1.72E+08	2.01E+13
N		E	E	COMPLETI	03.09.201(	15:50:14	01.01.2011	4:17:50	WF-BATCH	P00162839		5	0	0	1	120	NEW FC - I	1354	1.72E+08	2.01E+13
N		E	E	COMPLETI	06.09.201(	9:10:23	01.01.2011	4:17:50	WF-BATCH	P00010136		5	1	0	1	117	DISPUTE MC - LB fact		1.72E+08	2.01E+13

Figure G2, SAP Business Warehouse table containing event information

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Top Workitem ID	V	CHANGED	Source Sy	Company	Cost Cent	Controllin	Doc.numb	Documen	Posting da	Scan date	Clearing d	Leveranci	Valuta	Bedrag	Invoice da	Kalenderj	Closing da	Registratiedatum	
2	153480751	A		CP	3364	3364975	3364	1.9E+09	ZLB_FAC		03.05.2010		1030979		-1,566.60	08.04.2011	2010	03.03.2011	03.05.2010	
3	171717738	A		CP	3574	3574975	3574	1.9E+09	ZLB_FAC	03.03.2011	03.09.2010	04.03.2011	1043614		-1,694.08	05.11.2009	2011	03.03.2011	03.09.2010	
4	173838943	A		CP	1538	1538120	1538	1.9E+09	ZLB_FAC	02.03.2011	16.09.2010	04.03.2011	1040370		-1,098.27	10.09.2010	2011	02.03.2011	20.09.2010	
5	179169121	A		CP	5420	421560	RABO	1.9E+09	ZRNZB_FA	03.03.2011	13.10.2010	03.03.2011	1016263		-216.58	29.09.2010	2011	03.03.2011	13.10.2010	
6	179413966	A		CP	1538	1538120	1538	1.9E+09	ZLB_FAC	02.03.2011	13.10.2010	04.03.2011	1040370		-273.78	07.10.2010	2011	02.03.2011	14.10.2010	
7	181614933	A		CP	3437	3437975	3437	1.9E+09	ZLB_FAC	03.03.2011	22.10.2010	04.03.2011	1058732		-362.84	20.10.2010	2011	03.03.2011	25.10.2010	

Figure G3, SAP Business Warehouse table containing case information

Appendix H: Case study – Event Log Creation

A	B	C	D	E	F	G	H	I
Workitem I	Creation da	Creation	End date of	End time	Work item agent	/BIC/MRCHWITXT	Top Workitem ID	
153684178	03.05.2010	13:28:23	03.03.2011	13:30:21	PIAPPLUSER	Wachten op event 'STOPPROCES' van objecttype 'FIPP'	153480751	
171758754	03.09.2010	12:22:30	01.01.2011	4:17:49	P00010136	NEW CA	171717738	
171824770	03.09.2010	15:50:14	01.01.2011	4:17:50	P00162839	NEW FC	171717738	
171905623	06.09.2010	9:10:23	01.01.2011	4:17:50	P00010136	DISPUTE MC	171717738	
172026390	06.09.2010	13:11:27	01.01.2011	4:17:50	P00020049	DISPUTE MC	171717738	
173507933	15.09.2010	13:55:19	01.01.2011	4:17:51	P00120293	DISPUTE MC	171717738	

Figure H1, extracted data containing process events

	A	B
1	Top Workitem ID	/BIC/MRCHD
2	208189611	ZRNZB_FAC
3	208873335	ZRN_BL_FAC
4	202650814	ZRNZB_FAC
5	207218123	ZRNZB_FAC
6	208361756	ZRNZB_FAC
7	207689537	ZRF_EM_IV
8	205958911	ZRNZB_FAC
9	208710451	ZLB_FAC
10	206139427	ZRN_BL_FAC

Figure H2, extracted data containing invoice information

A	B	C	D	E	F
Case(ID)	**Case(type**	**Resource(ID)**	**Activity(ID)**	**Time(start)**	**Time(stop)**
153480751	ZLB_FAC	EMPsystem	Wachten op event 'STOPPROCES' van objecttype 'FIPP'	03.03.2011 13:30:21	03.03.2011 13:30:21
171717738	ZLB_FAC	EMP11315	NEW CA	01.01.2011 04:17:49	01.01.2011 04:17:49
171717738	ZLB_FAC	EMP64018	NEW FC	01.01.2011 04:17:50	01.01.2011 04:17:50
171717738	ZLB_FAC	EMP11315	DISPUTE MC	01.01.2011 04:17:50	01.01.2011 04:17:50
171717738	ZLB_FAC	EMP21228	DISPUTE MC	01.01.2011 04:17:50	01.01.2011 04:17:50
171717738	ZLB_FAC	EMP21472	DISPUTE MC	01.01.2011 04:17:51	01.01.2011 04:17:51
171717738	ZLB_FAC	EMP21228	DISPUTE MC	01.01.2011 04:17:51	01.01.2011 04:17:51

Figure H3, created event log

Appendix I: Case study – Process Mining

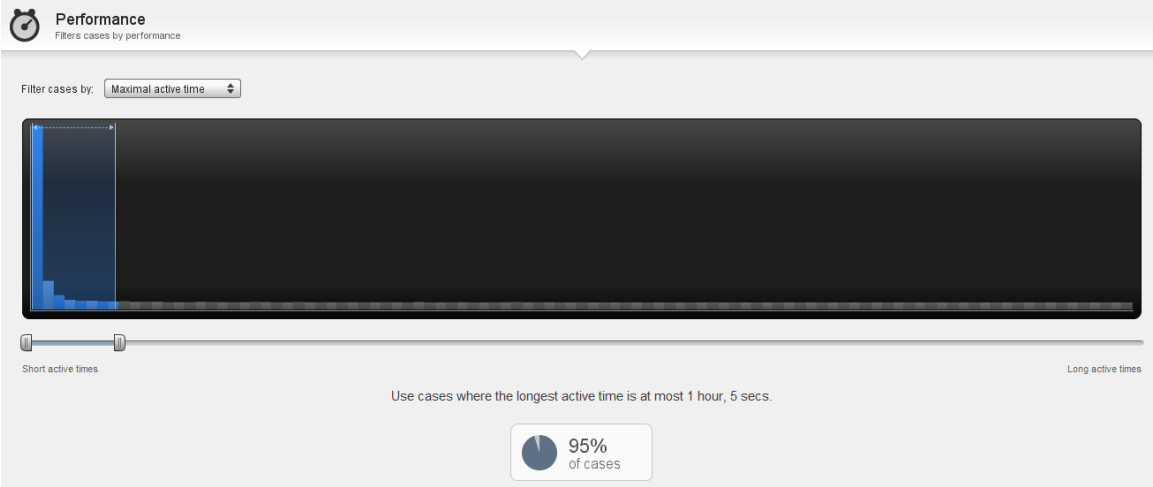


Figure I1, filter cases by the duration of activities

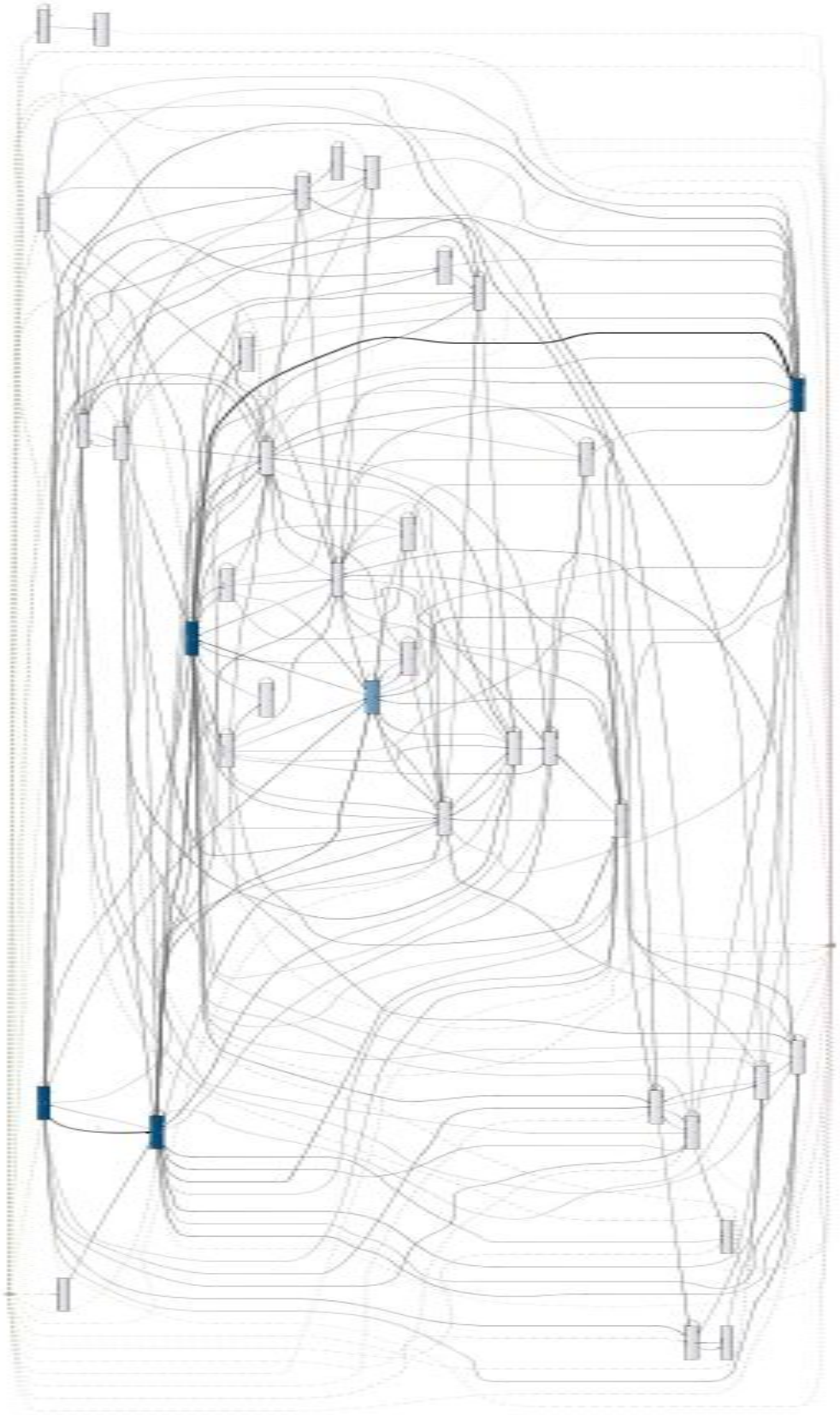


Figure I2, control-flow model, ‘spaghetti-like’

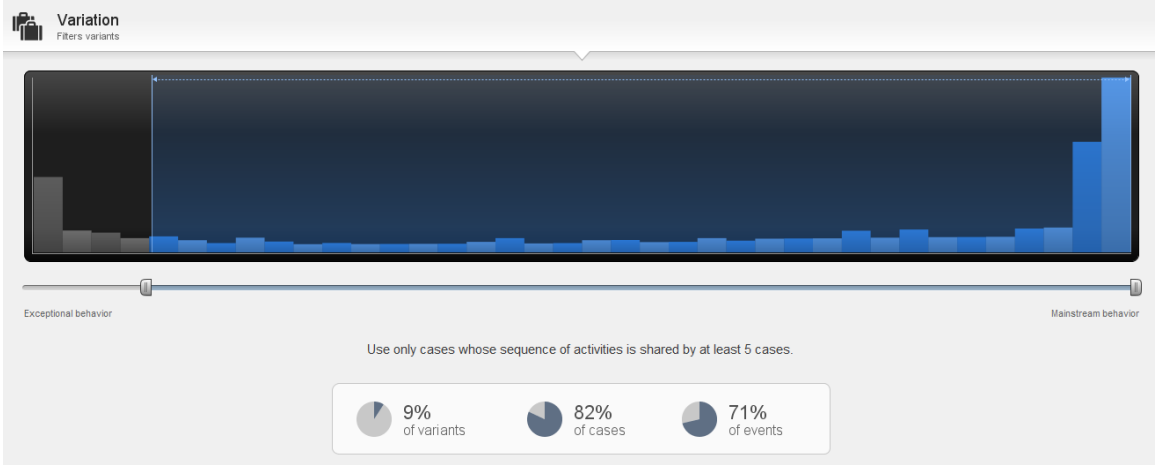


Figure I3, filter cases by occurrence of activity flows

Appendix J: Case study – Process Mining 'Process Discovery'

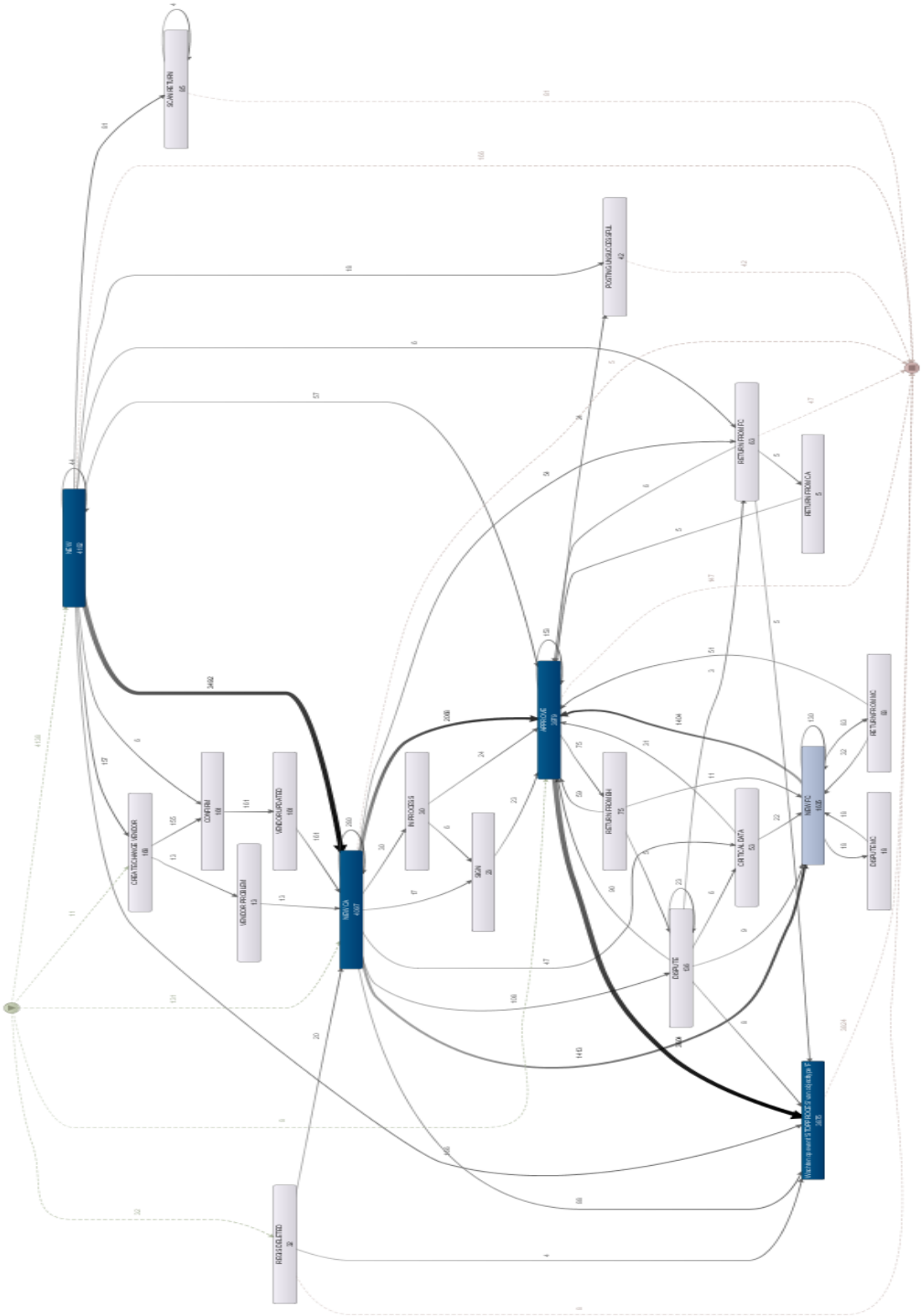


Figure J1, control-flow model, 79% of the cases.

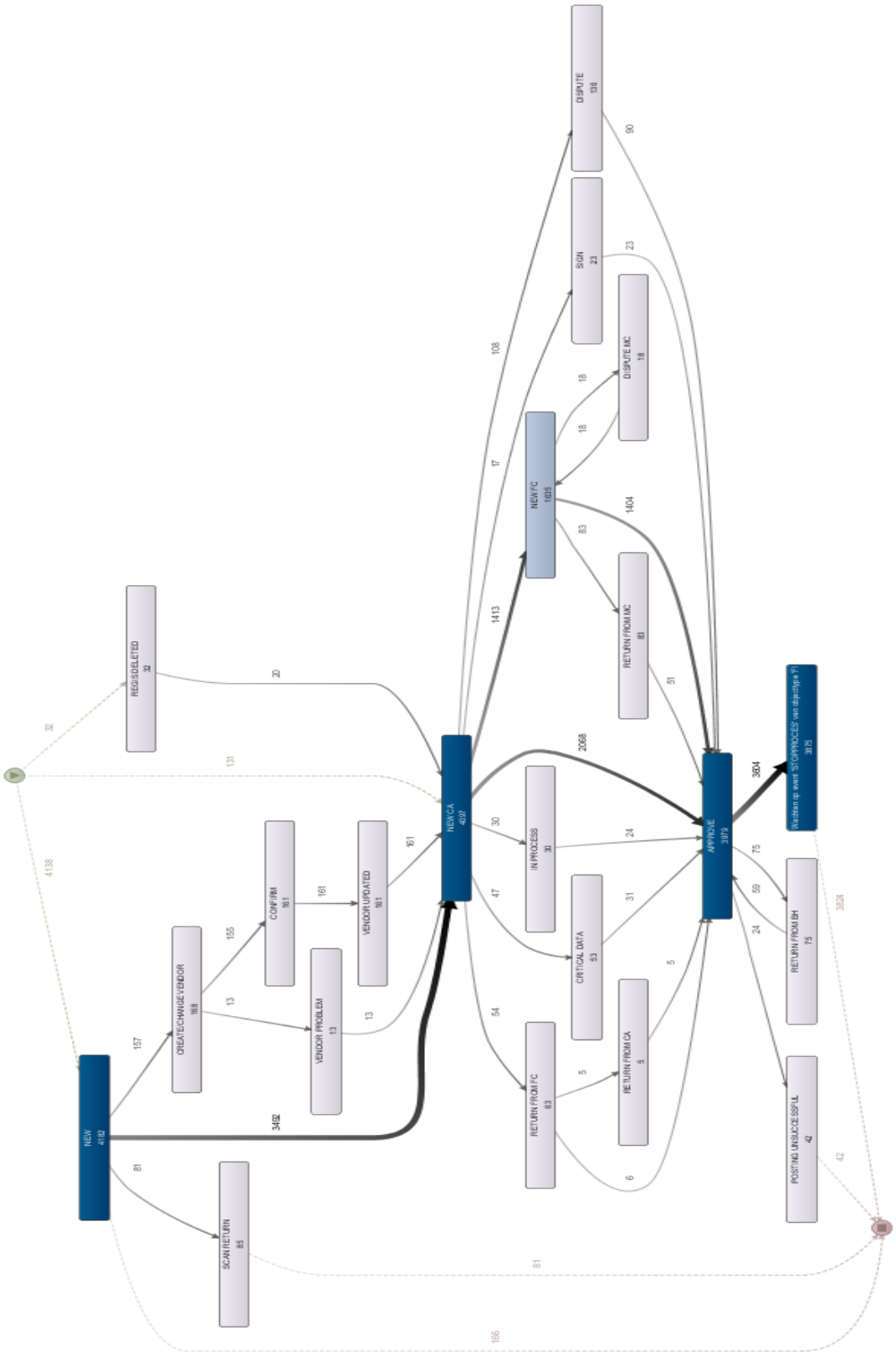


Figure J2, control-flow model, 79% of the cases and only frequent relationships

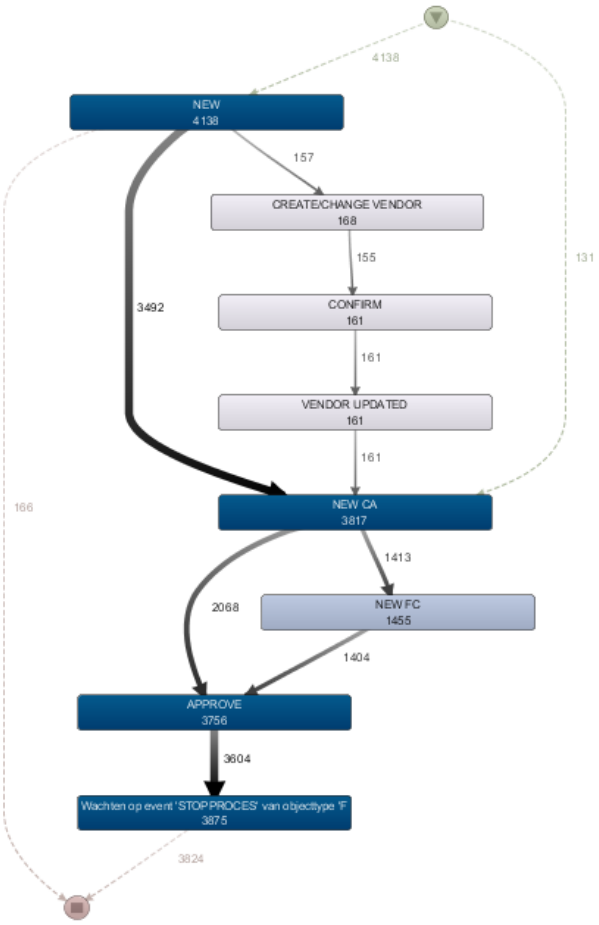


Figure J3, control-flow model, 79% of the cases and only frequent relationships and activities

Appendix K: Case study – Process Mining 'Process Efficiency'

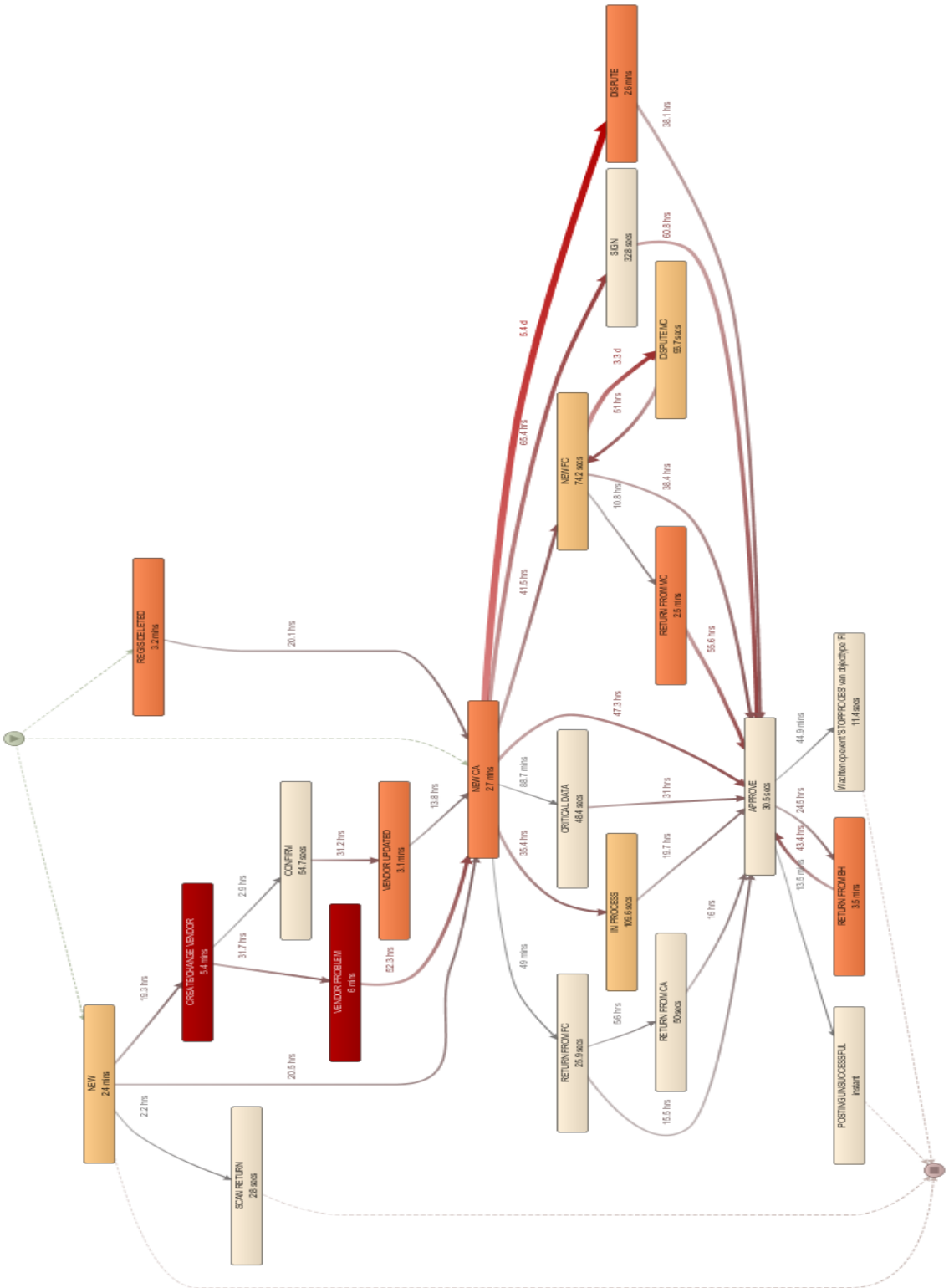


Figure K1, activities and their average handling time (darker colour means a higher average)

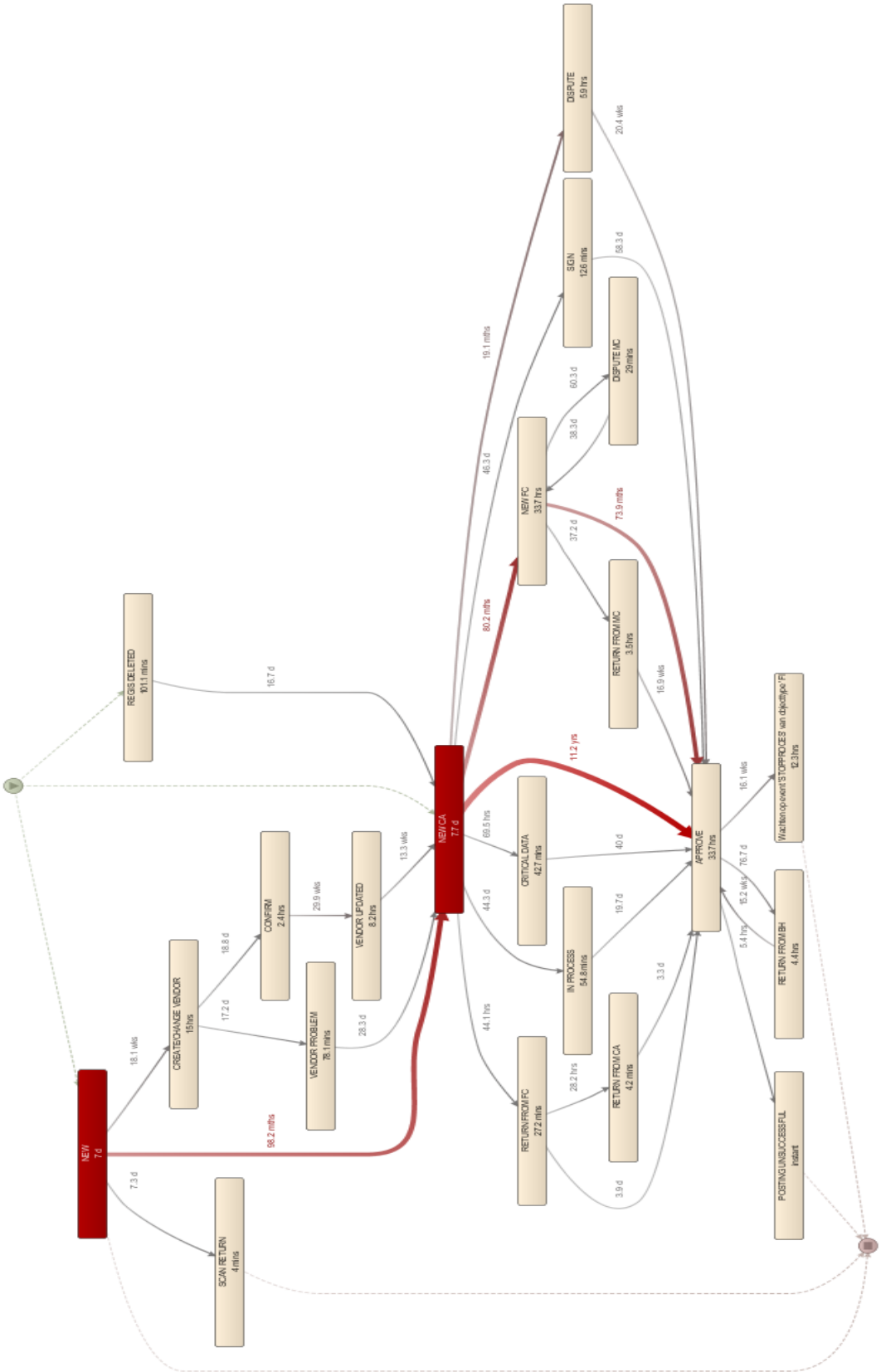


Figure K2, activities and their total handling time (darker colour means a higher impact)

Invoice type	% of cases (a)	events/case (b)	event duration (c)	Impact (a*b*c)
ZRNZB_FAC	10.8%	3.11	2.82	0.95
ZRN_EP_IV	17.8%	2.93	1.87	0.98
ZLB_FAC	61.5%	3.90	1.99	4.76
ZLB_EM_IV	6.1%	3.83	1.65	0.38
Other	3.8%	2.81	1.98	0.21

Table K1, overview of the impact of different types of invoices on total handling time

Value	Frequency	Relative frequency	Mean duration
EMP86234	351	8,39 %	0d 00:01:26
EMP66263	345	8,25 %	0d 00:01:35
EMP81916	303	7,25 %	0d 00:01:03
EMP84307	210	5,02 %	0d 00:02:23
EMP41429	206	4,93 %	0d 00:01:22
EMP81608	201	4,81 %	0d 00:02:11
EMP86776	197	4,71 %	0d 00:03:58
EMP80543	194	4,64 %	0d 00:02:05
EMP83823	194	4,64 %	0d 00:02:35
EMP75019	176	4,21 %	0d 00:02:35
EMP85058	168	4,02 %	0d 00:03:02
EMP77984	154	3,68 %	0d 00:02:48
EMP80479	148	3,54 %	0d 00:02:43
EMP76111	145	3,47 %	0d 00:01:59
EMP83531	140	3,35 %	0d 00:02:17
EMP81555	138	3,3 %	0d 00:02:42
EMP91109	136	3,25 %	0d 00:06:11
EMP74600	131	3,13 %	0d 00:02:21
EMP33513	107	2,56 %	0d 00:03:24
EMP78288	105	2,51 %	0d 00:02:26
EMP41229	92	2,2 %	0d 00:01:55

Table K2, handling time per resource for activity 'NEW'

Value	Frequency	Relative frequency	Mean duration
EMP11558	228	5,57 %	0d 00:01:48
EMP18934	216	5,27 %	0d 00:01:23
EMP14209	148	3,61 %	0d 00:01:11
EMP17273	82	2 %	0d 00:01:54
EMP40178	73	1,78 %	0d 00:02:31
EMP58217	52	1,27 %	0d 00:01:48
EMP11855	48	1,17 %	0d 00:02:32
EMP74267	47	1,15 %	0d 00:01:49
EMP93443	43	1,05 %	0d 00:02:16
EMP4539	43	1,05 %	0d 00:04:15
EMP6216	42	1,03 %	0d 00:02:19
EMP18614	40	0,98 %	0d 00:02:19
EMP10740	39	0,95 %	0d 00:01:47
EMP11574	38	0,93 %	0d 00:03:51
EMP16481	37	0,9 %	0d 00:03:46
EMP9423	36	0,88 %	0d 00:05:10
EMP20072	34	0,83 %	0d 00:01:47
EMP13685	31	0,76 %	0d 00:01:31
EMP17459	31	0,76 %	0d 00:01:10
EMP20143	30	0,73 %	0d 00:04:06
EMP30024	30	0,73 %	0d 00:02:03

Table K3, handling time per resource for activity 'NEW CA'

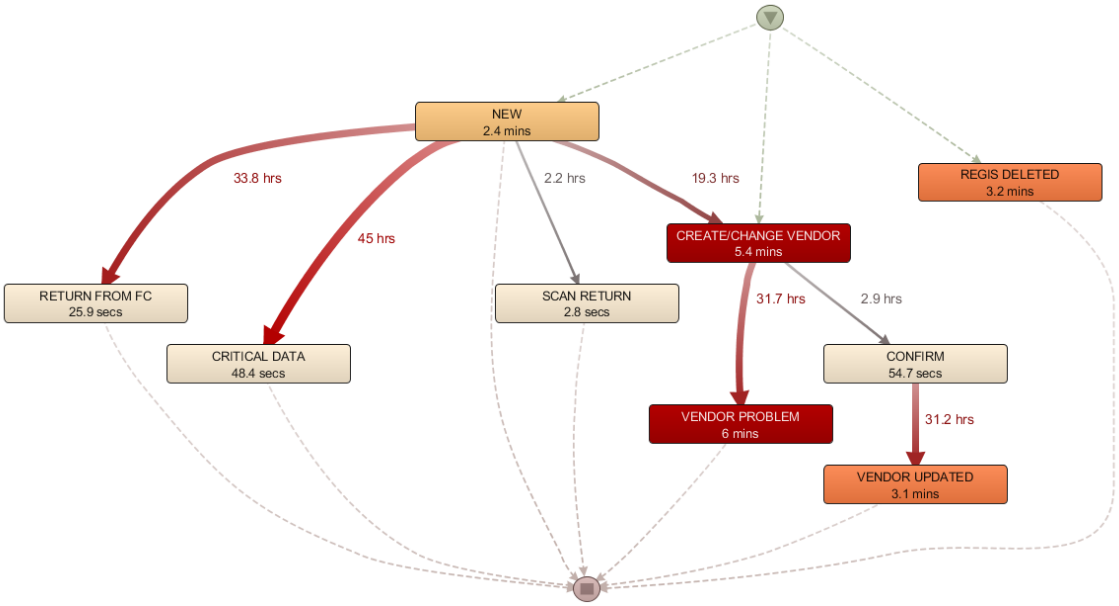


Figure K3, FS activities and their average handling time (darker colour means a higher average)

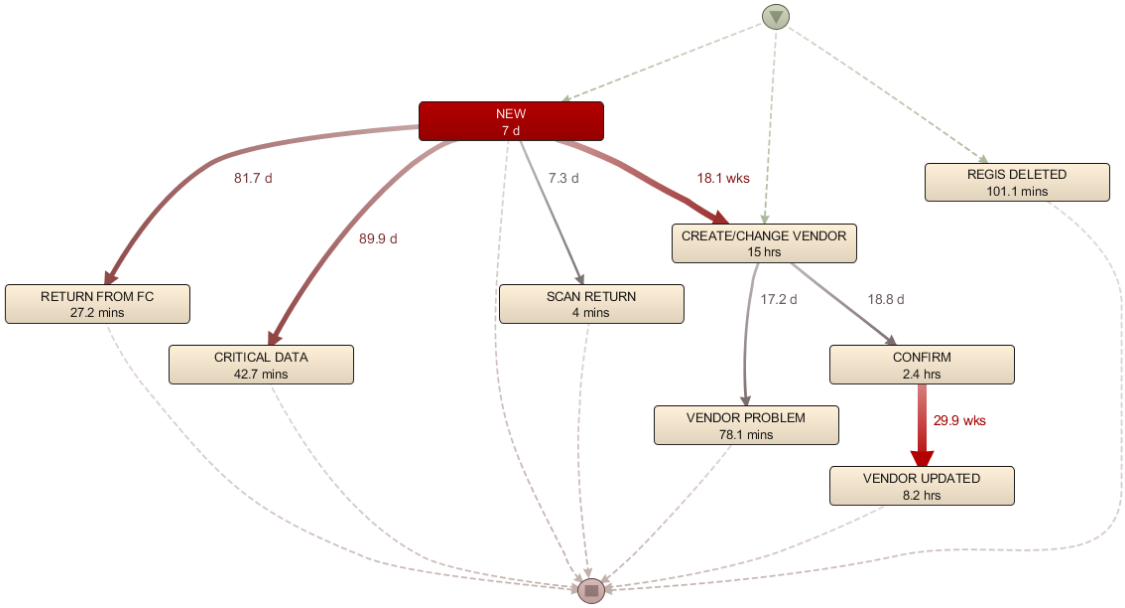


Figure K4, FS activities and their total handling time (darker colour means a higher impact)

Appendix L: Case study – Process Mining 'Risk Control'

Role	Activity	Role	Activity
BH	APPROVE	**CAFC**	RETURN FROM BH
BH	APPROVE 2	**CAFC**	RETURN FROM MC
CA	CRITICAL DATA	**FC**	DISPUTE BH
CA	GR DONE	**FC**	DISPUTE MC
CA	NEW	**FC**	GR NOT DONE
CA	Pas boekdatum aan in registratie	**FC**	IN PROCES
CA	POSTING UNSUCCESSFUL	**FC**	NEW CA
CA	RETURN FROM FC	**FC**	RETURN FROM CA
CA	SCAN OK	**FC**	RETURN FROM PL
CA	SCAN RETURN	**FC**	UD: Geen goederenontvangst
CA	UD: Verwerk gescand document	**MC**	NEW FC
CA	VENDOR PROBLEM	**PL**	SIGN
CA	VENDOR UPDATED	**SC**	CONFIRM EENM CRED
CAFC	DISPUTE	**SC**	CREATE/CHANGE VENDOR
CAFC	ERROR	**SC**	NOT CONFIRMED
CAFC	IN PROCESS	**SCCA**	CONFIRM
CAFC	REGIS DELETED		

Table L1, role for each activity (can be a combination of roles)

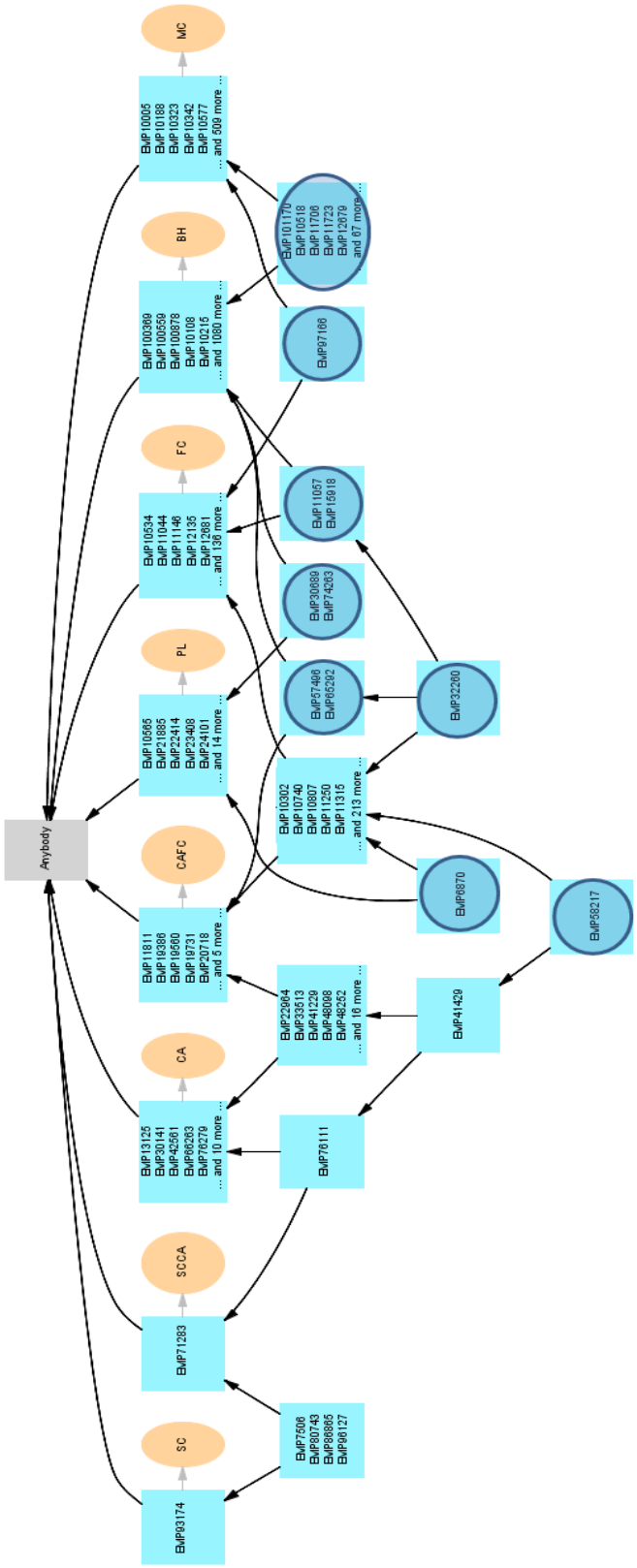


Figure L1, resources and their role according to executed activities, circles show conflicts

Appendix M Case study – Process Mining 'Process Quality'

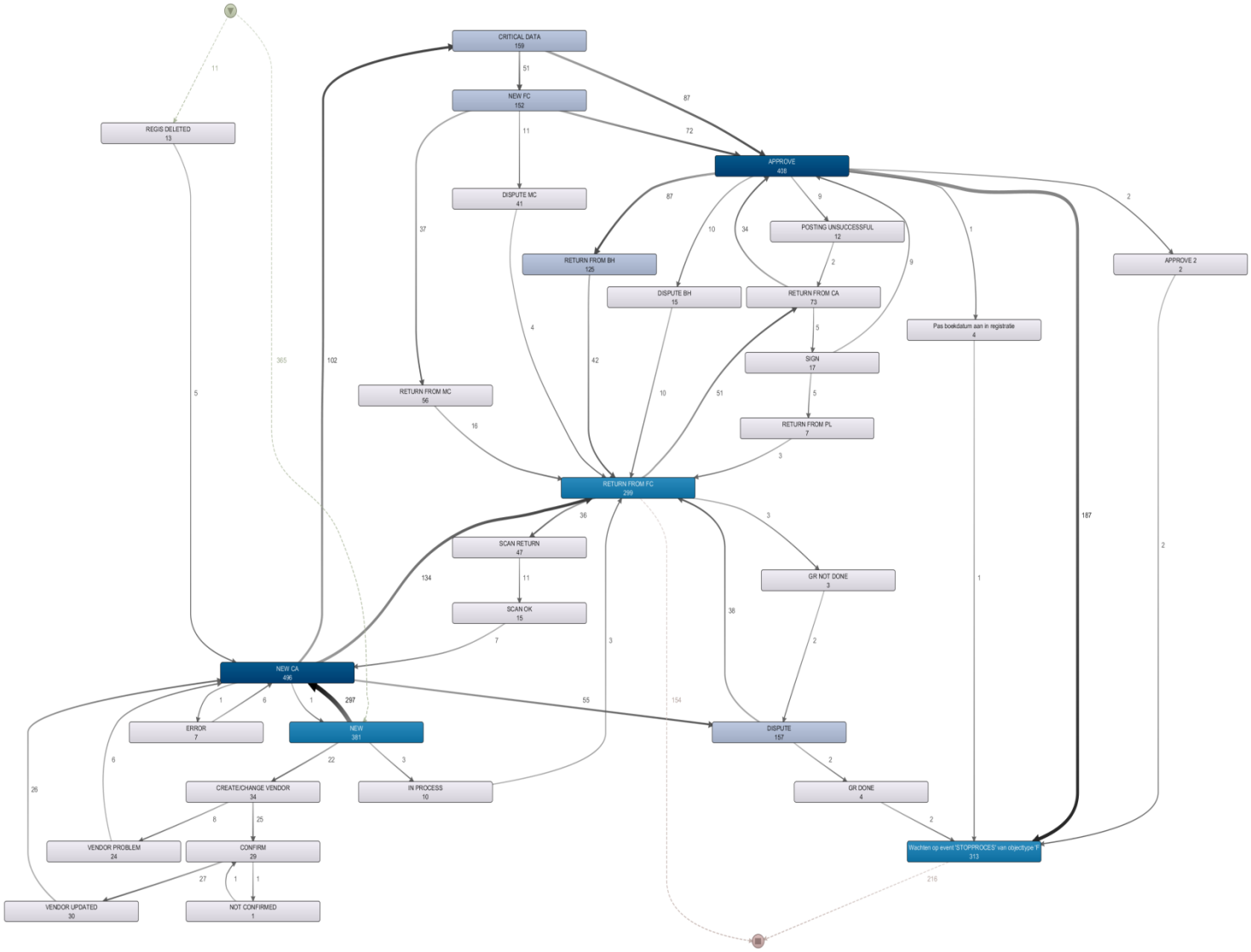


Figure M1, overview of the iterated cases in the process

Value	Frequency	Relative frequency
EMP77984	22	20,75 %
EMP81916	9	8,49 %
EMP86234	9	8,49 %
EMP41429	7	6,6 %
EMP80543	6	5,66 %
EMP76111	6	5,66 %
EMP81608	6	5,66 %
EMP86776	5	4,72 %
EMP83531	5	4,72 %
EMP41229	4	3,77 %
EMP66263	4	3,77 %
EMP91109	4	3,77 %
EMP78288	4	3,77 %
EMP74600	4	3,77 %
EMP84307	3	2,83 %
EMP85058	2	1,89 %

Table M1, occurrence of CA resources in iterations

Value	Frequency	Relative frequency
EMP86234	351	8,39 %
EMP66263	345	8,25 %
EMP81916	303	7,25 %
EMP84307	210	5,02 %
EMP41429	206	4,93 %
EMP81608	201	4,81 %
EMP86776	197	4,71 %
EMP80543	194	4,64 %
EMP83823	194	4,64 %
EMP75019	176	4,21 %
EMP85058	168	4,02 %
EMP77984	154	3,68 %
EMP80479	148	3,54 %
EMP76111	145	3,47 %
EMP83531	140	3,35 %
EMP81555	138	3,3 %
EMP91109	136	3,25 %
EMP74600	131	3,13 %
EMP33513	107	2,56 %
EMP78288	105	2,51 %
EMP41229	92	2,2 %

Table M2, occurrence of CA resources overall

Value	Frequency	Relative frequency
EMP15960	9	8,18 %
EMP64325	5	4,55 %
EMP11855	4	3,64 %
EMP6216	4	3,64 %
EMP9423	3	2,73 %
EMP93443	3	2,73 %
EMP40178	3	2,73 %
EMP11558	3	2,73 %
EMP13685	2	1,82 %
EMP4011	2	1,82 %
EMP94698	2	1,82 %
EMP7632	2	1,82 %
EMP12250	2	1,82 %
EMP17353	2	1,82 %
EMP40605	2	1,82 %
EMP8141	2	1,82 %
EMP5558	2	1,82 %
EMP17273	2	1,82 %
EMP58217	2	1,82 %
EMP28046	2	1,82 %
EMP87845	2	1,82 %

Table M3, occurrence of FC resources in iterations

Value	Frequency	Relative frequency
EMP11558	228	5,57 %
EMP18934	216	5,27 %
EMP14209	148	3,61 %
EMP17273	82	2 %
EMP40178	73	1,78 %
EMP58217	52	1,27 %
EMP11855	48	1,17 %
EMP74267	47	1,15 %
EMP93443	43	1,05 %
EMP4539	43	1,05 %
EMP6216	42	1,03 %
EMP18614	40	0,98 %
EMP10740	39	0,95 %
EMP11574	38	0,93 %
EMP16481	37	0,9 %
EMP9423	36	0,88 %
EMP20072	34	0,83 %
EMP13685	31	0,76 %
EMP17459	31	0,76 %
EMP20143	30	0,73 %
EMP30024	30	0,73 %

Table M4, occurrence of FC resources overall

Appendix N: Case study – Deployment

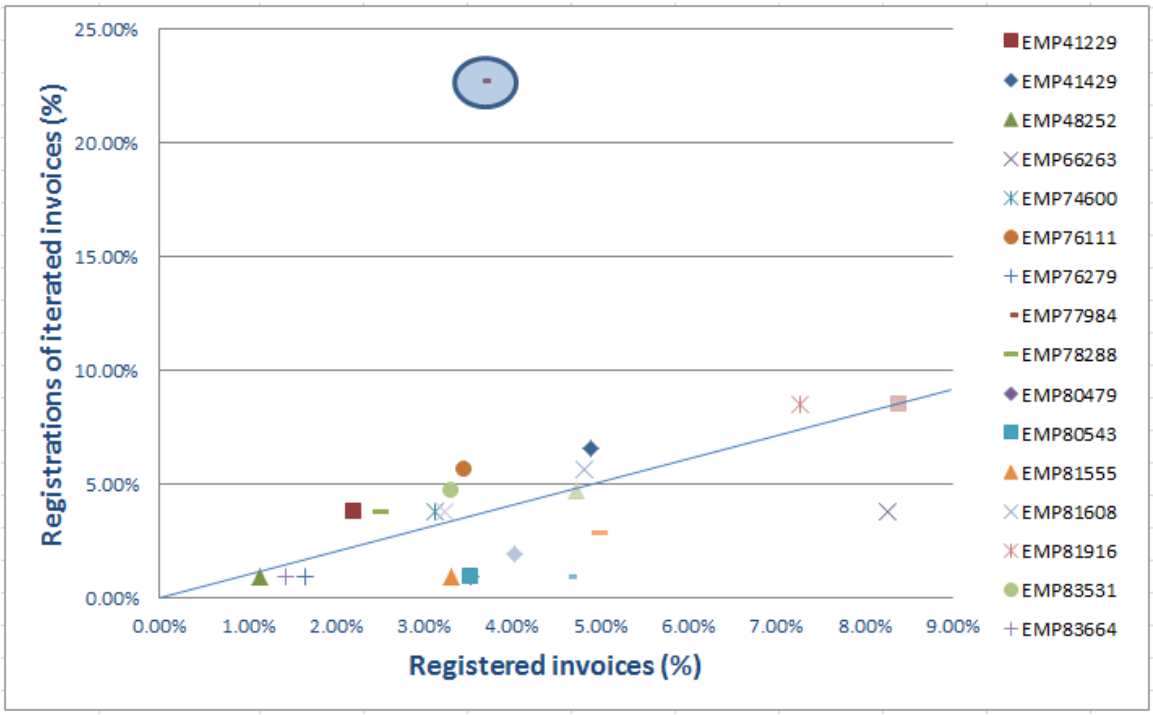


Figure N1, iterated invoices in proportion with all registered invoices, line is average