MASTER

# Business process mining success
developing knowledge on how to successfully conduct a process mining project

Prince, R.

*Award date:*
2011

Eindhoven, June 2011

# Business process mining success:

## Developing knowledge on how to successfully conduct a process mining project

BSc Industrial Engineering & Management Science — TU/e 2009

Student identity number 0554819

in partial fulfilment of the requirements for the degree of

**Master of Science**

**in Operations Management and Logistics**

Supervisors TU/e:

dr.ir. H.A. Reijers, TU/e, IS

dr.drs.ir. J.J. Berends, TU/e, ITEM

Company supervisors:

drs. H. van der Stok, T-Mobile Netherlands

ir. M. Snoep, T-Mobile Netherlands

TUE. School of Industrial Engineering.

Series Master Theses Operations Management and Logistics

Subject headings: business process mining, success factors, success measures, design principles

# Preface

At last, my master program at the Eindhoven University of Technology draws to a close with the completion of this master thesis project. It has not been an easy trajectory; at times it was hard to balance the practical part of the graduation project at T-Mobile Netherlands with the additional, more theoretical oriented research that is described in this thesis. An important lesson that I have taken from such a project is that is important to define a realistic, small scope at the outset of the project, go through the whole process, and then possibly adjust/refine the scope in an iterative manner. This is to avoid the pitfall of trying to do everything first time right, because changes are likely to be necessary as the project progresses.
Still, ultimately it is rewarding to have completed a research project that was both highly relevant for practice by conducting a specific process mining analysis for T-Mobile Netherlands as well as relevant to academics by making a contribution to process mining success research.

This achievement would not have been possible without the help of a number of people. First, I would like to thank my supervisors: Hajo, many thanks for your patience, and the (numerous) interesting meetings we had during which you provided me with useful and critical remarks. Thanks go to Hans Berends for discussing with me the methodological choices that had to be made and providing me with access to Nvivo. Also, I want to thank Ronny for his comments on the draft version of this report. Additionally, my appreciation goes to Marieke Snoep and Hans van der Stok for their time, commitment and opportunity for conducting the practical part of my master thesis project at T-Mobile. This provided me with an invaluable practical experience.

In close, I want to express my appreciation for the opportunities that both my mom and dad have given me in life. Without their support this would not have been possible. Of course, I would also like to thank my sister for her unconditional support and offering me a place to stay when I was in the Hague at T-Mobile. A special word of thanks goes to my grandma who can finally witness my graduation. Additionally, I would like to thanks my friends in Eindhoven and Middelburg for all the good times and my roommates for all the nice dinners we shared. Last but not least, I would like to thank a special someone; my girlfriend. Her support, patience and love helped me to bring this project to a good end.

Rogier Prince

Eindhoven, June 2011

## Abstract

Process mining is a method to extract process information from the data that is stored in information systems. It is a relatively young research area. Although quite some research has been conducted on the technical aspects and practical application of process mining, there has been little research that addressed the problem of how organizations can successfully conduct process mining projects.

To develop knowledge on process mining success relevant for practitioners, a theory-based analysis and practice based analysis were conducted. In the theory-based analysis, an a-priori model of process mining success was constructed, showing the relationships between the success measures, moderating factors and success factors of process mining projects. Theoretical design principles were distilled to aid practitioners in the realization of the success factors.

Subsequently, the a-priori model was evaluated and respecified through a multiple case study, which lead to the construction of a respecified model of process mining success. Additionally, design principles were distilled from these case studies to give prescriptive knowledge to practitioners how to conduct a successful process mining project.

# Summary

Process mining is a method to extract process information from the data that is stored in information systems. It is a relatively young research area. Although quite some research has been conducted on the technical aspects and practical application of process mining, there has been little research that addressed the problem of how organizations can successfully conduct process mining projects.

*Object of study & research questions*

In this thesis the main object of study was the process mining project. This includes the derivation of process models and process information through process mining, and the subsequent use of these derived models and information (adapted from Bandara, 2007, p. 30 & 139). Furthermore, the activities are conducted within an organization and organized in a project that has a certain business objective: in this case preferably business process improvement. Several research questions were formulated to guide the research efforts:

**How can organizations successfully conduct process mining projects?**

- What are the antecedent factors of process mining success?
    - Are there any contextual variables that influence the effects of these antecedent factors?
    - What are sub-constructs of the antecedent factors?
    - How can the antecedent factors be realized?
- What success dimensions are appropriate to measure process mining success?
    - What are sub-constructs of the success dimensions?

*Research design*

The research design was inspired by the research-design-development cycle of Van Burg et al. (2008), see figure 1 below.

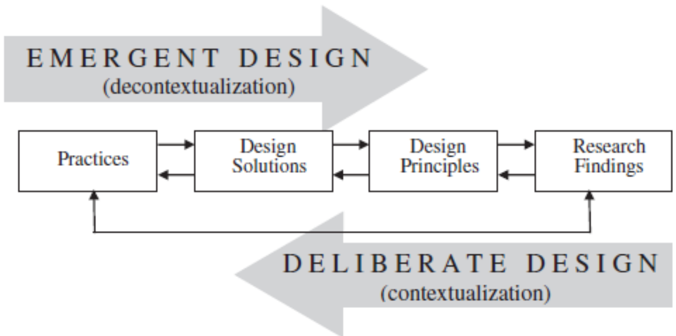


Figure 1 - The research-design development cycle (van Burg et al., 2008)

In the research design, practices (practice) and research findings (theory) were contrasted and linked through the creation of design propositions. To this end, first a **theory-based analysis** was conducted which consisted of:

- An analysis of important success factors and success measures of previously conducted process mining projects through open coding
- Construction of an a-priori of process mining success, borrowing constructs from the disciplines of business process modelling and data mining

- Distilling theoretical design principles from previously conducted process mining case studies

Next, a **practice-based analysis** was conducted which consisted of a multiple case study of four cases studies. This case study aimed to evaluate the completeness and validity of the a-priori process mining success model. Interview transcripts and documents were coded and analyze in the qualitative software tool Nvivo. After analyzing the case study the a-priori model was respecified. Additionally, specific design principles were formulated by investigating case study data on mentioned specific interventions to realize success factors.

*Results*

Through the investigation of theory and practice, an a-priori model of process mining project success was developed, tested and respecified. This model provides more insight on how to measure process mining project success and which factors contribute to this success. It is displayed in figure 2 below.

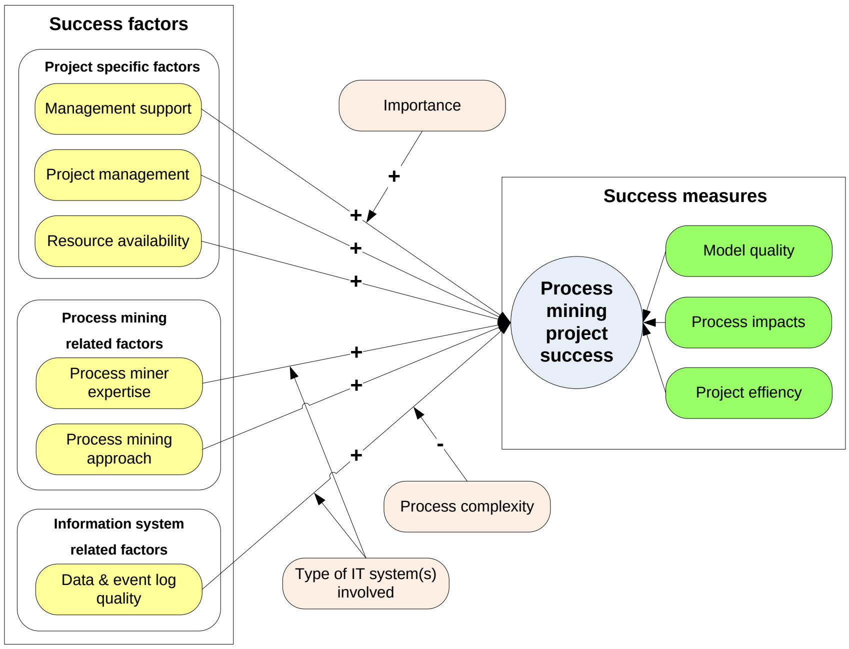


**Figure 2 - Process mining success model**

It was found that the antecedent factors of management support, project management, resource availability, process miner expertise, process mining approach and data & event log quality influence process mining project success. Important dimensions of process mining success are: model quality, process impacts and project efficiency. Several moderating variables were found, including: importance, type of IT system and process complexity. Some support was found for: importance influencing the relationship between management support and process mining success, type of IT system influencing the relationship between data & event log quality and process mining project success, type of IT system influencing the relationship between process miner expertise and process mining project success, and process complexity influencing the relationship between data & event log quality and process mining project success.

Additionally, to provide practitioners with prescriptive knowledge on how to conduct process mining projects, theory-based and research-based principles were formulated. These principles are linked to the success factors in the a-priori and respecified models of process mining success. Both theoretical and research-based design principles were formulated.

The specific contributions of this research project are as follows:

1. This thesis points to the need of proper success evaluation of process mining projects in process mining research
2. This thesis has generated new information on how process mining projects can be evaluated on what factors affect the success of a process mining project
3. This thesis can be seen as an initial step for developing design propositions that provide actionable, specific interventions for practitioners to increase the success of their process mining projects

# Table of Contents

# 1. Introduction to the research

## 1.1. General introduction

Process mining is a method to extract process information from the data that is stored in information systems. It is a relatively young research area. Although quite some research has been conducted on the technical aspects and practical application of process mining, there has been little research that addressed the problem of how organizations can successfully conduct process mining projects.

The objective of this study is to provide practitioners, and to lesser extent academics, with knowledge on process mining project success. It is less typical of process mining research which is often quite 'technical' in nature, focusing on (the application of) specific algorithms, techniques and tools that are used in process mining. In contrast, this thesis also focuses on other factors that affect process mining success and how this success is measures. This is a necessary step to developing prescriptive knowledge for practitioners on how to effectively conduct a process mining project. Developing prescriptive knowledge is the aim of design science research. As such, this research can be positioned as a mixture between process mining and design science research.

## 1.2. Research questions

In this thesis the main object of study was the process mining project. This includes the derivation of process models and process information through process mining, and the subsequent use of these derived models and information (adapted from Bandara, 2007, p. 30 & 139). Furthermore, the activities are conducted within an organization and organized in a project that has a certain business objective: in this case preferably business process improvement. Several research questions were formulated to guide the research efforts (see Table 1.1 below).

| Level 1 – Managerial question | |
|---|---|
| **How can organizations successfully conduct process mining projects?** | |
| Level 2 – Research questions | |
| ○ What are the antecedent factors of process mining success? | ○ What success dimensions are appropriate to measure process mining success? |
| Level 3 – Investigative questions | |
| ➢ Are there any contextual variables that influence the effects of these antecedent factors?<br>➢ What are sub-constructs of the antecedent factors?<br>➢ How can the antecedent factors be realized? | ➢ What are sub-constructs of the success dimensions? |

Table 1.1 – Overview of research questions

Following Bandara (2007, p. 31), questions were defined on the managerial, research and investigative levels (Table 1.1). On the managerial level, the research question addressed the practical problem of process mining success.
The managerial question was then translated to actual research questions on the second abstraction level. To know how to successfully conduct a process mining project, organizations need to

understand the mechanisms that produce success and how process mining success can be evaluated. Therefore, an important distinction was made between *success factors,* which are those inputs to the management system that lead directly or indirectly to the success of the project, and *success measures,* which are measures against which success or failure of a project will be judged (Cooke-Davies, 2002).
Additionally, questions on the investigation level were specified to answer the research question more specifically (Bandara, 2007). The first question was to look for contextual factors that would influence the relationship between the success factors and success measures. Secondly, another question was formulated to identify sub-constructs of the factors and measures, which could help to further specify these constructs and aid in the possible subsequent operationalization of the constructs (e.g. for the design of a survey). Lastly, the question was posed which interventions are available to project stakeholders to influence the antecedent factors of process mining success.

## 1.3. Research strategy

This research has adopted a design science approach where scientific knowledge was linked to the pragmatic knowledge of practitioners. As such, this research attempts to bridge the gap between managerial practice and academic research (after van Burg et al., 2008). Figure 1.1 below illustrates the activities that comprise the science-based design process.

As figure 1.1 illustrates, design science involves linking practices and research findings through design principles and design solutions. Design principles involve a coherent set of normative ideas and propositions, grounded in research, that serve to design and construct detailed solutions. Design solutions are representations of the practices being redesigned with help of the design principles, involving actions in the virtual world of drawings, models, narratives, simulations, and so forth (van Burg et al., 2008).

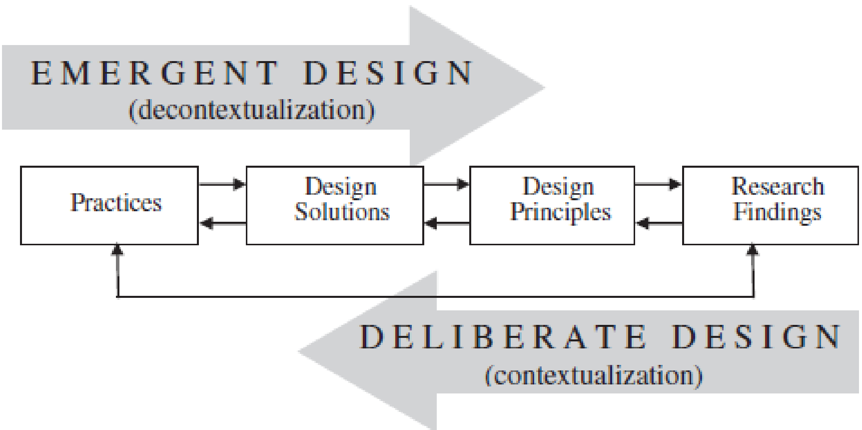


Figure 3.1 - The research-design development cycle (van Burg et al., 2008)

It is important to note that this thesis does not comprehensively cover the whole cycle. For the most part, it covers the interaction of research findings (theory) and practices (practice). This is because the field of process mining success lacks a widely accepted theoretical framework. Therefore, the deliberate and emergent dimensions of academic entrepreneurship need to interact and converge in building a cumulative body of knowledge and practice (adapted after van Burg et al., 2008). To a lesser extent this research aims to develop design principles (or propositions). However, this thesis does not cover design solutions due to scope limitations.

In order to construct a cumulative body of knowledge and practice, the following activities took place:

- A synthesized literature review (Chapter 4) was conducted, capturing potential process mining critical success factors and success measures, and an a-priori model of process mining success was derived
- An in-depth exploratory multiple case study (Chapter 5) was conducted to justify and possibly extend success factors and measurement dimensions of the a-priori model. The primary outcome of this phase was a conceptual process mining success model

Additionally, a limited set of design principles was distilled from theory and practice from the viewpoint of the project manager or project leader (which was often the process miner). Design principles or design propositions are often formulated in the so-called ‘CIMO’ logic which specifies the context (C), intervention type (I), mechanism (M) and outcome (O) of the proposition (Denyer et al. (2008), see section A.4.2. Through the construction of the a-priori success model an understanding was developed of context, mechanisms, intervention and outcomes. Specific design propositions were then formulated from theory and practice.

## 1.4. Thesis structure

The remainder of this thesis is structured as follows: chapter 2 describes the theoretical background of process mining and success measurement in related disciplines. Chapter 3 elaborates on the research design in more detail. Chapter 4 holds the results of theory-based analysis. Chapter 5 presents the practice-base case study results and the respecified model of process mining success, and distils design propositions from practice. Chapter 6 presents the conclusions and discusses the implications of the obtained results.

# 2. Theoretical Background

The following chapter serves to provide the reader with important background information on current research within the topic of process mining (success) and related fields. For the construction of this chapter, a literature review was undertaken which served to describe the current knowledge and possible gaps and important issues on the topic of process mining success. A systematic review procedure was used in this effort (Tranfield et al., 2003), which is described in Appendix A.

This chapter further consists of the following sections: section 2.1 describes some more general concepts that are important to business process mining. Section 2.2 provides a treatise of the different information systems that can be used to "produce" the data used in process mining. Section 2.3 is concerned with the basics process mining. Section 2.4 and 2.5 describe the related fields of business process modelling and data mining. Section 2.6 provides a treatise of success evaluation in general and related disciplines. Section 2.7 describes the success factors that are mentioned in related disciplines.

## 2.1 Business process (management)

Central to the topic of business process mining is the business process. The focus on business processes as a fundamental unit of analysis gained widespread attention with the advent of business process reengineering (BPR), at the beginning of the 1990s. However, its origins can be traced back to scientific management at the beginning of the previous century (Melao & Pidd, 2000).
In the BPR approach, the radical redesign of business processes is advocated through the power of information technology (Davenport & Short, 1990; Hammer & Champy, 1993). Nonetheless, the radical, IT-driven and mechanistic approach of BPR has proved to be less successful. That is why BPR has currently evolved to a more IT-enabled, holistic and systematic process management approach (Melao & Pidd, 2000).

The question is what exactly constitutes a business process. Many different definitions exist in literature, which makes it hard to distinguish a clear concept (Vergidis, 2008). One definition is given by Weske (2007), who defines a business process, as: "a set of activities that are performed in coordination in an organizational and technical environment. These activities jointly realize a business goal. Each business process is enacted by a single organization, but it may interact with business processes performed by other organizations." In essence, business process thinking can be seen as a way to deconstruct organizational complexity (Bandara, 2005). Several important dimensions of business processes can be distinguished (Weske, 2007):

- *Organizational vs. operational* - Organizational business processes are high-level business processes, they characterize coarse-grained business functionality. In contrast, operational business processes specify all the activities and their relationships.
- *Degree of automation* - On the one end of the spectrum, there are automated processes which require little to no human involvement in enactment, on the other end, one can think of processes which are fully enacted by people instead of information technology
- *Degree of structuring* - This concerns the way in which the process is executed. A process is characterized as structured, if activities are always executed in a certain order, adhering to specific execution constraints.

Nowadays, in order to stay competitive, more and more organizations are looking to improve business process performance in terms of quality, costs, flexibility and time. Business process management (BPM) is the field that deals with the realization of this goal through the (re)design and control of business processes. BPM is defined by Weske (2007) as: "concepts, methods, and techniques to support the design, administration, configuration, enactment and analysis of business processes." Typically, a BPM initiative would consist of a number of phases. These are logically related in the BPM Lifecycle model of Weske (2007), depicted in figure 2.1 below.

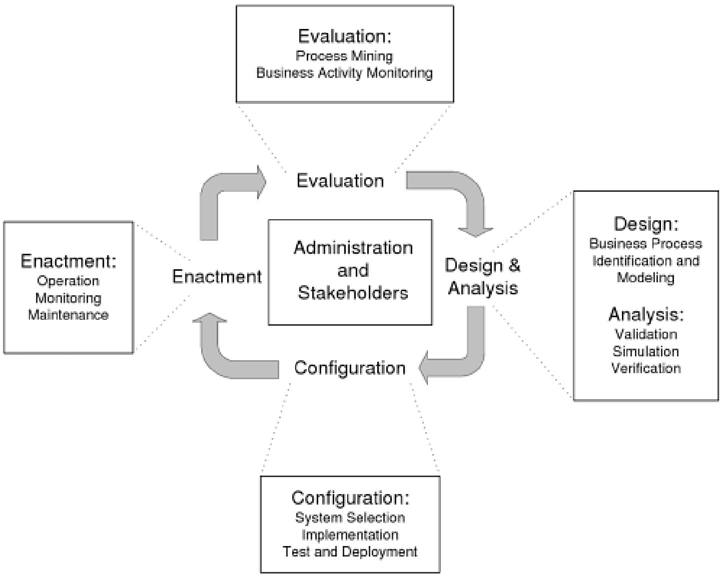


**Figure 2.4 - Business Process Management (BPM) Lifecycle (Weske, 2007)**

Figure 2.1 distinguishes the design & analysis, configuration, enactment and evaluation phases in the BPM Lifecycle. During the design and analysis phase, surveys of the organizational and technical environment are conducted, leading to the identification of certain business processes. Process modelling can be used in these phases. In the configuration phase, the implementation of the business process design takes place. Enactment concerns the actual run time of the business process. The evaluation phase uses information available to evaluate and improve business process models and their implementations. Process mining can be used for this purpose.

Furthermore, it is important to note that a business process requires proper administration and that the business process domain is characterized by several types of stakeholders, which include (Weske, 2007):

- *Chief process officer* – the person responsible for process management in the organization
- *Business engineer* – the business domain expert
- *Process designer* – the person responsible for modelling the business process
- *Knowledge worker* – the person with intricate knowledge about a certain IT application used by the process
- *System architect & developer* – the person responsible for developing and configuring the IT systems that support business processes

## 2.2 Information Systems

Process mining needs data which can originate from information systems that support business processes. It is important for the reader to understand that different types of information systems exist, which can affect process mining results. According to Alter (2002), an information system is: "a particular type of work system that uses information technology to capture, transmit, store, retrieve, manipulate, or display information, thereby supporting one or more other work systems." Whereas information technology is defined as (Alter, 2002): "the hardware and software used to store, retrieve and transfer information," and a work system as: "a system in which human participants perform a business process using information, technology and other resources to produce products for internal customers."

Van der Aalst (2007a) describes the spectrum of computer supported cooperative work (CSCW) information systems in the context of process mining. CSCW systems support work in all its forms. Two important dimensions are distinguished in the classification of these systems: on the one hand data centric systems focus on the sharing and exchange of data, whereas process centric systems focus on the ordering of activities. On the other hand, structured systems (a predefined way of dealing with things exists) and unstructured systems (no predefined way of dealing with things exists) are identifiable. Figure 2.2 presents a classification of some example CSCW systems:

- Staffware, a production workflow system
- InConcert, an ad-hoc workflow system
- Outlook, a groupware system
- SAP R/3, an enterprise resource planning system
- FlOWer, a case handling system

A slightly different classification of information systems is based on the concept of process-awareness (van der Aalst , 2009): "A *Process-Aware Information System* (PAIS) is a software system that manages and executes operational processes involving people, applications, and/or information sources on the basis of process models."
PAISs can be classified according to two dimensions: human-system orientation and the degree of structure of the processes to be automated. The first dimension can be classified into several categories: Person-to-Person (P2P), Person-to-Application (P2A) and Application-to-Application (A2A) processes. The second dimension is subdivided in: unframed (unstructured), ad hoc framed (a-priori model only executed a few times), loosely framed (there is an a-priori model, but deviations are allowed), and tightly framed processes (an explicit process model has to be followed).

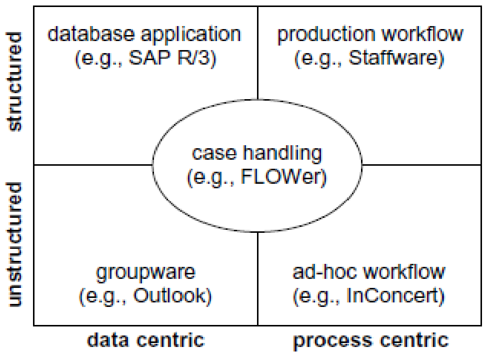


Figure 5.2 - Classification of the CSCW spectrum
(van der Aalst, 2007a)

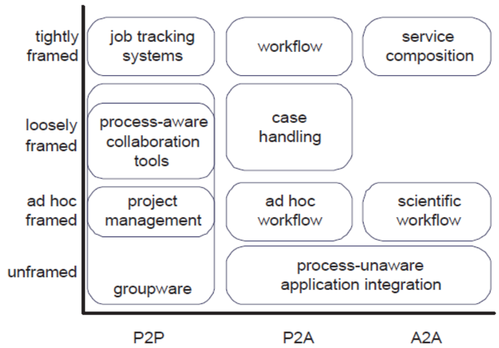


Figure 2.6 - Classification of PAIS
(van der Aalst, 2009)

## 2.3 Business Process Mining

For readers that are less familiar with process mining, the process mining basics are described in section 2.3.1. Next, section 2.3.2 describes the crucial step of creating event logs, something that is vital to conducting a process mining analysis. Section 2.3.3 describes the process mining tools that were considered in this thesis and mentions some important algorithms as an example. Lastly, known issues that affect process mining results are discussed in section 2.3.4.

### 2.3.1 Process mining basics

Process mining is a relatively young research area: the first publications on this subject appeared at the end of the 1990s (Agrawal et al., 1995; Cook and Wolf, 1998). Basically, process mining is a method to extract process knowledge from the data is recorded on transactions and events in the information systems that support business processes (van der Aalst & Weijters, 2004). Van der Aalst (2010b) distinguishes three different types of process mining (see figure 2.4 below):

- *Discovery* – There is no a-priori model. A model is constructed solely based on an event log
- *Conformance* – There is an a-priori model. This model is used to check if reality, as recorded in the data of the information system, conforms to the model and vice versa (van der Aalst, 2005a)
- *Extension* – There is an a-priori model. This model is extended with a new aspect or perspective, i.e., the goal is not to check conformance but to enrich the model. An example is the extension of a process model with performance data

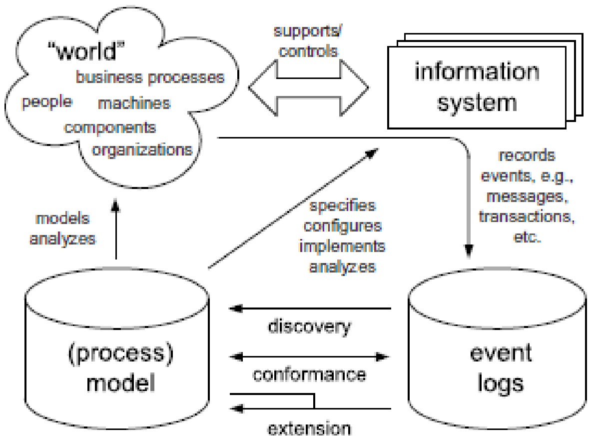


**Figure 2.4 - Schematic overview of process mining (van der Aalst, 2010b)**

It should be noted that process mining can also be used to 'predict the future' of running cases (van der Aalst, 2010a) and to build other types of simulation models (Rozinat et al., 2009). However, this thesis will only focus on the evaluation/diagnosis of enacted business processes. Process mining can be used to analyze three different performance perspectives (van Dongen et al., 2005):

- The *process perspective* focuses on the control flow. The goal of mining this perspective is to find a good characterization of all possible paths.
- The *organizational perspective* focuses on the originator field, i.e., which performers are involved and how are they related. The goal is to either to structure the organization by classifying people in terms of roles and organizational units or to show relations between individual performers.

- The *case perspective* focuses on properties of cases. Cases can be characterized by their path in the process or by the originators working on a case. However, cases can also be characterized by the values of the corresponding data elements.

### 2.3.2 Event log creation

Before process mining can be applied, the abundance of data that is stored in information systems needs to be converted to an event log. Basically, an event log is a way to record the events that take place in a specific process for a specific case (Buijs, 2010). Figure 2.5 below provides an overview of the general structure of an event log.

Different types of information systems store event log information in different ways. In order to standardize the storage of event log information, and to create a common input format for process mining, van Dongen & van der Aalst (2005a) have proposed the XML mining (MXML) format. In the MXML format, logged events are referred to as audit trail entries. These audit trail entries have to meet certain requirements (van Dongen & van der Aalst, 2005a) in order to be able to construct an event log in MXML format:

- Each audit trail entry should be an event that happened at a given point in time. It should not refer to a period of time. For example, starting to work on some work-item in a workflow system would be an event, as well as finishing the work-item. The process of working on the work-item itself is not.
- Each audit trail entry should refer to one activity only, and activities should be uniquely identifiable.
- Each audit trail entry should contain a description of the event that happened with respect to the activity. For example, the activity was started or completed.
- Each audit trail entry should refer to a specific process instance (case). We need to know, for example, for which invoice the payment activity was started.
- Each process instance should belong to a specific process.

The MXML format was mostly developed for structured processes. To be able to log information from unstructured processes better, the newer eXtensible Event Stream (XES) event log format has been developed. The main difference between MXML and XES is that attributes of logs, traces and events can be defined by the user through so-called extensions, offering a more flexible standard to construct event logs (Buijs, 2010).

Regarding the process of converting raw data to an event log, several important considerations should be made. Buijs (2010): “In general the conversion consists of two steps. The first step is the definition of the conversion. This conversion definition specifies how concepts of the data source are mapped onto concepts of the event log. The second step is to execute this conversion and actually convert the data from the data source to the event log as specified in the mapping. It is important to be aware of the influence of decisions made in the mapping phase on the resulting event log and hence the process mining results.”

The first step in a typical process mining project is to determine the goal of the project, as a result the scope and focus of the project can be determined. The scope determines what to include and the focus determines the level of detail needed for the conversion. Next, the traces pertaining to a specific process instance or case should be selected. Process instances are often business object such as claims, customers etc. For these process instances, relevant events should be selected. Additionally, important attributes for process instances and events can be selected (Buijs, 2010).

### 2.3.3 Process mining tools & algorithms

**Process mining tools**

For the execution of the conversion definition, i.e. converting raw data to an actual event log, different tools are available. The ProMImport framework supports the creation of MXML event logs from different types of data sources. Several standard adapters have been integrated into this framework, although it is possible for users to program their own adapters in Java. For the creation of XES or MXML logs, the 'XES Mapper' or 'XESMa' tool is available. The main advantage of the last tool is that it allows for the creation of XES logs and requires little to no programming (Buijs, 2010).

After the construction of an event log, it is possible to start the actual process mining. Van Dongen et al. (2005b) describe the generic, open-source framework ProM as a comprehensive tool for process mining. ProM 6.0 can use event logs in the MXML and XES log formats, while ProM 5.2 can only cope with the logs in the MXML format (Buijs, 2010). Although the ProM framework is targeted at an academic audience, it can be used by all kinds of organizations. It has been developed at Eindhoven University of Technology. The ProM 5.2 framework integrates over 280 different process mining plug-ins (van Dongen et al., 2005b), although it's version is currently at 6.0 (containing less plug-ins).

It should also be noted that some of the available plug-ins of the ProM framework have been transferred and further developed into the commercial tool Futura Reflect of Futura Process Intelligence. This tool shares a lot of the functionality with the ProM framework. In turn, elements of Futura Reflect have been integrated in the ReflectOne suite of Pallas Athena.

Other closely related tools are business process monitoring tools, such as the ARIS Process Performance Manager (PPM), developed by IDS Scheer. ARIS PPM offers functionality for the monitoring of business processes. However, it can only create process models if the process is known. Explicit process models have to be available beforehand (van Dongen & van der Aalst, 2005b). As such, these types of tools are not further considered in this research, since there is no real process 'discovery'.

**Process mining algorithms**

To give an example of some of the available plug-ins in the ProM framework, here is a list of some of the most frequently used algorithms for control-flow discovery:

- *α-algorithm* – Produces a place/transition net. The alpha algorithm is unable to mine certain constructs such as loops, duplicate and invisible tasks. Also, it cannot deal well with data that is incomplete or contains 'noise' (de Medeiros et al., 2003)
- *Heuristics miner* – The heuristics miner constructs a model based on the frequencies of tasks (Weijters et al., 2006). It is unable to deal with non-free choice and duplicate tasks. However, it is robust to noise in logs.
- *Fuzzy miner* – The fuzzy miner is best suited for mining less structured processes. It is able to abstract from details, although its design causes it to lack support for mining specific splits and joins in a process. The fuzzy miner can also deal well with noise (Gunther & van der Aalst, 2007).
- *Genetic algorithm* – The genetic algorithm constructs a process model according to an approach that is similar to the process of evolution in biological systems. It is able to deal with all constructs, apart from duplicate tasks. It is also robust to noisy logs. One drawback is the long computational time required (van der Aalst et al., 2005b)

## 2.4 Business process modelling

A research field that is closely related to process mining is business process modelling. Bandara (2007) defines business process modelling as: "an approach for visually depicting how businesses conduct their operations; defining and depicting entities, activities, enablers, events, states and the relationships between them." Green & Rosemann (2000) define a process model as: "the order of activities within a business transaction. A process model would typically include information about the control flow, the organizational units (*"internal and external persons / units that are involved in the process by having to execute certain tasks")*, input data, functions and output".

According to the author, business process modelling is similar to business process mining in two ways. First, regarding the BPM lifecycle of Weske (2002), see section 2.1, both process modelling and process mining can be used in the evaluation phase to reflect on past process performance and to possibly identify performance improvements. Secondly, process modelling shares a close similarity to the discovery dimension of process mining. In this respect, similar types of models can be constructed: control-flow models and organizational resource models. However, process modelling is less suitable for investigating conformance (although it could be used) than process mining, and extension is an analysis dimension which is not supported by process modelling.
Additionally, for discovery a large number of process modelling notations are also used in process mining. Frequently used techniques include: flowcharting, event-driven process chains (EPCs), business process modelling notation (BPMN) and Petri nets. With regard to process mining, the ProM framework also uses Petri nets and EPCs for instance to depict process models.

The main difference between business process modelling and business process mining is the approach that is used to construct the process models. According to Bandara (2007), there are six core phases in any process modelling initiative:

- goal identification
- process identification
- information gathering
- map/model generation
- analysis
- continuous improvement

The goal and process identification phases are straightforward, and are also important for conducting a process mining project (see section 2.3.3). The main difference between business process modelling and process mining lies in the information gathering, map/model generation and analysis phases.
Starting in the information gathering phase, business process modelling mostly uses qualitative data by conducting interviews with process experts, conducting workshops with process stakeholders, reviewing documentation and/or using simple flowcharting to reveal process information. These qualitative findings are integrated through the construction of a process model, which is done by a human process modeller (Green & Rosemann, 2000). However, few or no guidelines exist for conducting the information gathering, modelling and analysis phases (Bandara, 2007).
Van der Aalst et al. (2003) describe the issue of objectivity in this situation: "Modelling an existing process is influenced by perceptions, e.g., models are often normative in the sense that they state what 'should' be done rather than describing the actual process. As a result models tend to be rather subjective."

In comparison to business process modelling, process mining relies more heavily on 'objective' information system data to construct process models and less on qualitative information from process stakeholders. For process mining, process data is collected on the events that actually occurred, as registered by the information system. From this 'objective' data process models are then generated through explicitly defined algorithms, after which an analysis can be conducted in a more quantitative fashion.

In general, business process modelling is used for several purposes (Bandara, 2007): the identification of process weaknesses, adapting best business practices, the (re)design of businesses processes, end-user training, compliance and risk management and systems design and configurations. Regarding the BPM lifecycle of Weske (2002) in section 2.1, business process modelling can be used in the analysis phase to provide more insight into a process and ways of attacking potential problems. It can be used in the design phase to 'put an idea into concrete form' (for example, by defining the activities needed and their dependencies, data flows, roles, and goals etc.). Process modelling can also be used in the enactment phase to 'put design decisions into practice' (Bandara, 2007). In contrast, process mining is only used in the evaluation/diagnosis phase to analyse the performance of an enacted process.

## 2.5 Data mining

Another closely related field to business process mining is data mining. Data mining is defined by Fayyad et al. (1996) as: "the application of specific algorithms for extracting patterns from data."
Instead of data mining, the term knowledge discovery in databases (KDD) is often used in order to emphasize that 'knowledge' is the end product of data-driven discovery. As such, the field of KDD encompasses more than only data mining, which is only a particular step in the KDD process (Fayyad, 1996), see figure 2.7 below.

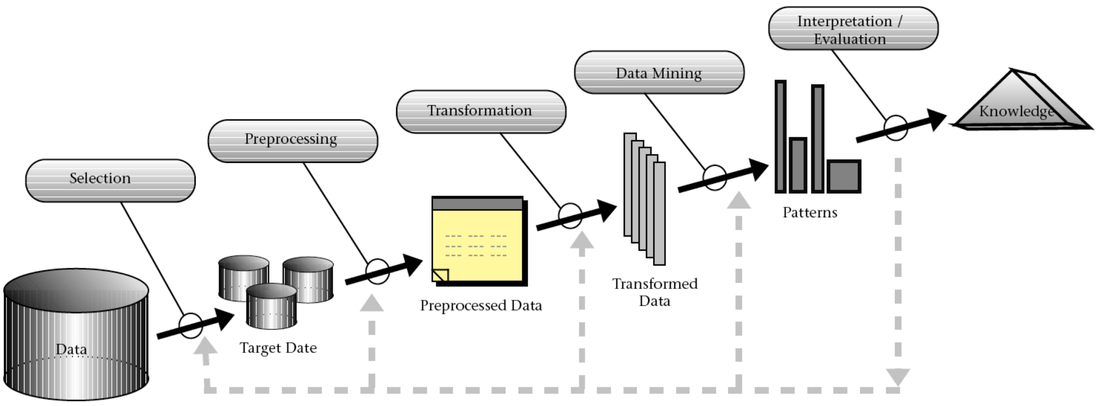


Figure 2.7 - An overview of the activities in the KDD process (Fayyad et al., 1996)

Figure 2.7 illustrates that relevant data should be selected first. Pre-processing should then take place to clean the data. Next, the data is transformed into the right format for analysis, followed by the actual data mining analysis. The last step is interpreting patterns to obtain knowledge. It should be noted that significant iteration occurs in the KDD process.
Basically, process mining can be seen as a specific form of data mining with a strong business process view. Some existing data mining techniques have been used in the context of process mining, such as clustering, while other algorithms such as the heuristics miner, fuzzy miner and genetic miner have been developed exclusively to mine processes. Additionally, process mining has its unique pre-processing standard in the form of event logs.

## 2.6 Success evaluation

In general, the measurement of information systems (IS) success is essential to understanding the value and efficacy of IS initiatives in organizations, and to justify investments in IS (adapted from DeLone & McLean, 2003). Process mining can be regarded as specific discipline within information systems research; as such it will also be important to measure the success of process mining initiatives. Furthermore, DeLone & McLean (1992) state that: "if information systems research is to make a contribution to the world of practice, a well-defined outcome measure (or measures) is essential". Seddon et al. (1998) agree that it is important to measure IS effectiveness, yet they argue against a single standard measure of IS success, and instead plead for more sharply-focused, context-specific success measures. The author feels this statement of Seddon et al. (1998) also holds for process mining, and that specific success measures for process mining projects are needed.

Unfortunately, after conducting the literature review described in Appendix A, no research could be found that was conducted on the topic of process mining success. For this reason, a general investigation of IS success measurement (section 2.6.1 & 2.6.2) and success measurement in the related disciplines of process modelling (section 2.6.3) and data mining (section 2.6.4) was conducted.

### 2.6.1. Perspective of success measurement

The importance of defining the perspective of IS success evaluation has been raised by several authors (Bandara, 2007; Seddon et al., 1998). Seddon et al. (1998) argue that: "in the measurement of IS effectiveness, researchers should always take care to identify, first, the stakeholder (in whose interest the evaluation of IS success is being performed) and, second, the specific system or class of system that is being evaluated. Different stakeholders are likely to use different criteria for evaluating IS effectiveness." Therefore, researchers should take care to specify which of the following stakeholder perspectives and systems are being evaluated (Seddon et al., 1998), see table 2.1 below:

| *Stakeholder perspectives* | *Information systems* |
|---|---|
| ○ an independent observer (who has no personal stake in the measure)<br>○ an individual user (who evaluates a system from his or her own point of view)<br>○ a group of users, e.g., of a group decision support system (GDSS)<br>○ the management or owners of the organization<br>○ a country, or mankind | ○ a single IT application (e.g., a spreadsheet, a PC, or a library cataloging system)<br>○ a type of IT or IT application (e.g., TCP/IP, a GDSS, a TPS, a data warehouse, etc.)<br>○ all IT applications used by an organization or suborganization<br>○ an aspect of a system development methodology<br>○ the IT function of an organization or sub-organization |

**Table 2.1 - Stakeholder and IS perspectives (Seddon et al., 1998)**

Bandara (2007) has adapted the work of Seddon et al. (1998) to identify the following stakeholders that are typically involved in a business process modelling project:

- *Model users* – Those who used, or will use the process models
- *Modellers* – Those who conduct the modelling project, where the role was to design the process models

- *Project sponsors* – those who provide the necessary resources to commence and sustain the process modelling project (often hold senior management positions)
- *Project leaders* – those who lobby for the process modelling project and drive the initiative from start to end

### 2.6.2 Information system success

DeLone & McLean (1992) were one of the first to develop a framework and model for measuring the complex dependent variable in information systems (IS) research. To this end, they proposed a multidimensional and interdependent concept of IS success. The authors attempted to reduce significantly the number of different measures used to measure IS success so that research results could be compared and findings validated. Since the initial publication of the IS success model, nearly 300 articles in refereed journals have referred to, and made use of, this IS success model (DeLone & McLean, 2003). In response to research efforts that have applied, validated, challenged, and proposed enhancements to the original IS success model. DeLone & McLean (2003) have made minor refinements to the model and propose an updated IS Success Model, see figure 2.8 below.

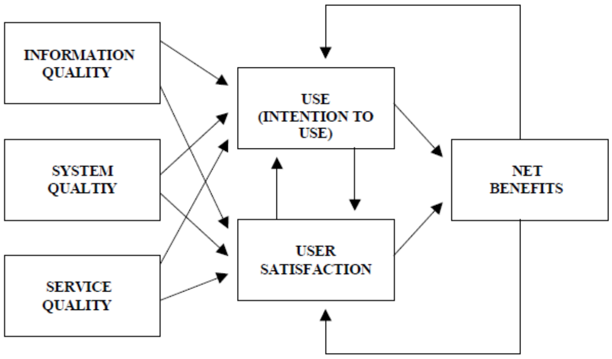


Figure 2.8 - Reformulated IS success model (DeLone & McLean, 2003)

In the IS success model, information quality is a measure of the information system output. System quality is a measure of the information processing system itself. Service quality is the provisioning of support for end user developers. Information use is the recipient consumption of the output of an information system. Given the difficulties in interpreting the multidimensional aspect of “use”, DeLone and McLean (2003) have suggested to measure the attitude “intention to use” instead of the behaviour “use”. User satisfaction is the recipient response to the use of an output of an information system. Net benefits entail the impact of the IS on certain chosen stakeholders or aspects, depending on the systems being evaluated and their purposes.

Delone and Mclean (1992) argue that researchers should systematically combine all measures from their six IS success categories in measuring IS success. However, Seddon et al. (1998) oppose this, because of: “the range of different systems, stakeholders, and issues involved in different studies, a wide diversity of sharply-focused dependent variables is essential.” The author agrees with this reasoning. This is why success evaluation in the closely related disciplines of process modelling and data mining was further investigated.

### 2.6.3 Process modelling success

For the field of business process modelling, Bandara (2007) has derived and adapted process modelling project success measures from the IS success model of DeLone & McLean (1992). Through a series of nine case studies, the initial a-priori process modelling success model was respecified.

Constructs of the refined model were operationalized through the design of a qualitative survey, which was distributed in the form of a questionnaire to two-hundred-ninety respondents. Using statistical validation and confirmation, a final, comprehensive model of process modelling success was developed. This model includes the following success measures (Bandara, 2007):

- *Model quality* – The extent to which all desirable properties of a model are fulfilled to satisfy the needs of the model users in an effective and efficient way.
- *Process impacts* – Refers to the overall effect of the initiative on the processes modelled.
- *Project efficiency* – The ratio of obtained outcomes over invested resources.

It was shown in section 2.4 that business process mining and business process modelling provide the same type of output in the form of business process models when it concerns the analysis dimensions of discovery. However, business process modelling is not often used for conformance analysis and effectively not at all for extension.
Regarding discovery, both methods only differ in the process of how to arrive at these results and the qualitative versus quantitative nature of results. Due to their similarity in these project outcomes, process modelling success measures could possibly be adapted to evaluate process mining project success.

### 2.6.4 Data mining success

Nemati & Barko (2003) propose to use the square root (TSR) framework of Atkinson (1999) to measure the success of data mining projects. The TSR frameworks integrates well-researched project success measures of the 'iron triangle' (cost, time and quality), with information system success measures and organizational and stakeholder community benefits (Atkinson, 1999). A breakdown of the specific success measures within the TSR framework is given in figure 2.9 below.

| Iron Triangle | The information system | Benefits (organisation) | Benefits (stakeholder community) |
|---|---|---|---|
| Cost | Maintainability | Improved efficiency | Satisfied users |
| Quality | Reliability | Improved effectiveness | Social and |
| Time | Validity | Increased profits | Environmental impact |
| | Information– | Strategic goals | Personal development |
| | quality | Organisational-learning | Professional learning, |
| | use | | contractors profits |
| | | | Capital suppliers, content project |
| | | | team, economic impact to |
| | | Reduced waste | surrounding community. |

Figure 2.9 - Breakdown of success criteria in the TSR framework (Atkinson, 1999)

## 2.7 Success factors

In this research, a distinction is made between success measures and success factors. Success measures are criteria to evaluate a project, whereas success factors are the set of influential forces that contribute directly or indirectly to the success of a project (adapted from Nemati, 2003). In contrast to success measures (dependent variables), success factors are the independent variables in the construction of a process mining success model (after Bandara, 2007). This section aims to summarize the different success factors that have been mentioned in research on: process mining (section 2.7.1), process modelling (section 2.7.2) and data mining (2.7.3).

### 2.7.2 Process modelling success factors

Bandara (2007) presents a model of process modelling project success in which certain success factors influence the success of a process modelling project. Figure 2.10 below displays the final process modelling success model.

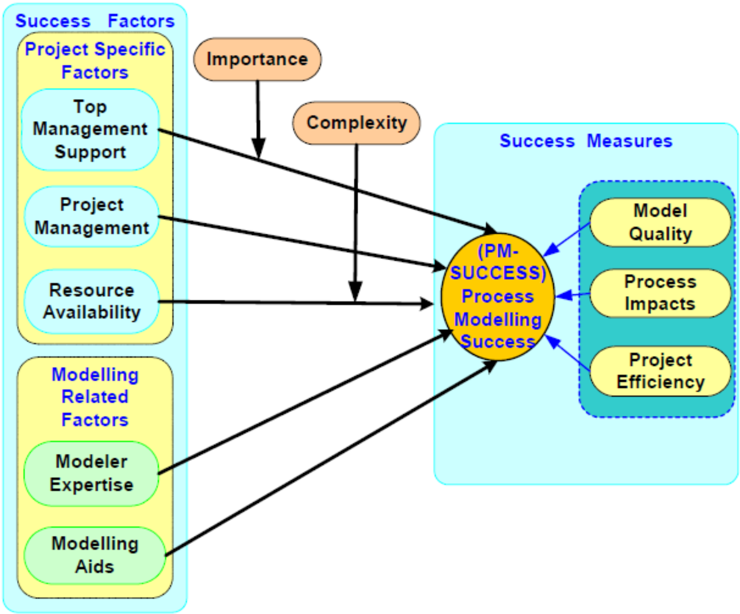


Figure 2.10 - Final process modelling success model (Bandara, 2007)

The process modelling success factors in figure 2.10 are defined by Bandara (2007) as follows:

| ***Project Specific factors*** | ***Definition*** |
|---|---|
| Top management support | The level of commitment by senior management in the organization to the process modelling project, in terms of their own involvement and the willingness to allocate valuable organizational resources. |
| Project management | The management of the process modelling project including defining the project scope, aims, milestones and plans. |
| Resource availability | Consolidation of user participation and information resources. Degree of information available from project stakeholders for the design, approval and maintenance of the models |
| ***Modelling related factors*** | ***Explanation*** |
| Modeller expertise | The experience of the process modellers in terms of conceptual modelling in general and process modelling in particular. |
| Modelling aids | Consists of modelling guidelines (detailed set of instructions for conducting process modelling), modelling tools and modelling languages |
| ***Moderating factors*** | ***Explanation*** |
| Process importance | Refers to the criticality of process modelling for the underlying purpose |
| Process complexity | The many different features of the processes modelled |

Table 2.5 - Process modelling success factors (Bandara, 2007)

### 2.7.3 Data mining success factors

Nemati & Barko (2003) found that in organizational data mining (ODM) projects, key success factors are: data quality, data integration, technical integration and expertise, an ODM outsourcing strategy and level of end-user expertise, and project scope, length and resources.

Data quality can be affected by missing values or incorrect values. Data integration is defined as the integration of data from multiple data sources. Technical integration is defined as the integration of technological components (such as systems integration, middleware, customer interfaces etc.). Technical expertise is the technological expertise to achieve technical integration. ODM outsourcing refers to the presence of an outsourcing strategy if the organization has no experience with data mining. The definition of end-user expertise used by Nemati & Barko (2003) remains unclear.

Feelders et al. (2000) complement these factors by stating that: "successful data mining projects require the involvement of expertise in data mining, company data, and the subject area concerned. Despite the attractive suggestion of `fully automatic' data analysis, knowledge of the processes behind the data remains indispensable in avoiding the many pitfalls of data mining".

## 2.8 Chapter conclusion

This chapter has introduced the basic concepts of process mining. Although success evaluation is an important topic, it was shown that little to no research has been conducted on the topic of process mining success. However, research on success evaluation does exist in the closely related research disciplines of information systems, process modelling and data mining. For instance, a comprehensive process modelling success framework was developed by Bandara (2007). Process mining success research could borrow concepts, methods and ideas from research in the aforementioned areas in order to develop knowledge on how to conduct process mining project successfully. The next chapter will describe the research design that was chosen to address this challenge.

# 3. Research design

Within this chapter the methods are described that were used to develop knowledge in the area of process mining success. Recall that the main research goal is to provide knowledge on how organizations can successfully conduct process mining projects. Little research has been conducted in this area.
The research design was inspired by the research strategy (described in section 1.3) which calls for a convergence of theory and practice through the creation of design principles. Since little theory was available in the area of process mining success, new theory had to developed. Subsequently, the theoretical findings were contrasted with practice to increase the validity of results and practical relevancy. For this reason, an important distinction was made between the theory-oriented research part (described in section 3.1) which aimed at developing theoretical knowledge, and the practice-oriented research part (described in section 3.2) which aimed to evaluate/respecify the outcomes of the theory-oriented research.

## 3.1 Theory-based research

It is evident that there is a lack of theory with respect process mining success. For this reason, several previously conducted process mining case studies were investigated first to look for implicitly mentioned success factors and success measures. This procedure is described in section 3.1.1. The next step was to combine this knowledge with the known success factors and measures from related disciplines in an a-priori model that could be used to focus subsequent research efforts. This is explained in section 3.1.2. Lastly, to aid practitioners with specific design knowledge on how to successfully conduct a process mining project, the procedures for extracting theory-based design principles are described in section 3.1.3.

### 3.1.1 Analysis of success factors and measures

During the systematic review process, described in Appendix A, 15 publications or cases studies were found that deal with the application of process mining in a practical setting. These publications were further analyzed to identify potential success measures. Regarding the section on process mining success measures and process mining success factors, a meta-synthesis approach was used that shares a great similarity with the grounded theory approach of Corbin & Strauss (2008).

To start the coding process, the set of 15 practice-oriented publications were imported in the qualitative data analysis tool NVivo. In light of the research questions, nodes for two important categories were formed: success factors and success measures. Through open coding, passages mentioning success factors or success measures were descriptively coded under a node with a name closely related to the actual piece of coded text (Miles & Huberman, 1994, p.57).
To create a more manageable set of success factors, similar nodes were (re)grouped in categories through a step called 'axial-coding' (Corbin & Strauss, 2008).
To reduce bias in the theoretical coding process, another Operations Management & Logistics student assisted in the coding of the first 5 practice-oriented publications. This resulted in an inter-coder reliability of 85%. Remaining coding issues and questions were resolved between the author and the additional coder. Coding was stopped when no additional concepts were found.

### 3.1.2 A-priori model construction

An a-priori model is basically a conceptual model of the relationships between the measures that are used to evaluate process mining success and the relationships they have with the success factors that affect these variables. Yin (2003) posits that the a-priori development of theory is essential in both theory-building and theory-testing research to focus research efforts. The author agrees with this view: due to the complexity of process mining it is important to know what you are looking for.

Through combining found process mining success factors, the process modelling success framework of Bandara (2007) and information systems and data mining success factors and measures, a hypothesized a-priori model of process mining success was constructed.

Following Eisenhardt (1989), it is important to recognize that the a-priori constructs were tentative in subsequent research. No construct was guaranteed a place in the resultant theory, no matter how well it was measured.

### 3.1.3 Distillation of theoretical design propositions

The identified success factors that contribute to process mining success merely present a 'static' description of what is necessary to achieve success. However, they do not indicate how to achieve a specific success factor. This is why an attempt was made to distil design principles from the previously conducted, practice-oriented process mining case studies (see Appendix A). Design principles can provide prescriptive knowledge to practitioners in the form of the so-called 'CIMO' logic, which stands for (C)ontext, (I)ntervention, (M)echanism and Outcome, see figure 3.1 below for a detailed explanation.

| Component | Explanation |
|---|---|
| Context (C) | The surrounding (external and internal environment) factors and the nature of the human actors that influence behavioural change. They include features such as age, experience, competency, organizational politics and power, the nature of the technical system, organizational stability, uncertainty and system interdependencies. Interventions are always embedded in a social system and, as noted by Pawson and Tilley (1997), will be affected by at least four contextual layers: the individual, the interpersonal relationships, institutional setting and the wider infrastructural system. |
| Interventions (I) | The interventions managers have at their disposal to influence behaviour. For example, leadership style, planning and control systems, training, performance management. It is important to note that it is necessary to examine not just the nature of the intervention but also how it is implemented. Furthermore, interventions carry with them hypotheses, which may or may not be shared. For example, 'financial incentives will lead to higher worker motivation'. |
| Mechanisms (M) | The mechanism that in a certain context is triggered by the intervention. For instance, empowerment offers employees the means to contribute to some activity beyond their normal tasks or outside their normal sphere of interest, which then prompts participation and responsibility, offering the potential of long-term benefits to them and/or to their organization. |
| Outcome (O) | The outcome of the intervention in its various aspects, such as performance improvement, cost reduction or low error rates. |

Figure 3.7 – Components of design propositions (Denyer et al., 2008)

The approach of Denyer et al. (2008) was followed to distill theoretical design propositions from the set of practice-oriented publications. The efforts in creating an a-priori process mining success model can actually be seen as an initial step for creating theory-based design principles since both are related in the following way:

- **(C)ontext** – is made up of the process mining project (unit of analysis) and the mediating/contextual variables within the a-priori model
- **(I)ntervention** – are the success factors in the a-priori model. They can be seen as high-level interventions of the design principle
- **(O)utcome** – are the success measures in the a-priori model

As such, the a-priori model already contains the basic ingredients for developing theoretical design propositions. High-level information was already coded under nodes in Nvivo. However, these design propositions needed to be more specific. Therefore, for each of the success factors in the a-priori model, text searches were conducted in the qualitative data analysis tool Nvivo to see if passages that mentioned specific interventions to realize these success factors could be identified. If this was the case, a separate node was created for this specific intervention. These specific interventions were then consolidated in one design principle for each of the success factors/interventions in the a-priori model.

## 3.2 Practical evaluation/respecification

The next step was to evaluate or respecify the theoretical a-priori model. Deliberately, the term respecification is chosen instead of model-building or model testing, since both activities took place (adapted from Bandara, 2007). The respecification of the a-priori model was performed through a multiple case study (section 3.2.1). Additionally, design principles that were based on the data from practice were distilled since they can provide practitioners with specific knowledge on how to conduct process mining projects successfully.

### 3.2.1 Multiple case study design

A case study design was chosen because it is generally considered a viable research strategy within information systems research when (Benbasat, 1987):

- The researcher wants to study information systems in a natural setting, learn about the state of the art and generate theories from practice
- The researcher wants to ask how and why questions, that is to understand the nature and complexity of the processes taking place
- Few previous studies have been carried in the research area

These conditions seemed to hold for this project: the aim of this thesis was to develop knowledge on the practical application of process mining. Therefore, it was important to study process mining projects in a natural setting. It was also critical to understand the complexity within the process mining projects to be able to build the cumulative body of knowledge on process mining success and to develop some design principles, Furthermore, as demonstrated in the previous chapter, little research has been conducted on this topic.

Several other considerations were addressed in the case study design. Firstly, a multiple case study design was chosen, instead of a single case study, since it is an appropriate method to use when the purpose is to build and/or test theory. Single case studies are often criticized for their lack of generalizability (Bandara, 2007, p.100).
Secondly, it was noted that this multiple case study had characteristics of both theory-testing and theory-building research. It was theory-testing in the sense that propositions relating to the project-related factors, some mediating factors and success measures (see Chapter 2) had already received strong support in the context of process modelling (Bandara, 2007), the question was if these propositions would hold for process mining projects as well. However, the multiple case study was

also theory-building in the sense that new, less developed propositions regarding process mining related factors, information systems related factors and other mediating effects were posited, and enough flexibility was retained to possibly formulate new propositions.
As the majority of case study methods either stress theory-building or theory-testing research, it was difficult to find the most appropriate case study approach. A mix of the approaches of Bandara (2007), Eisenhardt (1989) and Yin (2003) was used, which consisted of the following phases: case selection, data collection (design), data coding and analysis and shaping of hypotheses and sharpening of constructs.

#### *3.2.1.1 Case selection*

In this study, a case was defined as a process mining project (see section 1.2). Recall that process mining is used in the evaluation/diagnosis phase of the BPM lifecycle to evaluate the performance of an enacted business process. In the definition of the author, the process mining project includes the derivation of process models, mainly through control-flow discovery and possibly some conformance checking. Additionally, it had to be the case that the process mining activities were conducted within an organization and were organized in a project that had a certain business objective.
Recall that a literature review was conducted in Chapter 2. During the systematic review procedure, described in Appendix A, 15 previously conducted process mining case studies were identified that were well documented in a scientific publication. In these already conducted case studies, process mining was applied in practical setting. Initially, the author sought to select cases from this population. Theoretical sampling was used as a sampling method: cases were chosen to fill theoretical categories and provide examples of polar types (Eisenhardt, 1989).
The moderating factors 'type of IT system' and 'process complexity' of the a-priori model were chosen to guide the selection of case studies. These contextual variables were hypothesized to have an important effect on several success factors. For 'type of IT system', cases were sought where data originated from different types of information systems along the CSCW spectrum. These cases were hypothesized to be distributed along the CSCW spectrum as depicted in figure 3.2 below. Additionally, cases where sought that were hypothesized to vary in the degree of process complexity.
Due to scoping reasons only a limited number of cases could be selected. Additional selection constraints were specified as follows:

- Process mining stakeholders had to be available for interviews
- The project had to be conducted quite recently

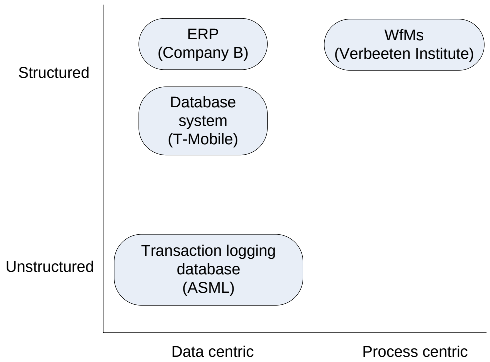


**Figure 3.2 - Classification of cases along the CSCW spectrum**

| *Case study* | *Process complexity* |
|---|---|
| ASML | High |
| Company B | Medium |
| T-Mobile | High |
| Verbeeten Institute | Low |

**Table 3.3 - Hypothesized case values on process complexity**

To fill the bottom left and top right quadrants of figure 3.2, the **Verbeeten Institute** and **ASML** case studies were selected from the initial population of previously conducted case studies. As figure 3.2

shows, the Verbeeten case and ASML cases both represent a distinct type of information system, although the Verbeeten case was hypothesized to have a low process complexity, as opposed to the high process complexity of ASML (see table 3.1).

Within the initial population, no case studies could be found for the ERP system, as none of the relevant case studies were conducted within a practical setting that had a clear business objective. Still, it was deemed important to include an ERP case study, as it represents a distinct situation in which process mining is quite difficult (as demonstrated in Chapter 2). After talks with process mining experts on the Eindhoven University of Technology, the only known available **Company B** case was selected. The Company B case was hypothesized to have medium process complexity (table 3.1).

No suitable cases could be found for the bottom right quadrant of figure 3.2 (unstructured and process centric). Since process mining within an ERP system represents quite a distinct situation in the sense that is notoriously difficult, the author wanted to select an additional case study to contrast results in this quadrant. Again, no additional, suitable cases could be found for an unstructured/process centric system with high process complexity. This is why the author himself conducted a process mining project at **T-Mobile Netherlands**.

#### 3.2.1.2 Data collection

Case study researchers commonly use data collection methods such as archives, interviews, questionnaires and observations. It is advised to combine different methods to achieve triangulation (Eisenhardt, 1989). The main data collection methods that were used in this multiple case study were: interviews, documentation and participant observation.

Interviews were the most significant data collection method used in this study. A semi-structured interview approach similar to Bandara (2007, p.114) was followed. All interviews followed the same format described in Appendix C. First, an open discussion was held on perceived success factors and measures in relation to the process mining project. Subsequently, the constructs of the a-priori model were introduced, and opinions on the relevance and importance of the constructs were collected. This approach enabled new ideas or constructs to emerge, but also enabled the validation of the a-priori model.

Whenever it was possible, different process mining project stakeholders were interviewed. Their respective role classification was adapted from Bandara (2007):

- *Model user* – Those who used, use or will use the models or results provided by process mining
- *Process miner* – Those who are conducting the actual process analysis within the process mining project, either internal (employed by the organization) or external
- *Project sponsor* – Those who provide the necessary resources to commence and sustain the process mining project (often have a senior management role)
- *Project leader* – Those who drive the process mining initiative from start to end

If it was possible, the project sponsor(s) and process miners were interviewed at a case site. However, project sponsor were not available at the ASML and Company B case sites. Most of the times, the project leader role was taken by a process miner (except for ASML). Model users were hard to identify and were therefore not interviewed. Table 3.2 below shows the specific stakeholders that were interviewed at each case site

| Case site | Stakeholders |
|---|---|
| ASML | ○ External process miner<br>○ Project leader |
| Company B | ○ External process miner<br>○ Internal process miner |
| T-Mobile | ○ Project sponsor 1<br>○ Project sponsor 2<br>○ External process miner (author by participant observation) |
| Verbeeten Institute | ○ External process miner<br>○ Project sponsor |

Table 3.4 - Interviewed stakeholders at each case site

It should be noted that participant observation was used to collect data for the external process miner role at T-Mobile. The author kept a journal of his process mining experiences at T-Mobile, which was adapted and used in further analysis. Documentation, including a master thesis concerning the process mining at the Verbeeten Institute (Staal, 2010) and a journal publication on process mining at ASML (Rozinat, 2006b) were used in support of the interview transcripts.
All of the interviews were conducted by a single researcher, and lasted about 45 to 60 minutes. The researcher took field notes during the conduct of these interviews. Furthermore, all interviews were electronically recorded and transcribed later on. The interview transcripts and documentation were maintained through a case database in Nvivo.

#### *3.2.1.3 Data coding and analysis*

**Codifying the data**

This section describes the procedures that were used for the coding the collected data. Miles and Huberman (1984) state that*:* "Codes are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study".
A similar approach to coding as executed by Bandara (2007, p. 146) was followed. Prior to any data analysis, a node structure was created in Nvivo to mirror the constructs of the a-priori model. All eight interview transcripts and the observation report were imported into Nvivo and stored in their respective case site folders. The coding of the interview transcripts then took place in the following order:

1. Coding only took place when relevant passages (mentioning something similar to the definition of the a-priori constructs) in the documents were found
2. Passages of text that generally mentioned an existing construct in any way were coded under the relevant node of the a-priori model. If new ideas or constructs were found that could not be easily placed under an existing construct, they were coded under a new node
3. When new ideas or constructs were found, previous transcripts were scanned again to look if all data relating to these new constructs was captured
4. General coded citations within each node were further investigated to distinguish between citations that stressed the importance or unimportance of constructs. Positive or negative citations were coded as sub-nodes within the general construct
5. Furthermore, the general citations were screened for potential sub-constructs, to identify more specific descriptions of the construct (specific interventions of the design principles)

**Analyzing the data**

Case study data was analyzed to answer the following questions (after Bandara, 2007, p. 121):

- Are all the important constructs captured in the a-priori model?
- Are there any constructs in the a-priori model that are not critical or relevant for process mining success?
- Are any of the constructs identified redundant?
- What type of relationship exists among the constructs (e.g. moderation)?

Typically, data analysis in case study research consists of within-case analysis and cross-case analysis. Within-case analysis mainly serves to become intimately familiar with the case site data. A method commonly used for this is the construction of detailed write-ups for each case site. These write-ups are often quite descriptive in nature (Eisenhardt, 1989). Section 4.1 holds these case site descriptions.

After the within-case analysis, a cross-case search for patterns was undertaken. Often, this is a critical step in qualitative case study research, which is underlined by Eisenhardt (1989): "the tactics here are driven by the fact that people are notoriously poor processors of information. The key to good cross-case comparison is counteracting these tendencies by looking at the data in many divergent ways. One tactic is to select categories or dimensions, and then to look for within-group similarities coupled with intergroup differences. Dimensions can be suggested by the research problem or by existing literature, or the researcher can simply choose some dimensions. A second tactic is to select pairs of cases and then to list the similarities and differences between each pair. A third strategy is to divide the data by data source."

The cross-case analysis was performed by comparing the number of general citations (merely mentioning the construct), important citations (mentioning the importance of the construct) and unimportance citations (mentioning the unimportance of the construct) for each construct across the different case sites. Coded data from the interviews and participant observation yielded counts of the number of citations for each construct. The number of citations for each construct was summarized in a cross-case comparison matrix which was created through a matrix coding query in Nvivo.

To test for redundancy, relatedness and possible moderating effects of constructs, a different matrix intersection search was conducted Nvivo which served to identify passages that were coded under multiple constructs. If a passage of text is coded under multiple constructs, it might indicate possible overlap or a moderating relationship. The Nvivo tool allows checking these specific instances to let the qualitative researcher judge what is actually happening.

#### *3.2.1.4 Shaping hypotheses and sharpening constructs*

The results of all these data analysis procedures where taken together to enable the respecification of the a-priori model, based on the case study data (adapted after Bandara, 2007, p.380). Following the same logic of Bandara (2007, p.380) an a-priori construct could be deleted for one or more of the following reasons:

- having only a few general citations, and /or
- having one or two strong citations for its irrelevance as a construct, and /or
- the data coded under this construct also having been coded under another – hence depicting possible redundancy or overlap

In a similar fashion, a new construct could be included based on one or more of the following reasons Bandara (2007, p.380):

- having many general citations, stating this new construct, and /or
- having one or two strong citations for its high relevance as a construct and /or
- an overarching theme (construct) identified that captures more than one of the a-priori constructs

#### *3.2.1.5 Validation*

The methodological quality of any case study research design can be evaluated on the basis of four tests (Yin, 2003):

- *Construct validity* – establishing correct operational measures for the concepts being studied
- *Internal validity* – establishing a causal relationship, whereby certain conditions are shown to lead to other conditions, as distinguished from spurious relationships
- *External validity* – establishing the generalizability of findings
- *Reliability* – ensuring that a later researcher following the same procedures and conducting the same case studies would arrive at the same findings and conclusions

### 3.6.1 Construct validity

To increase construct validity, multiple sources of evidence (interviews, documentation and participant observation) from multiple cases were used in data collection to achieve data triangulation.

Additionally, the researcher tried to maintain a clear 'chain of evidence' by enabling readers to trace back conclusions to the actual research questions (Yin, 1994). Also, the creation of a case database in NVivo helped the researcher to maintain a 'chain of evidence', enabling the review of actual passages of coded text within a certain node or within a cell of the matrix coding query table.

Although the author performed the majority of the coding, to increase the validity of the construct coding process, the author enlisted the help of another process mining graduate student (Jeroen Veldhoen). He was given the definition of constructs of the a-priori model and coded 4 interview transcripts from different case sites. Differences in coding were discussed to improve the coding performed by the author.

### 3.6.2 Internal validity

Internal validity is of concern for causal or explanatory case studies during data analysis. Pattern matching and explanation building are seen as appropriate analytical strategies for increasing internal validity (Yin, 2003).

Pattern matching compares an empirically based pattern with a predicted one. If the patterns match, the internal validity of the study is enhanced. Explanation building is a special type of pattern matching where a phenomenon is explained by stipulating a presumed set of causal links about it (Yin, 2003). Both pattern matching and explanation building were used in data analysis. However, explanation building was mostly used when a change had to be made to the a-priori model.

### 3.6.3 External validity

External validity was increased by having constructed a theoretical a-priori model beforehand, which could be used to contrast the findings in the respecified model. The generalizability of findings was increased by selecting multiple cases that had different environmental (or mediating) factors.

### 3.6.4 Reliability

Although an explicit case study protocol was not developed, this chapter and the interview protocol in Appendix C comprehensively describe the steps that were taken to arrive at the research findings and conclusions. This should enable a different researcher conducting this research to arrive at the same findings and conclusions. Also, the construction of a case database helped to increase the reliability of results.

### 3.2.2 Derivation of design principles

The method of Plsek et al. (2007) was used to derive a set of practice-based design principles. These principles help to convert the tacit knowledge of organizational change agents into explicit, actionable knowledge and are often formulated in the 'CIMO' logic. The efforts in creating a respecified model can actually be seen as an initial step to creating practice-based design principles. Both are related in the following way:

- **(C)ontext** – is made up of the process mining project (unit of analysis) and the mediating/contextual variables within the a-priori model
- **(I)ntervention** – are the success factors in the a-priori model. They can be seen as high-level interventions of the design principle
- **(O)utcome** – are the success measures in the a-priori model

Several way of deriving practice-based principles are presented by Plsek et al. (2007):

- Reviewing written documentation of change programs in order to extract design rules
- Convening groups of change experts and asking them to describe what they do, or see themselves as doing, in the form of design rules
- Listening to stories of change efforts told by change leaders, operational managers, and front-line staff and then extracting design rules off-line (e.g., via review of transcripts or notes)
- Posing hypothetical scenarios to those experienced in organizational change, asking them to "think aloud" about how they would approach the situation, and then extracting design rules off-line.

Reviewing written documentation and listening to change efforts were used to construct the respecified model of process mining success. The respecified model already contained the basic ingredients for developing practice-based design propositions. High-level information on interventions was already coded under nodes in Nvivo. However, these design propositions needed to be more specific. Therefore, for each of the success factors in the respecified model, text searches were conducted in the qualitative data analysis tool Nvivo to see if passages that mentioned specific interventions to realize these success factors could be identified. If this was the case, a separate node was created for this specific intervention. These specific interventions were then consolidated in one design principle for each of the success factors/interventions in the a-priori model (where possible).

## 3.3 Chapter conclusion

This chapter presented the methods that were used to develop more knowledge on the topic of process mining success. In general, these methods consisted of a theory-oriented and a practice-oriented part. The results of the theory-oriented research will be presented in chapter 4. The results of the practice-oriented part are given in Chapter 5.

# 4. Theory-based research results

This chapter presents the results of the conceptualization of process mining success and the distillation of theory-based design principles for achieving process mining success, based on scientific literature found in the systematic review (Appendix A). Not much has been published in the research area of process mining success. Therefore, the results of this chapter were invaluable for the subsequent practice-oriented research.

To start off, section 4.1 deals with the analysis of implicitly mentioned success factors and success measures in process mining literature which could possibly be included in the a-priori model. Section 4.2 deals with the construction of the a-priori model of process mining success.

## 4.1 Analysis of success factors and measures

Although critical success factors or success measures were not found to be explicitly mentioned in process mining publications, the previously conducted process mining case studies were coded for implicitly mentioning of success measures (section 4.1.1) or success factors (4.1.2). Coding occurred when a construct was merely mentioned.

### 4.1.1 Process mining success measures

Most previously conducted case studies only concluded something about the applicability of process mining in a practical setting. In the view of the author, applicability can be seen more as a necessary step to achieving process mining project success. This raises the question if applicability alone is a suitable evaluation criterium. Often, the evaluation of project outcomes was done from the perspective of the person conducting the process mining.
Since no explicit success measures were found, the practice-oriented publications were coded for mentioning specific project outcomes. This was done with the use of the NVivo tool. Results are displayed in table 4.1 below.

| Outcome | Number Of Coding References |
|---|---|
| Increased process insights | 4 |
| Usefulness of process models | 2 |
| Objective support for organizational measures | 1 |
| Understandability of models | 1 |

Table 4.1 - Descriptive coding of process mining project outcomes

As indicated in table 4.1, several publications mentioned the creation of new insights into the process as an important process mining project outcome. Usefulness of the process model, objective support for organizational measures and understandability of models can be seen as measures to evaluate how satisfied a specific model user is with the process mining results.
Still, it remains unclear if this is a set of comprehensive measures that can be used to properly evaluate the success of process mining projects in organizations. For this reason, success evaluation in related disciplines is investigated.

### 4.1.2 Process mining success factors

The case studies were also coded for process mining success factors with the use of NVivo when they were mentioned. The open coding phase resulted in the following set of descriptive codes for potential success factors:

| # | Concept | Coding References | # | Concept | Coding References |
|---|---|---|---|---|---|
| 1 | Domain knowledge | 4 | 12 | Event log construction | 1 |
| 2 | Involving the right people | 2 | 13 | Representativeness of the data | 1 |
| 3 | Understanding of the process | 1 | 14 | Type of information in data set | 1 |
| 4 | Ontological knowledge of the IT system(s) | 3 | 15 | Structure of the process | 2 |
| 5 | General process description available | 2 | 16 | Process variability | 2 |
| 6 | Understanding of the system | 2 | 17 | Logging functionality | 3 |
| 7 | Amount of data available | 3 | 18 | Type of IT system | 2 |
| 8 | Characteristics of the data | 3 | 19 | Combining mining perspectives | 1 |
| 9 | Characteristics of the event log | 1 | 20 | Detailed explanation of results | 1 |
| 10 | Conformance to process mining requirements | 3 | 21 | Discuss process mining results with client | 1 |
| 11 | Data quality | 1 | | | |

Table 4.2 - Descriptive coding results of open coding phase

The entries in table 4.2 were reorganized and condensed through a more interpretive coding step into the following categories (see table 4.3 below):

- Number 1 till 6 – Domain knowledge
- Number 7 till 14 – Data and event log quality
- Number 15 till 16 – Process complexity
- Number 17 till 18 – Type of IT system
- Number 19 till 21 – Process mining approach

| Concept | Coding References | Concept | Coding References |
|---|---|---|---|
| Domain knowledge | 9 | Type of IT system | 2 |
| Data and event log quality | 7 | Process mining approach | 2 |
| Process complexity | 4 | | |

Table 4.3 - Interpretive coding results

Table 2.3 illustrates that 5 general process mining success factors were found, namely domain knowledge, data and event log quality, process complexity, type of IT system and process mining approach. These factors are defines as follows:

*Domain knowledge* – is mentioned in two different forms: business knowledge and systems knowledge**.**
Business knowledge consists of process expertise (Maruster & van Beest, 2009) and organizational expertise. Often, this type of knowledge is held by process owners. It seems crucial to be involved with these people (van der Aalst, 2005).
Systems knowledge is defined as the ontological knowledge of the data structure and relationships within the IT system. It seems that for complex information systems, e.g. SAP, the need for systems knowledge increases (van Giesssel, 2004; Segers, 2007; Ingvaldsen, 2008).

*Data and event log quality* – process mining needs data of sufficient quality. First of all, there needs to be enough data available to be able to draw any conclusions (Driessen, 2006; Rusu, 2010), although Ingvaldsen (2008) also notes that too much data can complicate the analysis. Furthermore, the data needs to be representative of the actual process under consideration (Driessen, 2006). To be able to construct the event logs from raw data, the data should meet minimum process mining requirements (as described in section 2.3.2). The degree of noise in the data can also complicate the process mining analysis (section 2.3.4). Noise refers to the fact that certain parts of the event log might be incorrect, incomplete or refer to exceptions.

*Process complexity* – an important process characteristic in the context of process mining is the structuredness of the process, which was already mentioned in sections 2.1 and 2.2 as well. Unstructured or highly variable processes provide a challenge for traditional process mining approaches (Mans, 2009). Structuredness of the process has to be taken into account to choose the proper process mining algorithms (Gunther, 2008). Another important point to consider is the rate of change in a business process, which might outdate process mining results (De Boer, 2010).

*Type of IT system* – Several authors stress the difficulty of applying process mining in an ERP system (van Giessel, 2004; Ramesh, 2006; Segers, 2007; Ingvaldsen, 2008). This is mainly due to the way in which these ERP systems log the transactional data that is used in process mining (van Giessel, 2004). It helps if a system has a logging functionality in place that enables an easy event log conversion. Maruster & van Beest (2009) show that different systems have different logging functionalities, which lead to different preprocessing/event log conversion steps.

*Process mining approach* – Some authors (Rusu, 2010; Ingvaldsen, 2008) mention the importance of the iterative nature of process mining, much like the importance of iteration in data mining (see section. Furthermore, van der Aalst et al. (2005) stress the importance of combining different (control-flow, organizational, case) mining perspectives. Bozkaya et al. (2009) indicate that is helps to discuss the process mining results with the client to create a better understanding of the system and avoid misinterpretation. Furthermore, it is important to take the existence of noise into account in choosing the right process mining algorithms (van der Aalst et al., 2005).

## 4.2 A-priori model construction

Success factors and measures in the areas of process mining, business process modelling and data mining have been derived in previous sections. The aim of this section is to integrate these findings into an a-priori model of process mining project success that can be further used in the subsequent research (after Bandara, 2007), as advocated by Eisenhardt (1989): "A-priori specification of constructs can also help to shape the initial design of theory-building research. Although this type of specification is not common in theory-building studies to date, it is valuable because it permits researchers to measure constructs more accurately. If these constructs prove important as the study progresses, then researchers have a firmer empirical grounding for the emergent theory."

Section 4.2.1 deals with the inclusion of potential process mining success factors. Section 4.2.2 describes the inclusion of potential success measures. Section 4.2.3 presents the a-priori model of process mining success.

### 4.2.1 A-priori success factors

Naturally, all of the process mining success factors that were found in section 4.1.2 are included in the a-priori process mining success model:

- Domain knowledge
- Data and event log quality
- Process complexity
- Type of IT system
- Process mining approach.

The question is how to organize these factors into common themes of factors in a similar fashion to Bandara (2007). In the process modelling success model of Bandara (2007), success factors are organized into: project specific, modelling related and mediating factors.

The **project specific factors** are included in the process mining success model due to the close similarity in project outcomes of process modelling and process mining projects, and the fact these factors are expected to hold for different types of projects since these originated from project management literature. The following factors are included under project specific factors:

- *Top management support*
- *Project management*
- *Resource availability*

*Top management support* - is kept as important process mining success factor (after Bandara, 2007).

*Project management* - is kept as important process mining success factor. Project management is validated as an important success factor, since it also mentioned as an important factor influencing data mining success (Nemati & Barko, 2003).

*Resource availability* - is defined by Bandara (2007) as: “The degree of information available from project stakeholders for the design, approval and maintenance of the models”. In previous sections we have seen that process mining is very much dependent on information system data, in contrast to process modelling. So this factor would be extended with the sub-construct of data availability.
Additionally, one specific element of domain knowledge, the availability of business knowledge from process stakeholders, is categorized under resource availability as well. It seems valuable to be able to use stakeholder knowledge when conducting process mining. This information has to be available from stakeholders, something that is confirmed in data mining success research as well (Feelders et al., 2000). Additionally, systems knowledge is taken from the process mining success factor domain knowledge as another important sub-construct of resource available. Systems knowledge is defined as the ontological knowledge of the data structure and relationships within the IT system(s). This knowledge needs to be available to construct valid event logs.

The modelling related factors category of Bandara (2007) us renamed to **process mining related factors,** because the process of delivering outcomes differs between process mining and process modelling (for discovery). In section 2.3.2 it was shown that event log construction is an important step in process mining. Also, section 2.3.3 illustrated the different tools and algorithms that are available for conducting a process mining analysis. The following constructs are included under process mining related factors:

- *Process miner expertise*
- *Process mining approach*

*Proces miner expertise* - In the process mining success model, modeller expertise is renamed to process miner expertise, and redefined in the following way: "The experience of the one conducting the process mining in terms of event log construction, and process mining in particular with the aid of certain tools and techniques."

*Process mining approach* - Modelling aids are not considered as separate success factor, since there are not as many different modelling tools available in process mining as there are in process modelling. Process mining mainly uses the same set of tools (ProM framework, or e.g. Futura Reflect which is based on the ProM framework) and standards (MXML or XES). For this reason it is hard to differentiate between different tools, guidelines and such. Instead, this success factor is renamed to process mining approach, and is also placed within the process mining related factors. Already, we saw that an iterative approach is necessary when conducting data mining projects (Fayyad et al., 1996), this can also be true for process mining, since it is a specific form of data mining. Process mining approach is then defined as follows: "The way in which the process miner conducts the process mining project."

A new and separate category of **information systems related factors** is introduced in the process mining success model. It is obvious that process mining relies heavily on the data that originates from information systems in order to reconstruct process models. This is somewhat different from process modelling, since process modelling relies more on (qualitative) information from process stakeholders. The following construct is part of this category:

- *Data & event log quality*

*Data & event log quality* - the success factor of data and event log quality is placed under this new category since it is a product of the information system(s). Data and event log quality influence the subsequent process mining analysis and the model outcomes. This factor would also encompass the previously mentioned data mining success factors data quality, data integration and technical integration (Nemati & Barko, 2003)

Finally, the **moderating factors** of Bandara (2007) are included and adapted in the process mining success model. The following constructs are included under this category:

- *Process complexity*
- *Importance*
- *Type of IT system*

*Process complexity* - The process modelling success factor of process complexity is described by Bandara (2007) as the many different features of the process modelled. It mediates the relationship between resource availability and process modelling success. Bandara (2007): "when process complexity was low the resources made available were low in comparison to when the complexity was higher and resource availability increased as process complexity increased." Processes are the unit of analysis for both process modelling and process mining. It was already found that process characteristics can have an important influence on process mining success as well. Important process complexity characteristics include the structuredness, complexity and rate of change of the process. These features are somewhat more specific than the definition given by Bandara (2007). It is also chosen as moderating factor on the relationship between resource availability and process mining success. Furthermore, it also chosen as a moderating factor between data and event log quality and process mining project success. For example, a more unstructured, complex process will also produce an event log that will be hard to analyze since it will produce "spaghetti-like" models.

*Importance* - Additionally, importance is included as a moderating factor in the process mining success model. Bandara (2007): "when the perceived importance was low, the amount of top management support received was also low and the degree of process modelling success was low too." This is expected to hold for projects in general, and also for process mining projects.

*Type of IT system* - Finally, the type of IT system is chosen as a new and important moderating factor. Already in section 2.2 it was shown that information systems can be classified along the data-centric vs. process-centric and unstructured vs. structured dimensions (van der Aalst, 2007a), or the unframed till framed and person-to person till application-to-application dimensions (van der Aalst, 2009). From this it can be derived that the type of system has an effect on the structure of the process as it is enacted in the system, and thus influences the process characteristics. Also, each system has a certain way in which it logs the data. For instance, WfMs often provide very suitable logs of data for conducting, while this is less evident for SAP systems. So, the type of system determines to a large extent the data and event log characteristics. This is why the type of system is chosen as a moderating factor between data & event log quality and process mining success. Lastly, complexity of the type of IT system seems to influence the need for having systems knowledge. It seems that for complex information systems like SAP, the need for systems knowledge increases (van Giesssel, 2004; Segers, 2007; Ingvaldsen, 2008). For this reason the relationship between systems knowledge and process mining project success is mediated by the type of IT system.

### 4.2.2 A-priori success measures

Since process modelling projects and process mining process share the same objective of deepening organizational knowledge, and deliver similar results, the success measures of the final process modelling success model of Bandara (2007) are adapted for process mining purposes. Model quality and project efficiency would have a similar definition, whereas process impacts would be the effect of process mining on process performance, leading to the following success measures:

- *Model quality* – the extent to which all desirable properties of a model are fulfilled to satisfy the needs of model users in an effective and efficient way
- *Process impacts* – measures the effect of process mining on process performance
- *Project efficiency* – the ratio of obtained outcomes over invested resources

Section 2.5.2 demonstrated that no explicit success evaluation criteria are used in evaluating process mining projects. Some implicitly described success outcomes were found: 'Increased process insights' and 'Model user satisfaction' (usefulness of process model, objective support for organizational measures and understandability of models).
Increased process insights would best fit under model quality (more specifically to the importance or relevance of the results). It has to be noted that it would probably fit even better under the construct 'Individual impact' (Bandara, 2007, p.178), which refers to how process mining has influenced the process stakeholders. However, Bandara (2007) showed that this construct had some overlap with model quality and therefore it was removed for reasons of creating a parsimonious model. The other outcomes could also be included under model quality.
Nemati & Barko (2003) mention several data mining success measures. The 'iron triangle' (cost, time and quality) measure can be integrated under project efficiency. The information system and stakeholder community success measures can be summarized under model quality, and the organization benefits under process impacts.

### 4.2.3 A-priori model

Figure 4.1 below presents the a-priori model that was derived from literature, which can be seen as an extension/adaptation of the process modelling success model of Bandara (2007). Specified definitions of these a-priori constructs, including sub-constructs, can be found in Appendix B.

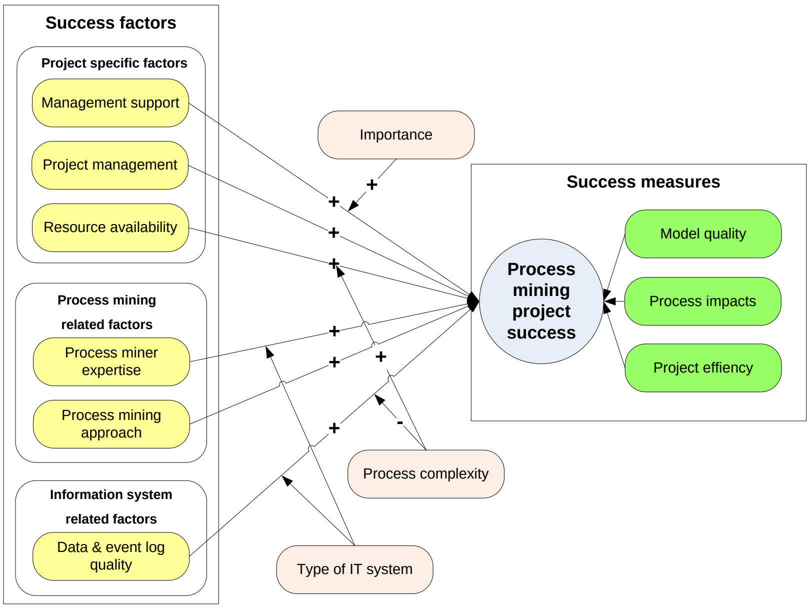


Figure 4.1 - A-priori model of process mining success

### 4.2.4 Research-based design principles

The a-priori model of process mining success should be seen as a first step to understanding how practitioners can successfully conduct process mining projects. It represents a 'static' picture of the high-level interventions (success factors) that are needed to achieve successful outcomes of a process mining project. However, practitioners can benefit even more from detailed, prescriptive knowledge in the form of design principles. The following specific interventions were found for specific project stakeholders when analyzing the previously conducted process mining case studies:

***Principle 1* (project leader)** – To successfully conduct a process mining project (outcome), there needs to be good project management in place (high-level intervention) which can be implemented through involving the right people knowledgeable about the process. This helps to explain the meaning of activities in the process and to determine the appropriate level of granularity of the analysis (mechanisms).

***Principle 2* (process miner)** – To conduct a successful process mining project (outcome), the process miner needs to have sufficient expertise (high-level intervention) by having or obtaining knowledge of the business process under consideration and having knowledge of the data structure and contents of the relevant information systems (specific interventions). The last part is especially true for process mining in SAP R/3 systems (context). This is necessary to be able to construct valid event logs and correctly interpret the results (mechanisms).

***Principle 3 (process miner)*** – To enable the success of a process mining project (outcome), a good process mining approach (high-level intervention) should be adopted, which consists of (specific interventions):

- Discussing results with the client; never give the analysis results without explanation

- Combining mining perspectives to get different views on the process

***Principle 4*** **(project leader/process miner)** – The success of a process mining project (outcome) depends on having data and event logs of sufficient quality (high-level intervention), which is characterized by (specific interventions):

- Having the right amount of data
- Conformance of data to process mining requirements
- The amount of 'noise' in the data
- The representativeness of the data

## 4.3 Chapter conclusion

This chapter sought to develop more knowledge on process mining success, based on process mining case studies and the closely related disciplines of process modelling and data mining. An a-priori model of process mining success was developed which shows the relationship between several hypothesized success factors (management support, project management, resource availability, process miner expertise, process mining approach and data & event log quality), moderating factors (importance, process complexity, type of IT system) and success measures (model quality, process impacts, project efficiency). Four theoretical design propositions have been distilled to realize some of the success factors.

The question now is how the a-priori success model and theoretical design propositions hold in practice. For this reason a multiple case study was conducted. Results of the multiple case study are presented in the next chapter to test if the a-priori model is valid, if all the relevant constructs are captured in the model, and if the design propositions can be adapted/extended with direct insights from practice.

# 5. Practice-based research results

In this chapter, the previously constructed a-priori model is evaluated and possibly respecified through a multiple case study. The results of the multiple case study analysis are presented here. Case study data included interview transcripts, an observation report and documentation. Additionally, practice-based design principles were distilled from the case study data. By contrasting the theoretical results with the results from practice, confidence is increased in the validity and completeness of the final results.

Section 5.1 presents the within-case analysis. Section 5.2 elaborates on the cross-case analysis that was conducted to look for patterns and similarities across the case sites and the respecification of constructs in the a-priori process mining success model. Section 5.3 presents the respecified model. Lastly, section 5.4 complements the respecified model by listing the relevant design principles that were found.

## 5.1 Within-case analysis

Please recall that four different case studies were selected for analysis. This section includes a detailed write-up for each of the case sites:

- ASML

- Company B
- Verbeeten Institute
- T-Mobile

### 5.1.1 ASML

ASML is a leading manufacturer of advanced technology systems for the semiconductor industry. The company offers an integrated portfolio for manufacturing complex integrated circuits (ICs). ASML designs, develops, integrates and services these advanced manufacturing systems, better known as wafer scanners. Wafer scanners use a photographic process to image nanometric circuit patterns onto a silicon wafer. Because of competition and fast innovation in this sector, the time-to-market is very important for ASML (Rozinat et al., 2006b).

The testing of manufactured wafer scanners is an important but also time-consuming process which consists of three phases: (1) the calibration phase, (2) the test phase (the actual testing), and (3) the final qualification phase. The whole process typically takes a few weeks. When finished, the wafer scanner is partly taken apart and shipped to a customer. A part of the calibration and test phase is repeated at the customer site, after re-assembling the wafer scanner. ASML operates in a market where the time-to-market of system enhancements and the time-to-market of new system types is critical (Rozinat et al., 2006b).

The number of manufactured wafer scanners of a single type is typically less than 50. And with each new type, parts of the calibration and test phase are adjusted. On average five different system types are manufactured in parallel. The short time-to-market, the constant innovation, and the high value of wafer scanners make testing very important. On the one hand, spending too much time on testing will result in high inventory costs and lost sales. On the other hand, inadequate tests will result in systems which are malfunctioning (Rozinat et al., 2006b).

For the process mining project at ASML, the aim was to find process bottlenecks and ideas for process improvement for the wafer scanner test process. Also, the executed test sequences were compared to the given reference sequence, i.e., the “ideal process model” defined by ASML (Rozinat et al., 2006b). The project was initiated by a PhD student working at ASML, whose objective it was to look at the test processes and test methods employed at ASML. He thought that process mining could answer some of his questions relating to how much time is spend on actually testing the machines and how much time as idle time in the process. He coordinated most of the ASML related project activities, so he was more or less the project leader. A clear project sponsor within ASML was lacking. The project leader involved several process mining experts from Eindhoven University of Technology (two PhD students and one professor). They were interested in the ASML case study, because it presented the opportunity to test their algorithms on a highly unstructured process, originating from real-life data. One PhD student was mostly involved in doing the process mining, so she can be seen as the process miner.

A choice was made to analyze the specific testing process at the ASML site (not at the customer site) since appropriate logging was available in a straightforward format. Data originated from a single database. Each wafer scanner in the ASML factory produces a log of the software tests which are executed. The logging contains the start and stop moment of each test. Sets of calibration and test actions are grouped into so called job steps. These job steps are executed according to a certain sequence. Raw test log data was converted to MXML event logs by using a purposely created filter for the ProMimport framework. These logs were imported into ProM for further analysis. Because the tests were logged on a very detailed level, control-flow discovery algorithms delivered process models that were not very readable and were not understandable. A mapping of low level test codes to job-steps had to take place. To successfully conduct this mapping, a team of four production engineers assisted in determining which low-level test belonged to which job-step in the test process.

On the one hand, the event log contained few process instances which contained many activities (due to the fact that ASML only produces a limited amount of wafer scanners of a certain type). On the other hand, the event log illustrated that the test process is very flexible, since tests can fail and all kinds of orderings of activities are possible then. That is why the heuristics miner was chosen as the control-flow mining algorithm, since it can deal with unstructured processes. The mined process model was then compared to the reference model (or the ideal model), which showed a very low fit between the two models, showing that the reference model used by ASML was incorrect and more complex than imagined. Additionally, the order of certain activities in the test process was investigated to suggest possible process improvements (Rozinat et al., 2006b).

### 5.1.2 Company B

Company B is a very large German multinational that is active in the energy, healthcare and industry sectors. The organization consists of a very large number of subsidiary companies. Company B uses internal auditing to evaluate the performance of these separate organizational entities with respect to risk management, control and governance. Commonly, internal auditors use reports to investigate possible frauds or other errors, or they make use of test samples to see the whole trajectory of a certain case, as it follows a certain process. At company B, four hundred internal auditors worldwide are supported by a team of twenty IT auditors that deliver data reports to the internal auditors.

The process mining project was conducted within the IT auditing department of Company B. Initially, they were made an offer by a German consulting company to try out process mining (through the use of Futura Reflect). The IT auditing department was very much interested to try out process mining. They already had strong capabilities with conventional data analysis (with SQL server) and wanted to try out new data analysis techniques. The main processes of interest for the analysis were the purchase-to-pay and order-to-cash processes of SAP, an enterprise system which is used by most of Company B's subsidiaries.

Three small projects were conducted with process mining. Together, they can be seen as one larger process mining project. The people that were involved in the process mining project were a data analyst from Company B who conducted a part of the process mining and coordinated the project (more of a combined process miner/project leader role), and a process mining expert from the process mining tool vendor which assisted in creating the event log and provided additional process mining support. A budget for the process mining project was provided by the department manager, which also coordinated a bit of the project (a project sponsor/project leader role). Unfortunately, it was not possible to interview this person.

In the first project, the concrete goal was to try out the new technique and evaluate the software. Since Company B has large volumes of data, they were interested to see if the software could handle these data volumes. Also, they wanted to know what information was available in the SAP system for the creation of the event logs.
The focus of the second project was on segregation of duties (SOD) analysis. This means that one user should not be allowed to perform two critical tasks in the process, like changing the vendor bank details and creating an outgoing payment, because then he could easily enter his own bank details and create payment to his own bank account. Typically there are a few SOD conflicts for each business process. SOD analysis was performed for the purchase-to-pay and order-to-cash processes.

During the third project, Company B was also interested in process discovery and analyzing the process model in more detail, albeit only for the purchase-to-pay process.

The event log was created by the tool vendor in the first and second project. But event logs were different from each other. The first event log contained the complete purchase to pay process, from purchase requisitions to outgoing payments.
The second event log was quite minimal; it was optimized for SOD analysis. It contained a few more steps regarding the creation of vendor master data and vendor posting and vendor payments which were not related to the purchase to pay process. It was not possible to see the whole purchase to pay process in the second event log. It was only adequate for seeing the social network.
For the third project the event log of the internal process miner was used. He merged the first two event logs and corrected some mistakes and included a few more activities. He standardized the whole event log and SQL script, and included much more details of interest when analyzing the results.

### 5.1.3 Verbeeten Institute

The Verbeeten Institute is a specialist hospital delivering clinical care in the field of radiotherapeutic oncology and nuclear healthcare. The treatment trajectories of radiotherapeutic oncology are usually much longer than those of nuclear healthcare treatment.
At the Verbeeten Institute, a lean management philosophy was adopted to reduce unnecessary waiting times between activities in the radiotherapeutic oncology process, more specifically in the treatment preparation phase. Well-known lean management techniques such as value mapping and standardization of processes were already being used for this purpose (Staal, 2010).

The process mining project originated after a new workflow management system (WfMs) was implemented at the Verbeeten institute. After this system was implemented, the organization was interested to see if certain standard work patterns could be distilled from the logged data.
Several people were involved in the process mining project at the Verbeeten Institute. Firstly, a master thesis student from Eindhoven University of Technology conducted the process mining analysis, so he acted as an external process miner. The process manager radiotherapy within Institute Verbeeten acted as a project sponsor and provided the necessary resources for conducting the project.

Both process mining and data mining were used to group patients on the basis of control-flow and performance characteristics. Data from three different data sources, including the WfMs, was consolidated into a Microsoft Access database, and converted with the help of the ProMImport framework to an MXML event log. In the ProM framework, the heuristics miner was used to obtain the control-flow model of the preparation phase from the event log. This model showed a relatively linear process, confirming the designed process model in the WfMs that guides process execution.
To investigate the performance of the preparation phase, an analysis was conducted using the performance analysis with Petri net plug-ins. Because a large number of start times were not registered by the WfM system, it was not possible to properly identify process bottlenecks. This also forced the process miner to use data from other systems.

Aggregation of activities had to take place so that dotted chart analysis and performance sequence diagram analysis could be used to identify prevalent patterns in the dataset. A pattern which 77% of cases confirmed to was found, which confirmed the existence of highly standardized process. The case data extraction plug-in provided a way to organize certain average throughput times for certain types of patients. The case data extraction plug-in was further used to extract data for classification and prediction (data mining), to classify specific classes of patients and predict throughput time of patients.

### 5.1.4 T-Mobile Netherlands

T-Mobile Netherlands B.V. is a provider of mobile communications offering products and services for consumers as well as for business users in Netherlands. It is part of the Deutsche Telekom group, which is the largest telecommunications company in Europe. In the Netherlands, T-Mobile serves more than 4.5 million customers, and their offering includes post-paid and pre-paid subscriptions to (mobile) telephone and internet services. The company also acts as a reseller of mobile (phone) and internet products that are offered in combination with a T-Mobile subscription.

The process mining project at T-Mobile Netherlands was conducted within the Sales Operations & Process management department. Here, the aim was to obtain more insight into the activation process of customer services for existing customers with a new Iphone subscription; referred to by T-Mobile as the Iphone renewal process. This process was selected because there were some indications that customer expectations were not always met in this process.

Three people were mostly involved in the project. The author himself conducted the process mining analysis and coordinated the process mining efforts as a part of this master thesis project. The Manager Process management acted as the responsible of the project, which can be characterized as a project sponsor role. The Director Sales Operations & Process management facilitated the project by providing the necessary resources and evaluating the final results of the project, so she can also be seen as a project sponsor. Additionally, a process integration manager provided all the relevant business domain knowledge, since she was very knowledgeable about the provisioning process. A senior IT developer assisted with obtaining all the necessary knowledge about the relevant information systems.

The project was scoped to the part of the process was logged by the Oracle database containing information about customer contracts, orders and order lines. This database acted as a repository between the different front-end and back-end systems. Initially, data was provided by an IT developer, however, at a certain point in time it was decided that the process miner would query a copy of this database himself to speed up the analysis. Something that posed a challenge was the filtering of consumer customers from the original dataset, since a coupling had to be made with a different IT system. Overall, obtaining the relevant data seemed to be quite difficult in this project.

For converting raw data to an event log, Microsoft Access and the ProMImport framework were used. Several different event logs were made out of the different initial datasets. Once the event logs were constructed, the process mining analyses went quite fast. There was some noise in the data/event log, since it originated from a test environment. Therefore, the heuristics miner was used because it can deal well with noise. As it turned out, the process model was quite structured and comprehensive. It showed that the process was a relatively straightforward and linear one, which was expected since the process is mostly automated. This is also why a social network analysis was less relevant. The obtained models were then converted to a Petri net, so a performance analysis with Petri net could be conducted. This performance model was insightful since it showed potential bottleneck activities in the process on which to focus further analysis. It showed that some bottlenecks existed. One bottleneck that was specifically interesting occurred at the beginning of the process.
In addition to the process mining analysis, an analysis was made of the customer interactions regarding the process instances with a longer than usual throughput time. This verified the impact of the process bottlenecks on the customer, and increased the importance of the results.

## 5.2 Cross case analysis & respecification

After the within-case analysis, a cross-case search for patterns was conducted to look for emergent patterns across the different case studies (as described in section 3.4). A combination display of group similarities and group differences across different respondents was constructed by exporting the results of an Nvivo matrix coding query to Excel. In this matrix, the rows represent the transcripts of the different respondents. The respondents are grouped per case site. The columns represent the a-priori and newly found success factors, moderating factors and success measures. These columns are further distinguished in the general mentioning of the constructs, and the specific mentioning of the importance or unimportance of the construct (where relevant). The cell contents in the matrix refer to the number of coding references, which is the number of passages in the transcripts that were coded under each of the constructs.

The number of coding references was summarized for each of the case sites and for each of the constructs. Table 5.1 and 5.2 show the cross-case analysis results. Table 5.1 includes the initial success factors of the a-priori model (F1-F6) and the newly found success factors (F7-F9). The importance or unimportance of these constructs was included as well, where mentioned. Table 5.2 includes the moderating factors (M1-M3), the initial success measures (S1-S3) and the newly found success measures (S4-S5).

To test for redundancy, relatedness and possible moderating effects of constructs, a different matrix intersection search was conducted Nvivo which served to identify passages that were coded under multiple constructs. If a passage of text is coded under multiple constructs, it might indicate possible overlap or a moderating relationship. These are shown in table 5.3 as counts that are not within the diagonal. The Nvivo tool allows checking these specific instances to let the qualitative researcher judge what is actually happening. Results of this search are displayed in table 5.3.

| | Success factors | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A-priori | | | | | | | | | | | | | | New | | | | |
| | **F1. Management support** | F1. Management support (important) | F1. Management support (not important) | **F2. Project management** | F2. Project management (important) | **F3. Resource availability** | F3. Resource availability (important) | **F4. Process miner expertise** | F4. Process miner expertise (important) | **F5. Process mining approach** | F5. Process mining approach (important) | **F6. Data & event log quality** | F6. Data & event log quality (important) | F6. Data & event log quality (not important) | **F7 - Data privacy & authorization** | F7 - Data privacy & authorization (important) | **F8 - Personal commitment** | **F9 - Personal skills** | F9 - Personal Skills (important) |
| **ASML** | | | | | | | | | | | | | | | | | | | |
| 1. ASML - Project leader | 2 | 2 | 0 | 1 | 0 | 2 | 1 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2. ASML - External process miner | 3 | 3 | 0 | 1 | 1 | 4 | 3 | 4 | 4 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Total case site* | 5 | 5 | 0 | 2 | 1 | 6 | 4 | 6 | 6 | 1 | 1 | 3 | 2 | 0 | 0 | 0 | 1 | 0 | *0* |
| **Company B** | | | | | | | | | | | | | | | | | | | |
| 3. Company B - Internal process miner | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 3 | 5 | 1 | 1 | 4 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| 4. Company B - Two external process miners | 2 | 2 | 0 | 4 | 2 | 0 | 0 | 3 | 3 | 3 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Total case site* | *3* | *2* | *1* | *5* | *2* | *1* | *1* | *6* | *8* | *4* | *3* | *6* | *3* | *1* | *1* | *1* | *0* | *0* | *0* |
| **T-Mobile** | | | | | | | | | | | | | | | | | | | |
| 5. T-Mobile - Project sponsor 1 | 1 | 1 | 0 | 6 | 1 | 3 | 2 | 5 | 4 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 6. T-Mobile - Project sponsor 2 | 1 | 1 | 0 | 4 | 1 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 7. T-Mobile - External process miner | 3 | 2 | 0 | 3 | 3 | 4 | 4 | 5 | 5 | 5 | 5 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Total case site* | *5* | *4* | *0* | *13* | *5* | *7* | *6* | *12* | *11* | *5* | *5* | *3* | *3* | *0* | *1* | *0* | *0* | *2* | *2* |
| **Verbeeten Institute** | | | | | | | | | | | | | | | | | | | |
| 8. Verbeeten institute - Project sponsor | 1 | 0 | 0 | 1 | 1 | 2 | 2 | 5 | 1 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9. Verbeeten Institute - External process miner | 1 | 1 | 0 | 2 | 1 | 2 | 3 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Total case site* | *2* | *1* | *0* | *3* | *2* | *4* | *5* | *5* | *1* | *2* | *1* | *4* | *3* | *0* | *1* | *0* | *0* | *0* | *0* |
| **Total of all case sites** | **15** | **12** | **1** | **23** | **10** | **18** | **16** | **29** | **26** | **12** | **10** | **16** | **11** | **1** | **3** | **1** | **1** | **2** | **2** |

Table 5.5 - Matrix cross-case display for success factors

| | Moderating factors | | | | | | | Success Measures | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | A-priori | | | | | | | | | New | | |
| | **M1. Importance** | M1. Importance (important) | **M2. Type of IT system(s)** | M2. Type of IT system(s) involved (important) | **M3. Process complexity** | M3. Process complexity (important) | M3. Process complexity (not important) | **S1. Model quality** | S1. Model quality (important) | S1. Model quality (not important) | **S2. Process impacts** | S2. Process impacts (important) | S2. Process impacts (not important) | **S3. Project efficiency** | S3. Project efficiency (important) | S3. Project efficiency (not important) | **S4. Recognizability of results** | S4. Recognizability of results (important) | **S5. Translating results to customer experience** |
| **ASML** | | | | | | | | | | | | | | | | | | | |
| 1. ASML - Project leader | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | 4 | 2 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 2. ASML - External process miner | 1 | 0 | 1 | 0 | 2 | 1 | 0 | 2 | 1 | 1 | 4 | 3 | 0 | 2 | 2 | 0 | 1 | 1 | 0 |
| *Total case site* | *1* | *0* | *1* | *0* | *4* | *3* | *0* | *3* | *2* | *1* | *8* | *5* | *0* | *3* | *3* | *0* | *2* | *2* | *0* |
| **Company B** | | | | | | | | | | | | | | | | | | | |
| 3. Company B - Internal process miner | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 5 | 4 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 4. Company B - Two external process miners | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 2 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| *Total case site* | *3* | *3* | *4* | *1* | *0* | *0* | *0* | *7* | *5* | *2* | *4* | *4* | *1* | *2* | *2* | *0* | *2* | *2* | *0* |
| **T-Mobile** | | | | | | | | | | | | | | | | | | | |
| 5. T-Mobile - Project sponsor 1 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 6. T-Mobile - Project sponsor 2 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| 7. T-Mobile - External process miner | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Total case site* | *5* | *5* | *3* | *2* | *1* | *1* | *0* | *3* | *2* | *0* | *6* | *6* | *0* | *2* | *2* | *0* | *1* | *1* | *4* |
| **Verbeeten Institute** | | | | | | | | | | | | | | | | | | | |
| 8. Verbeeten institute - Project sponsor | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 2 | 1 | 0 | 2 | 1 | 1 | 1 | 1 | 0 |
| 9. Verbeeten Institute - External process miner | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 4 | 2 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| *Total case site* | *2* | *2* | *2* | *1* | *2* | *2* | *0* | *5* | *3* | *0* | *3* | *1* | *1* | *3* | *1* | *1* | *2* | *2* | *0* |
| **Total of all case sites** | **11** | **10** | **10** | **4** | **7** | **6** | **0** | **18** | **12** | **3** | **21** | **16** | **2** | **10** | **8** | **1** | **7** | **7** | **4** |

**Table 5.6 - Matrix cross-case display for moderating factors and success measures**

| | F1. Management support | F2. Project management | F3. Resource availability | F4. Process miner expertise | F5. Process mining approach | F6. Data & event log quality | F7 - Data privacy & authorization | F8 - Personal commitment | F9 - Personal skills | M1. Importance | M2. Type of IT system(s) | M3. Process complexity | S1. Model quality | S2. Process impacts | S3. Project efficiency | S4. Recognizability of results | S5. Translating results to customer experience |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **F1. Management support** | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 5 | 0 | 0 | 0 |
| **F2. Project management** | 1 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **F3. Resource availability** | 0 | 0 | 18 | 3 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **F4. Process miner expertise** | 0 | 0 | 3 | 29 | 3 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 |
| **F5. Process mining approach** | 0 | 0 | 1 | 3 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| **F6. Data & event log quality** | 0 | 0 | 2 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| **F7 - Data privacy & authorization** | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **F8 - Personal commitment** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **F9 - Personal skills** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **M1. Importance** | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **M2. Type of IT system(s)** | 0 | 0 | 1 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| **M3. Process complexity** | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 1 | 0 | 0 | 0 |
| **S1. Model quality** | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 1 | 0 | 0 | 0 |
| **S2. Process impacts** | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 21 | 0 | 0 | 0 |
| **S3. Project efficiency** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 |
| **S4. Recognizability of results** | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 |
| **S5. Translating results to customer experience** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |

Table 5.7 - Matrix intersection search results

### 5.2.1 Re-specification of success factors

Table 5.1 shows that **Management support (F1)** was repeatedly cited across respondents and case sites, both in general (15 citations) and as an important success factor (12 citations). In contrast, 1 citation specifically mentioned the unimportance of the construct:

"Question: Was it important to get management support for this project?
Answer: Not in this case. The application of process mining in internal audit is different from the application for business process improvement. I think this is audit specific, at least for internal audit. It's not as important as for other departments." – *Internal process miner, Company B case site*

Still, the respondent mentioned that there was some involvement from senior management:

"Question: Were there also other people from your company involved in the project?
Answer: My boss was responsible for requesting the budget for the project and coordinated the project a little bit. He communicated between the internal auditors and our team, and he managed the work. But he was not involved in the project, outside of managing the tasks. " – *Internal process miner, Company B case site*

Although management support seems to be less important for internal auditing than for business process improvement, it seems that management support is still crucial to obtain the necessary resources. Another mentioned sub-constructs of management support included informing stakeholders on the purpose of the project. What became evident through the matrix intersection search (table 4.3) is that management support seemed to have a direct effect on process impact, as evidenced by this citation:

"Question: How important was it to get top management support for this project?
Answer: Support is important to be able to have impact. Without support nothing will be done with the results. But also perhaps to get the buy-in from people so they work with you and provide you with domain knowledge." – *External process miner, ASML case site*

Due to the number of citations mentioning the importance of management support, it is kept as a success factor, and existing sub-constructs are extended with the aforementioned sub-constructs.

**Project management (F2)** was cited 23 times in general, and was specifically mentioned as important in 10 citations across case sites and respondents. There were no citations stating that project management was not important. An example citation stressing the importance of the construct was this:

"Question: Was project management important for the success of this project?
Answer: I think so. In a project like this it is easy to continue with creating improved analyses. At a certain point you have to determine that 80% is also good enough. Project management is important to know when to stop. The question is what you want to achieve and to know when this has been achieved." – *Project sponsor, Verbeeten Institute case site*

Citations for project management did not overlap with any other constructs, as is evidenced by table 5.3. A lot of different sub-constructs were mentioned for project management, including: communication management, data management, definition of objectives and scope, people management. Due to the ample support for this construct, it is included in the respecified success model.

The next construct, **Resource availability (F3)**, was mentioned in a total of 18 citations across case sties and respondents, of which 16 stressed the importance of the construct. No citations were found that mentioned the unimportance of the resource availability.

At the ASML case site, the available domain knowledge from the test engineers was vital for success:

"Question: Did you have access to all the necessary resources to make this project a success?
Answer: Yeah, eventually we did. We did not claim anyone. Most of it happened through lobbying. We did talk a while to be able to involve the four test engineers in the project. That was not formally arranged. If we would not have had those four people, we would not have gotten any results." – *Project leader, ASML case site*

Sub-constructs that were mentioned include: the availability of people knowledgeable about the process and the system(s) involved in the process, available process documentation, available data models and data availability. From the large number of citations, it can be concluded that resource availability should be kept as a success factor.

**Process miner expertise (F4)** was the most cited construct across case sites and respondents, with a total of 29 citations, 26 citations mentioning its importance, and no citations mentioning its irrelevance. A lot of different sub-constructs were mentioned under the heading of process mining expertise, including:

- tool & algorithm expertise:

"Question: Were there things that facilitated the process mining project?
Answer: Yes, the fact that the process miner already had the process mining tools up and running and knew how they worked. In my opinion the process miner already had a detailed knowledge of process mining, enabling us to convert the data to an analysis quite quickly." – *Project sponsor 1, T-Mobile case site*

- business (process) knowledge

"You always need a basic understanding what these data are and what kind of process they are coming from, what it means, in order to be able to interpret the results and to see whether you are doing the right thing. There is judgement coming in, for example when you have to decide how to filter the process and which parts to filter on. You really need to have someone like Ivo to give you feedback in an iterative way or you need to have this domain knowledge yourself." – *External process miner, ASML case site*

- systems knowledge & data analysis expertise

"Question: Did you encounter any issues or concerns in the application of process mining?
Answer: The first issue is maybe that creating the event log is very difficult. You need a deep knowledge of the source system when you are analyzing an SAP system." – *Internal process miner, Company B case site*

"Question: What kind of skills really help in process mining?
Answer: The most critical part is creating the event log. You must understand the data structure of the SAP system. This is the most crucial point here.
You already know the right tables when you are developing data analysis. It was not difficult for me to identify the right tables. You really have to look into the data to find the right way of linking the

day. This is less difficult when you are already familiar with the SAP system." – *Internal process miner, Company B case site*

The need for systems knowledge was mentioned mostly at the Company B case site in relation to the SAP system. Another important point to note here is that there is overlap occurring between process mining expertise and resource availability. Since process mining expertise is such a widely cited construct, it is included in the respecified success model.

**Process mining approach (F5)** was specifically mentioned by process miners across the different case studies. In total there were 12 citations mentioning process mining approach of which 10 mentioned the importance of the construct and there were no citations stressing the unimportance of the construct. As expected, event log construction turned out to be an important step in conducting the process mining efforts. Also the iterative way of constructing and analyzing event logs was stressed:

"It took me quite a while to construct an event log out of the raw data, since I did not have any previous experience with constructing event logs. For the conversion of raw data to an event log, I used Microsoft Access and the ProMImport framework.
I noticed that event log creation was a very iterative process: after constructing an event log, I would analyze it in ProM only to find that it could be improved or adapted, and that I needed to query the raw in a different way. Having 'direct' access to the data significantly speeded up the iterative process of creating and analyzing event logs. " – *Observation report of the author (process miner), T-Mobile case site*

"Question: What issues did you encounter while conducting the process mining project?
Answer: Every dataset is different. The way in which I queried the dataset was very important for the results. One time I had sorted the dataset wrongly, after which I got really strange results. You always have to look at yourself: did I create the dataset in the right way and did I construct the appropriate joins between tables? I created a lot, really a lot of different datasets, which were time and time again an improvement upon the previous one. As a process miner, you are constructing dataset in a very iterative manner." – *External process miner, Verbeeten Institute case site*

Although iterative cycles seem important, it was also stressed that this is dependent on the process miner expertise:

"Question: How important is an iterative approach in process mining?
Answer: I think it is important. It depends on how much domain knowledge the analyst has."
– *External process miner, Company B case site*

Another important point to note is that in the Company B, T-Mobile and Verbeeten Institute case studies SQL was used to extract data from the source system. It is specifically mentioned as a laborious process:

"What we did up until now is to manually construct SQL queries. In the first project I created an SQL script of 600 lines, which were 30 or 40 queries. Company B extended this, and the script has doubled in size in the meantime, and it only involves the purchase-to-pay process. You do not want to have to adjust all these lines for a different SAP installation. That is why a graduate student has developed a tool for us to do this more easily." – *External process miner, Company B case site*

"What I also found was that it was quite laborious to create and adapt all the SQL queries myself, both for extracting the data from the test database and converting it into the right format in the

Access database. Perhaps I could have used a tool like XESame that generates these SQL queries automatically." *– Observation report of the author (process miner), T-Mobile case site*

Apart from the way of the (iterative) approach of constructing and analyzing the event logs, other mentioned sub-constructs included the choice of appropriate algorithms and combining different mining perspectives. Process mining approach is kept as a success factor in the model since it supported by a large number of citations stressing the importance of the construct.

**Data & event log quality (F6)** received 16 general citations, with 11 citations stressing its importance:

"Question: How would you characterize the quality of the raw data that you obtained at ASML?
Answer: The quality was very good. It was just not complete. So we had a number of actions that were not in the log and it was a problem because one of the questions we had was related to idle time: when was the machine not used, when test process time was wasted. Because of that we could not conclude that every idle time that we observed was indeed idle time. In that sense it was flawed. Also the mapping was quite ambiguous; it made it difficult to get to the higher abstraction level. We had to make some pragmatic choices. So yeah the quality of the log is also important, not just how easy it is to get, but also how complete or reliable is it." *– External process miner, ASML case site*

Although there were many citations for the importance of the construct, there was one citation stressing the irrelevancy of data quality for internal auditing:

"Question: How would you characterize the quality of the data and did it influence project success?
Answer: The quality of the data is not important. In internal audit we are interested in reality, we don't have to pre-process the data, we want to see 100% of the data. We used the original data. For business process improvement this might be different. So data quality is not an issue for us." - *Internal process miner, Company B case site*

This might suggest that data & event log quality is less important for internal auditing than for business process improvement purposes, or it could be the case that the respondent did not understand the meaning of the question properly. Still, it has to be said that pre-processing always has to take place before event logs can be constructed and analyzed: certain minimum data requirement have to be met. Also, there always needs to sufficient data.
A more plausible explanation could be that certain dimensions of data & event log quality are less important for internal auditing than for process improvement purposes.
Different sub-constructs mentioned for data & event log quality included: sufficiency, level of detail, accuracy, completeness and reliability of the data/event log. The dimensions of data accuracy or reliability might be less important for internal auditing. Nevertheless, data & event log quality is included in the respecified model.

**Data privacy & authorization (F7)** was a new factor that was mentioned 3 times, but only 1 time as crucial or important:

"The problem of process mining is maybe that we have an audit approach, which is mandatory; it's difficult to modify the audit approach. Because Company B has a lot of incorporations, you always need authorization for getting the data, but you only get the data when specific preconditions are met. It's difficult to use process mining when planning an audit." *– Internal process miner, Company B case site*

This construct seems to only hold for internal auditing in combination with a specific data governance policy, therefore, it is not included in the respecified model.

**Personal commitment (F8)** was identified by one citation stressing its importance:

"Question: What lessons have you taken from this process mining project?
Answer: I have noticed, and this is not only with process mining, but also with the things we have created ourselves that you can only get it a step further when you commit to it personally until it has been implemented in the organization. Within an operational organization like this, you have to be convinced that it helps. Although I am convinced of process mining, to really embed it, you have to invest effort personally for a longer period. That's not always possible." *– Project leader, ASML case site*

Also, since this construct was not mentioned in the other case studies, it is not included in the respecified model.

The new **Personal skills (F9)** success factor was mentioned a total of 2 times, but solely in the T-Mobile case study.

"Question: Would you perhaps like to add a success measure we did not discuss yet?
Answer: An important factor is that you have to be communicative as a process miner. You need have some consultancy skills regarding accessibility and being able to provide an explanation to people. I think the process miner had this, which contributed to the result of this project." *– Project sponsor 2, T-Mobile case site*

Mentioned sub-constructs of personal skills include communication skills, accessibility and independence. However, since there are only 2 citations for this construct within a single case study, it is not included in the respecified model.

### 5.2.2. Re-specification of moderating factors

**Importance (M1)** had 11 general citations across case sites, of which 10 stated its importance. No citations mentioned its irrelevance.

"Our audit department is very interested in trying out new data analysis techniques, since we want to enhance our abilities of data analysis. We were very interested in trying process mining out. The interest in doing this is very important." *– Internal process miner, Company B case site*

The most important sub-construct of importance that was found was the need for transparency:

"Transparency is very important for me, and in my case I want to have these results, but I can imagine that some people do not have this need. Not every manager benefits from increased transparency." *– Project sponsor 2, T-Mobile case site*

Some support was also found for the moderating effect of importance of the relationship between management support and process mining success.

"Question: How important was it in this project to obtain management support?
Answer: In our organization we have been paying attention to standardization and obtaining process insights for years now. It is embedded within the organization. Support for this is available within

every level of the organization. So it wasn't really an issue for us." – *Project sponsor, Verbeeten Institute case site*

The way in which the above citation is interpreted is that because of the importance of process improvement and standardization within this organization, the project automatically received a high degree of management support, so there was no need to actively seek additional support.

"It became clear (also during later talks with other people) that the provisioning process was quite important for T-Mobile, as they received indications that the performance of this process had an important impact on the customer. I do believe that it is because of this importance that such senior staff (e.g.  a director and manager) were involved in this project." – *Observation report of the author (process miner), T-Mobile case site*

Since there is ample support for this construct, and some support for the moderating effect of importance, it is included in the respecified model.

**Type of IT system(s) (M2)** was mentioned 10 times in total across the different case sites, and 4 times it was specifically mentioned as important.
As expected, it was confirmed that applying process mining on SAP data is a challenging task. By examining all citations, it became clear that certain system related issues affect process mining success indirectly:

"Question: Did you encounter any issues or concerns in the application of process mining?
Answer: The first issue is maybe that creating the event log is very difficult. You need a deep knowledge of the source system when you are analyzing SAP system. The second issue that you will only see these activities which you explicitly include in the event log. It is not possible download an event log out of the system which contains all tasks which are possible. You must know which activities are available in the process and you must add them to the event log. So this is a drawback." – *Internal process miner, Company B case site*

"In SAP you have change tables, it is important to know if you can use the tables: you can configure SAP in a way that it records all changes. This is very useful information for process mining, but it has to be available. Not everyone has put this logging functionality into place." – *External process miner, Company B case site*

The type of IT system(s) construct showed overlap with the process miner expertise and data & event log quality constructs. This indicated some support for the moderating effect of type of IT system on the relationship between process miner expertise and process mining project success.
As a process miner, having knowledge of the source system was mentioned as crucial for the SAP system, but less so for the other systems. This could be attributed to the inherent complexity of its data structure and the logging functionality of SAP.
Also the overlap with data & event log quality somewhat supported the moderating effect of type of IT system on the relationship between data & event log quality and process mining project success. For this reason, type of IT system was kept as construct in the respecified model.

**Process complexity (M3)** was found in 7 general citations, 6 of which underlined the importance of the construct.

"Question: How does the structuredness of the process influence the analysis?

Answer: What makes it difficult is on the one hand the sheer size, so for example if you look at the detailed test codes and look at the process at that level, then you have a model containing all of these tests. It's just enormously big and unreadable. If you then include all the connections, because this process was not following a linear line, but instead there were a lot of different possibilities. The order of activities is very flexible and people can do things in almost any order then you quickly arrive at a very spaghetti-like process if you show all of these possibilities that have been observed. Simplification is really necessary to be able to read something. If you can't see anything, then it's also useless. This really correlates to the usefulness of the results that you get." – *External process miner, ASML case site*

More support was found in the ASML case study for the moderating effect of process complexity on the relationship between data & event log quality and process mining project success. No support was found for process complexity influencing the relationship between resource availability and process mining project success. Therefore, process complexity is included in the respecified model only with a moderating effect on the relationship between data & event log quality and process mining project success. The link between process complexity and resource availability is removed from the respecified model.

### 5.2.3. Re-specification success measures

**Model quality (S1)** received mixed citations. The construct had 18 citations in total, 12 citations stating its importance, and 3 citations regarding its unimportance.
It was noticed that the term model quality was often interpreted in the interviews as either the technical 'fit' of the presented models (how well the model fitted the event log), or the understandability of the models, although the a-priori definition of the construct entails more dimensions.

"Question: Would model quality have been a good success measure for this project?
Answer: At first, I would say no. If it answers your question, it is already a success. That doesn't have to be the most perfect mined model or picture." – *External process miner, Company B case site*

"Question: Do you think that model quality would be a good measure of process mining success?
Answer: I would say no, because you are interested in the AS-IS process model. You cannot say process mining is good when it delivers a structured model. Process mining just displays the actual model." –*Internal process miner, Company B case site*

"Question: Do you think that model quality is an important success measure to consider?
Answer: I don't think it's a success measure. It should be a necessary step, good behaviour or a good work method. It's a more a problem in the method, it should be in the way the process mining expert works. It is part of the methodology." – *External process miner, ASML case site*

On the other hand, the sub-constructs of relevance (importance, relevance and completeness) and conformance to user requirements were deemed important.

"Question: Which criteria have you used to evaluate the success of this project?
Answer: I think I evaluate it myself on the axis of completeness: how thorough have you been in your analysis within the scope we have chosen." – *Project sponsor 2, T-Mobile case site*

"We have found interesting things, but the problem was that we could have found them with conventional data analysis techniques as well. Also the unexpected results were not really critical. If you are already using other analysis techniques like us, then the added value of process mining is lower. If you don't use data analysis (much), then you may be fine with process mining". – *Internal process miner, company B*

In conclusion, model quality is kept as a construct in the respecified model. It became clear that understandability and 'technical accuracy' or 'fit' of the models should not be included as sub-constructs of model quality. Rather, information relevance (new process insights) and conformance to user requirements are the important sub-constructs.

**Process impacts (S2)** was referenced 21 times in total, 16 times as important and 2 times as unimportant.

"Question: Do you think that evaluating the impact of the project on the process is a good success measure?
Answer: I don't think this is part of process mining. It has to be seen separately. I view process mining more as business intelligence. The way in which the management copes with that is something separate. It is a different project step for change. I see process mining more as generating information. The decision to do something with it is something different." – *External process miner, Verbeeten Institute case site*

This made the researcher doubt if it was valid to use process impacts as a success measures. It might not always be right to measure success by realized improvements due to its dependency on other factors, such as implementation. It can only be done if results can be clearly attributed to the process mining project. Through further analysis it was found that respondents do evaluate process mining project on the basis of concrete performance improvement suggestions.

"Question: Would process impact be a success measure to consider?
Answer: Yeah, I think so, that would be the ideal case. If the results of the process mining just confirm everything, than, usually there are surprises. The ideal scenario is that you find improvement ideas and that you can implement these changes that have an impact on the goals of the project." – *External process miner, ASML case site*

"Question: Is the way in which the process mining results influence the process under consideration an appropriate measure of success?
Answer: I would not look at the way in which the process mining results influence the process. I would say that the results of the project are related to the degree in which you can specify operational KPI performance improvement suggestions according to your analysis. Only then it will be actionable. For me, the added value lies in the operational activities that can be undertaken, and the effect these will have on the KPI." – *Project sponsor 2, T-Mobile case site*

The suggestion of process improvements as a sub-construct of process impact was made more often. Effectively, process impacts would not only entail the actual realisation of process improvements but also concrete suggestions on how to improve the process. Something that also became clear is that process impact is less relevant to internal auditing:

"Question: Have the process mining results been used in some way within the German multinational?
Answer: Internal auditing is a bit different from process improvement. Internal audit can only determine the problem, but they cannot improve the process. They can't offer themselves as consultant saying that they have discovered the problem and know a solution for it. This is to keep

their independence; else they might have to audit the process that they themselves have improved. “ *– External process miner, Company B case site*

Since the success measure of process impacts has received a lot of support, especially by taking into account the number of references mentioning the suggestion of possible process improvement, it is kept in the respecified success model.

**Project efficiency (S3)** was mentioned 10 times in general, 8 times as important and 1 time as unimportant.

“Question: Would project efficiency have been a good measure of success for this project?
Answer: That depends what you are interested in as organization. ASML is more focussed on time, everything has to go as quickly as possible. Time-to-market is crucial; costs for instance are subordinate.” *–Project leader, ASML case site*

Apart from throughput time, return on investment and the achievement of objectives were mentioned. Although one citation stressed it’s unimportance (related to costs of the project), this was mainly due to the fact that it was a master thesis project. The respondent indicated an efficiency evaluation would have been more important for a commercial project.
Therefore, project efficiency is included as a success measure in the respecified model.

**Recognisability of results (S4)** received 7 citations across all sites, which all stressed the importance of the construct. Something that was interesting was that the construct was first mentioned after a question about model quality:

“Question: Is model quality a relevant success measure for you?
Answer: It is important that the produced results are acceptable, that you can indeed acquire new insights. When the level of detail is too low it is obvious. Acceptable means that the results have to be recognizable and realistic for people. Additionally, I would want the process miner to indicate how he obtained the results.” *– Project sponsor, Verbeeten Institute case site*

“Question: Would the recognisability of results be a good success measure?
Answer: I think that’s a good one, although the assumption is that the analysis is correct. How the process miner came to his conclusion has to be good. How I solved this myself, is that I cross-checked the process mining results against the customer data in the CRM system. If the one thing strengthens another, then the confidence in the results increases. Recognisability in practice is important, and traceability of results is necessary.” *– Project sponsor 1, T-Mobile case site*

From these and other citations it was concluded that recognisability of results has significant overlap with model quality. Therefore, it is included a sub-construct within model quality. It could be seen as more of an external quality measure (as opposed to the internal ‘fit’ of the model).

**Translating results to customer experience (S5)** was mentioned 4 times in general and as an important measure in the T-Mobile case study only.

“Question: What measures would you use to evaluate the process mining project?
Answer: You have to be able to present the advice in regard to the drivers of the specific organization you are working in. I thought that was a good success measure that the process miner did that. That the process miner translated the results towards customer impact (how is it affecting the customer).” *– Project sponsor 2, T-Mobile case site*

At first, it was though that this could be a separate success measure. However, through additional analysis and the fact that it was only supported at the T-Mobile case site, it was found that translating results to customer experience would be better included as a sub-construct of model quality.

## 5.3 Re-specified process mining success model

Figure 5.1 below presents the initial a-priori model.

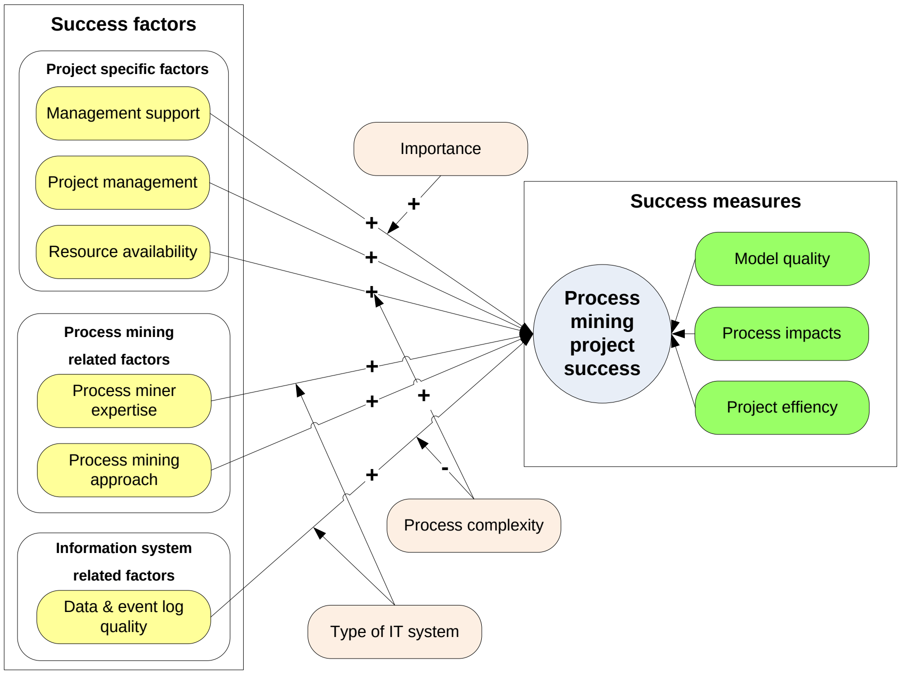


Figure 5.2 below presents the respecified process mining success model.

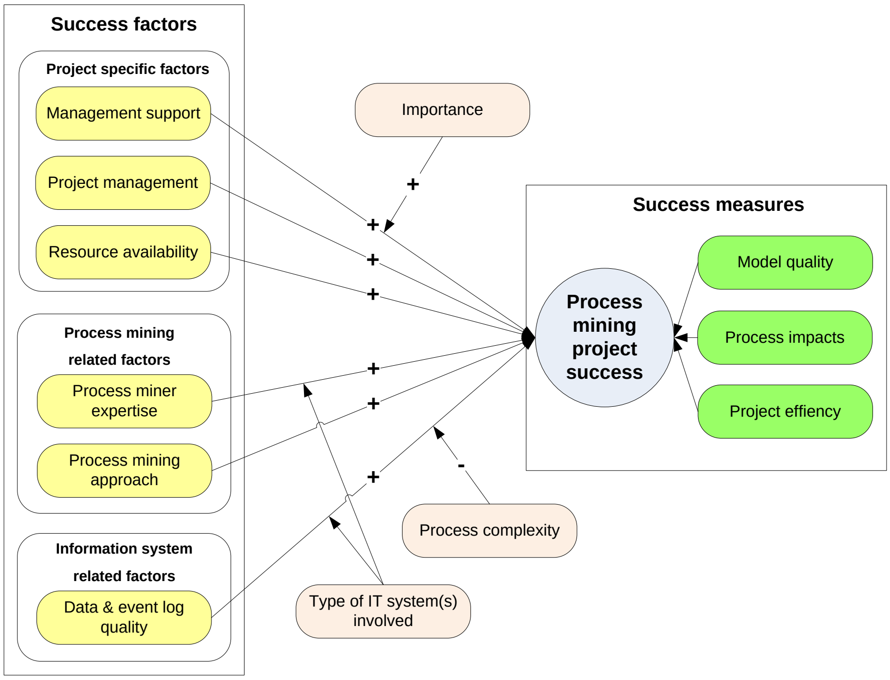


**Figure 5.2 - Respecified Process Mining Success Model**

As can be seen from both figures, there is little difference between the a-priori model and the respecified model. The main difference is that the moderating effect of process complexity on the relationship between resource availability and process mining success was deleted. Apart from that it can be seen that all success factors in the a-priori model were supported in the respecified model.

It has to be noted that management support was less relevant for the purpose of internal auditing than for process improvement purposed. Furthermore, some support was found for the following moderating effects: importance on the relationship between management support and process mining success, type of IT system on the relationship between process miner expertise and process mining success, type of IT system on the relationship between data & event log quality and process mining success and process complexity on the relationship between data & event log quality and process mining success. The success measure of process impacts was not supported for purposes of internal auditing. Lastly, different sub-constructs for each of the constructs were defined. Appendix D provides definitions and sub-constructs for each of the constructs depicted in figure 5.2.

## 5.4 Practice-based design principles

The respecified model can be seen as a 'static' picture of what is needed to achieve process mining success. However, it does not enable practitioners to act on this information. What is needed is prescriptive knowledge in the form of design propositions. The efforts in creating a respecified model can actually be seen as an initial step to creating practice-based design propositions. Both are related in the following way:

- **(C)ontext** – is made up of the process mining project (unit of analysis) and the moderating/contextual variables within the a-priori model
- **(I)ntervention** – are the success factors in the a-priori model. They can be seen as high-level interventions of the design principle
- **(O)utcome** – are the success measures in the a-priori model

As such, the respecified model already contains the basic ingredients for developing theoretical design propositions. High-level information was already coded under nodes in Nvivo. However, these design propositions needed to be more specific. Therefore, for each of the success factors in the a-priori model, text searches were conducted in the qualitative data analysis tool Nvivo to see if passages that mentioned specific interventions to realize these success factors could be identified. If this was the case, a separate node (sub-construct) was created for this specific intervention. These specific interventions were then consolidated in one design principle for each of the success factors/interventions in the respecified model.

*Principle 1 (project leader)* – For the process mining project to have an impact on the process (outcome), it is important to obtain management support (intervention), this is even more important in a situation where the importance of the project is consider low as opposed high (context). Management support works through providing access to resources and informing process stakeholders on the purpose of the project (mechanisms)

*Principle 2 (project leader)* – For the process mining project to be successful (outcome), it is important to have good project management in place (high-level intervention), through (specific interventions):

- Determining clear objectives and scope of analysis of the project beforehand
- Deciding which important process stakeholders to involve in the project and making clear arrangements as to what is expected from them. This is done beforehand
- Having frequent discussions of findings with a business domain expert and a data analysis/information system expert to validate the findings
- Planning an intermediate discussion of results with the project sponsor

*Principle 3 (project leader)* – To successfully conduct a process mining project (outcome), it is important to have sufficient availability of resources (high-level intervention), which is further specified as follows (specific interventions):

- Having the data available for analysis
- Having people knowledgeable about the process available
- Having people knowledgeable about the system (data) available
- Having (relational) data models and documentation of the process available to use

*Principle 4 (process miner)* – For a successful process mining project (outcome), it is important for the process miner to develop a certain degree of process mining expertise (high-level intervention) through (specific interventions):

- Obtaining knowledge of the business process
- Obtaining knowledge of the source system and its data structure (especially important for the context of ERP systems)
- Having general data analysis skills (SQL)
- Knowing how to construct an event log from source data
- Knowing which process mining tools and algorithms to use and how to use them
- Being able to correctly interpret the process mining results

*Principle 5 (process miner)* – To conduct a successful process mining project (outcome), process miners should take care to adopt an appropriate process mining approach (high-level intervention), through (specific interventions):

- Choosing the best approach for obtaining the data, often this involves directly accessing a copy or extract of the production data
- Use the appropriate tools for event log conversion
- Use the appropriate process mining algorithms to deal with real-life data (e.g. noise)
- Combining mining perspectives
- Use an iterative way of working: present initial results to process stakeholders, construct an improved event log, analyze the event log, present the results again etc.

*Principle 6 (project leader/sponsor)* – For a successful process mining project (outcome) one needs to ensure a sufficient quality of the data & event logs (high-level intervention) by (specific interventions):

- Choosing an appropriate amount of data to analyze
- Ensuring that the data meets the requirements for correctly constructing an event log
- Ensuring the data/event log is complete (represents the relevant activities in the process)
- Ensuring that the data/event log has the right level of aggregation
- Ensuring that the data is reliable (in the sense that it accurately represents what actually happened)

## 5.5 Chapter conclusion

This chapter evaluated the a-priori process mining success model through the means of a multiple case study. It was shown that most of the a-priori process mining success model is valid, only the moderating effect of process complexity on the relationship between resource availability and process mining success was removed. In order to provide actionable knowledge for practitioners, specific design propositions were formulated so that the process mining success factors can be realized. The next chapter will deal with the implications of these findings and the limitations of this study.

# 6. Conclusion & discussion

This chapter discusses the main conclusions, contributions, suggestions for future research and limitations of this study. Section 6.1 presents the conclusions, section 6.2 discusses the contributions and section 6.3 elaborates on the possible directions for future research.

## 6.1 Conclusion

This research attempted to address the research problem related to process mining project success: it was unclear how organizations can successfully conduct process mining projects. Little or no research has been conducted in the area of process mining success. Therefore, the following research questions were formulated to address this problem:

- How can organizations successfully conduct process mining projects?
  - What are the antecedent factors of process mining success?
    - Are there any contextual variables that influence the effects of these antecedent factors?
    - What are sub-constructs of the antecedent factors?
    - How can the antecedent factors be realized?
  - What success dimensions are appropriate to measure process mining success?
    - What are sub-constructs of the success dimensions?

Through the investigation of theory and practice, an a-priori model of process mining project success was developed, tested and respecified. This model provides more insight on how to measure process mining project success and which factors contribute to this success. It was found that the antecedent factors of management support, project management, resource availability, process miner expertise, process mining approach and data & event log quality influence process mining project success.
Important dimensions of process mining success are: model quality, process impacts and project efficiency. Further statistical testing will have to be conducted to see if these dimensions form a comprehensive measure of process mining project success.

Several moderating variables were found, including: importance, type of IT system and process complexity. Some support was found for: importance influencing the relationship between management support and process mining success, type of IT system influencing the relationship between data & event log quality and process mining project success, type of IT system influencing the relationship between process miner expertise and process mining project success, and process complexity influencing the relationship between data & event log quality and process mining project success. It has to be said that this evidence regarding moderating factors is quite 'weak': further statistical testing will have to be conducted to validate these mediating effects.

It should be noted that the respecified model results mostly relate to process mining projects where the objective is business process improvement, as top management support and process impacts seem less relevant for auditing purposes.

Additionally, to provide practitioners with prescriptive knowledge on how to conduct process mining projects, theory-based and research-based principles were formulated. These principles are linked to the success factors in the a-priori and respecified models of process mining success. Both theoretical and research-based design principles were formulated. In order to link these two, they are contrasted and condensed in table 5.1 below to construct the final design propositions, as the research-design development cycle described in section 1.3 advocates. The first column of table 5.1

represents the found theoretical design principles, in the second column theoretical and practical principles are merged, the third column represents the found practice-based design principles. These final design propositions will need further testing as they are not validated outside of this study.

## 6.2 Research contributions

The specific contributions of this research project are as follows:

4. This thesis points to the need of proper success evaluation of process mining projects in process mining research
5. This thesis has generated new information on how process mining projects can be evaluated on what factors affect the success of a process mining project
6. This thesis can be seen as an initial step for developing design propositions that provide actionable, specific interventions for practitioners to increase the success of their process mining projects

*Identifying the need for success evaluation in process mining research*
In chapter 1 it was already mentioned that this thesis is not typical of process mining research. Most research in the area of process mining is quite ‘technical’ in nature, focussing on (the application of) specific algorithms, techniques and tools that are used in process mining. Little to no attention has been paid to evaluating process mining project success, and to other types of factors that can affect process mining success. Yet, it has been demonstrated that if process mining is to show its value and efficacy in practice, that this is necessary.

*Improving understanding of process mining success*
Through the construction, testing and respecification of the a-priori success model, it has become clear how process mining project success can be measured and which factors contribute to process mining success. This model contributes by facilitating discussion on the topic of process mining by visualizing all the important relationships.

*Development of design propositions*
This research is also not typical for process mining research in the sense that it can be characterized for a part as design science research. This type of research seeks to provide practitioners with prescriptive knowledge, which is not common in process mining research. The value of these propositions is that they represent actionable knowledge for practitioners on how to realize specific interventions, which in turn can lead to process mining success. As such, it has aimed to increase the practical relevancy of process mining research.

| Theory-based principle | Design proposition | Research-based principle |
| --- | --- | --- |
| *Success factor 1: Management support (project leader)* | | |
| | It is important to obtain management support to be able to have an impact on the process. This is even more important in a situation where the importance of the project is consider low as opposed high. | It is important to obtain management support to be able to have an impact on the process. This is even more important in a situation where the importance of the project is consider low as opposed high. |
| *Success factor 2: Project management (project leader)* | | |
| To successfully conduct a process mining project there needs to be good project management in place which can be implemented through involving the right people knowledgeable about the process. This helps to explain the meaning of activities in the process and to determine the appropriate level of granularity of the analysis. | The project stakeholders need to ensure good project management through:<br>○ Determining clear objectives and scope of analysis of the project beforehand<br>○ Involving important process stakeholders in the project and making clear arrangements beforehand as to what is expected from them<br>○ Having frequent discussions of findings with a business domain expert and a data analysis/information system expert to validate the findings<br>○ Planning an intermediate discussion of results with the project sponsor | It is important to have good project management in place, through:<br>○ Determining clear objectives and scope of analysis of the project beforehand<br>○ Deciding which important process stakeholders to involve in the project and making clear arrangements as to what is expected from them. This is done beforehand<br>○ Having frequent discussions of findings with a business domain expert and a data analysis/information system expert to validate the findings<br>○ Planning an intermediate discussion of results with the project sponsor |
| *Success factor 3: Resource availability (project leader/project sponsor)* | | |
| | It is important to have sufficient availability of resources, which is further specified as follows:<br>○ Having the necessary data available for analysis<br>○ Having people knowledgeable about the process available | It is important to have sufficient availability of resources, which is further specified as follows:<br>○ Having the necessary data available for analysis<br>○ Having people knowledgeable about the process available |

| | | |
|---|---|---|
| | ○ Having people knowledgeable about the system (data) available<br>Having (relational) data models and documentation of the process available to use | ○ Having people knowledgeable about the system (data) available<br>○ Having (relational) data models and documentation of the process available to use |
| *Success factor 4: Process miner expertise (process miner)* | | |
| To conduct a successful process mining project, the process miner needs to have sufficient expertise by having or obtaining knowledge of the business process under consideration and having knowledge of the data structure and contents of the relevant information systems. The last part is especially true for process mining in SAP R/3 systems. This is necessary to be able to construct valid event logs and correctly interpret the results. | It is important for the process miner to have a certain degree of process mining expertise by:<br>○ Obtaining knowledge of the business process<br>○ Obtaining knowledge of the source system and its data structure (especially important for the context of ERP systems)<br>○ Having general data analysis skills<br>○ Knowing how to construct an event log from source data<br>○ Knowing which process mining tools and algorithms to use and how to use them<br>○ Being able to correctly interpret the process mining results | It is important for the process miner to have a certain degree of process mining expertise by:<br>○ Obtaining knowledge of the business process<br>○ Obtaining knowledge of the source system and its data structure (especially important for the context of ERP systems)<br>○ Having general data analysis skills<br>○ Knowing how to construct an event log from source data<br>○ Knowing which process mining tools and algorithms to use and how to use them<br>○ Being able to correctly interpret the process mining results |
| *Success factor 5: Process mining approach (process miner)* | | |
| To enable the success of a process mining project a good process mining approach be adopted, which consists of:<br>○ Discussing results with the client; never give the analysis results without explanation<br>○ Combining mining perspectives to get different views on the process | Process miners should take care to adopt an appropriate process mining approach, which consist of:<br>○ Choosing the best approach for obtaining the data, often this involves directly accessing a copy or extract of the production data<br>○ Use the appropriate tools for event log conversion<br>○ Use the appropriate process mining algorithms to deal with real-life data (e.g. | Process miners should take care to adopt an appropriate process mining approach through:<br>○ Choosing the best approach for obtaining the data, often this involves directly accessing a copy or extract of the production data<br>○ Use the appropriate tools for event log conversion<br>○ Use the appropriate process mining algorithms to deal with real-life data (e.g. noise) |

| | | |
|---|---|---|
| | noise)<br>○ Combining mining perspectives<br>○ Use an iterative way of working: present initial results to process stakeholders, construct an improved event log, analyze the event log, present the results again etc.<br>○ Never give the results without explanation | ○ Combining mining perspectives<br>○ Use an iterative way of working: present initial results to process stakeholders, construct an improved event log, analyze the event log, present the results again etc. |
| *Success factor 6: Data & event log quality (project leader/project sponsor)* | | |
| The success of a process mining project depends on having data and event logs of sufficient quality which is characterized by:<br>○ Having the right amount of data<br>○ Conformance of data to process mining requirements<br>○ The amount of 'noise' in the data<br>○ The representativeness of the data | One needs to ensure a sufficient quality of the data & event logs (high-level intervention) by (specific interventions):<br>○ Choosing an appropriate amount of data to analyze<br>○ Ensuring that the data meets the requirements for correctly constructing an event log<br>○ Ensuring the data/event log is complete (represents the relevant activities in the process)<br>○ Ensuring that the data/event log has the right level of aggregation<br>○ Ensuring that the data is reliable (in the sense that it accurately represents what actually happened) | One needs to ensure a sufficient quality of the data & event logs (high-level intervention) by (specific interventions):<br>○ Choosing an appropriate amount of data to analyze<br>○ Ensuring that the data meets the requirements for correctly constructing an event log<br>○ Ensuring the data/event log is complete (represents the relevant activities in the process)<br>○ Ensuring that the data/event log has the right level of aggregation<br>○ Ensuring that the data is reliable (in the sense that it accurately represents what actually happened) |

**Table 5.8 - Construction of final design propositions**

## 6.3 Limitations & suggestions for further research

**Research Limitations**

Obviously, this research also has its share of limitations. The most important ones were deemed as follows (adapted from Bandara, 2007):

- *Immature research field* - Process mining success is a relatively immature research field. This is the first study to adress this topic. That is why it draws heavily on closely related fields (adapted from Bandara, 2007). An attempt was made to clearly demonstrate the relatedness of process mining and other fields
- *A-priori conceptualization* - Quite strong a-priori conceptualization occurred, possibly filtering out important contextual factors. However, care was taken to also leave enough flexibility to add or remove constructs in the case study phase
- *Introduction of bias* - Introduction of bias into the multiple case study phase
    - Through the selection of case studies, although an effort was made to define a strict population of case studies
    - Coding and analysis was mostly done by one researcher. However, an effort was made to validate the coding by having an additional coder code some of the documents and transcripts
- *Timing aspect* – There is a difference in the amount of time that has passed between the execution and completion of the different case studies. For this reason, respondents could respond in different ways because they are more knowledgeable about the project (results are more recent), or because they already have had time to do something with the process mining results. For this reason, cases were selected to be as recent as possible.

**Suggestions for further research**

After conducting this research, several future research suggestions can be formulated:

- *Statistically validation of the process mining success model* – Through the construction of a survey, administering of this survey to a large group of different process mining project stakeholders, and subsequent statistical analysis of the respecified model of process mining success, confidence can be further increases in the validity of the process mining success, especially to increase the validity of the moderating relationships in the model for instance. It should be noted that a large enough population of process mining case studies needs to exists

- *Testing of design principles* – The design principles that were formulated need to be further tested in practice in process mining project to increase their validity

- *Formulation of new design principles* – Both practice and research are encouraged to develop new design principles to achieve process mining success to increase the practical relevancy of process mining research and to provide input from practice for process mining research. The design principles that were presented in this thesis do by no means represent an exhaustive, comprehensive set of principles.

## References

Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. Proceedings of the 11th International Conference on Data Engineering, Taipei: Taiwan.

Aguilar-Savén, R. S. (2004). Business process modelling: Review and framework. *International Journal of Production Economics, 90,* 129-149.

Alter, S. (2002). Information systems: The foundations of E-business. Fourth edition. Pearson Education: New Jersey.

Atkinson, R. (1999). Project Management: cost, time, quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17, 6, 337-342.

Bandara, W., Gable, G. C., & Rosemann, M. (2005). Factors and measures of business process modelling: model building through a multiple case study. *European Journal of Information Systems, 14,* 347-360.

Bandara, W. (2007). *Process Modelling Success Factors and Measures.* PhD thesis Queensland University of Technology, Brisbane.

Benbasat, I, Goldstein, D., & Mead, M. (1987). The case study research strategy of information systems. MIS Quarterly, 11, 3, p.369-386.

Bozkaya, M., Gabriels, J., & van der Werf, J. M. (2009). Process Diagnostics: a Method Based on Process Mining. In *International Conference on Information, Process, and Knowledge Management*.

Buijs, J. (2010). Mapping Data Sources to XES in a generic way. Master thesis Eindhoven University of Technology, Eindhoven.

Cooke-Davies, T. (2002). The “real” success factors on projects. *International Journal of Project Management*, 20, p.185-190.

Cook, J., & Wolf, A. (1998). Discovering models of software processes from event-based data**.** *ACM Transactions on Software Engineering and Methodology,* 7, 3, 215-249.

De Medeiros, A.K., van der Aalst, W.M.P., & Weijters, A.J.M.M. (2003). Workflow mining: Current status and future directions, Lecture Notes in Computer Science, 2888, 389-406.

Davenport, T., & Short, J. (1990). The New Industrial Engineering: Information Technology and Business Process Redesign. Sloan Management Review, 31, 4, 11–27.

de Boer, W. (2010). *Data- and Process Mining of Educational Data.* Master Eindhoven University of Technology.

DeLone, W., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3, 1, 60-95.

DeLone, W. H. & McLean, E. R. (2003). The DeLone and McLean Model of IS Success: A Ten-Year Update. *Journal of Management Information Systems, 19,* 9-30.

Denyer, D, Tranfield D, van Aken J. “Developing Design Propositions through Research Synthesis”. Organization Studies 2008, 29; p. 393-413.

Driessen, S. (2006). *Finding Process Bottlenecks from Logged Process Data using Data Mining,* master thesis*,* Eindhoven University of Technology.

Eisenhardt, K. (1989). Building theories from case study research. *The academy of management review,* 14, 4, p.532-550.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996c). Knowledge discovery and data mining: Towards a Unifying Framework. In (pp. 82-88).

Feelders, A., Daniels, H., & Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Information & Management, 37,* 271-281.

Gerke, K., Mendling, J., & Tarmyshov, K. (2009). Case Construction for Mining Supply Chain Processes. *Lecture Notes in Business Information Processing, 21,* 181-192.

Green, P., & Rosemann, M. (2000). Integrated process modelling: An ontological evaluation. *Information Systems,* 25, 2, 73-87.

Günther, C. W. & van der Aalst, W. M. P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. *Lecture Notes in Computer Science, 4714,* 328-343.

Gunther, C. W., Rozinat, A., & van der Aalst, W. M. P. (2008). *Monitoring Deployed Application Usage with Process Mining*.

Hammer, M., & Champy, J. (1993). Reengineering the corporation: a manifesto for business revolution. New York: Harper Business.

Hinojosa, E. U. (2008). *Process Mining Applied to the Change Control Board Process.* Eindhoven University of Technology.

Ingvaldsen, J. E. & Gulla, J. A. (2007). Preprocessing Support for Large Scale Process Mining of SAP Transactions. In *International Conference on Business process management*.

Lassche, R. (2010). *Care Pathway Analysis and Redesign: A Methodologu.* Master Thesis Eindhoven University of Technology.

Manilla, H. (2007). Methods and Problems in Data Mining. In *Database Theory — ICDT '97*, 41-55.

Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M. P., & Bakker, P. J. M. (2009). Application of Process Mining in Healthcare - A Case Study in a Dutch Hospital. *Communications in Computer and Information Science, 25,* 425-438.

Maruster, L. & van Beest, N. R. T. P. (2009). *Redesigning business processes: a methodology based on simulation and process mining techniques,* 267-297.

Melão, N. & Pidd, M. (2000). A conceptual framework for understanding business processes and business process modelling. *Information Systems Journal, 10,* 105-129.

Nemati, H. R. & Barko, C. D. (2003). Key factors for achieving organizational data-mining success. *Industrial Management & Data Systems, 103,* 282-292.

Mulrow C, Cook D, Davidoff F. “Systematic Reviews: Critical Links in the Great Chain of Evidence”. Annals of Internal Medicine 1997; 126; p. 389-391.

Plsek, P., Bibby, J., & Whitby, E. (2007). Practical methods for extracting explicit design rules grounded in the experiences of organizational managers. Journal of applied behavioural science, 43, p.153-170.

Ramesh, A. (2006). *Process Mining in Peoplesoft.* Master Thesis Eindhoven University of Technology: Eindhoven.

Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W. M. P. (2007). Discovering simulation models. *Information Systems, 34,* 305-327.

Rozinat, A., Alves de Medeiros, A. K., Gunther, C. W., Weijters, A. J. J. M., & van der Aalst, W. M. P. (2008). The Need for a Process Mining Evaluation Framework in Research and Practice. *Lecture Notes in Computer Science, 4928,* 84-89.

Rusu, S. (2010). *Discovery and Analysis of Field Service Engineer Process Using Process Mining.* Eindhoven University of Technology.

Seddon, P.B., Staples, S.D., Patnayakuni, R., & Bowtell, M.J (1998). The IS effectiveness matrix: the importance of stakeholder and system in measuring IS success. Proceedings of the international conference on Information systems, Atlanta, USA.

Segers, I. E. A. (2007). *Investigating the application of process mining for auditing purposes.* Eindhoven University of Technology.

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

Tranfield, D., Denyer, D., Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of a Systematic Review. *British Journal of Management* 2003; 14; p. 207 – 222.

Van Burg, E., Romme, G., Gilsing, V., & Reymen, I. (2008). Creating university spin-offs: a science based design perspective. *Journal of product innovation management*, 25, p. 114-128.

Van der Aalst, W.M.P, van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A.J.M.M. (2003). Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering, 47*, 2003.

van der Aalst, W. M. P. & Weijters, A. J. J. M. (2004). Process mining: a research agenda. *Computers In Industry, 53,* 231-244.

van der Aalst, W. M. P. (2005a). Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing. *Requirements Engineering, 10,* 198-211.

van der Aalst, W. M. P., Alves de Medeiros, A. K., & Weijters, A. J. J. M. (2005b). Genetic Process Mining. *Lecture Notes in Computer Science, 3536,* 48-69.

van der Aalst, W. M. P. (2007a). Exploring the CSCW spectrum using process mining. *Advanced Engineering Informatics, 21,* 191-199.

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

Van der Aalst, W.M.P. (2009). Process-Aware Information Systems: Lessons to be learned from Process Mining. Lecture Notes in Computer Science, 5460, 1-26.

van der Aalst, W. M. P. (2010). Process Discovery : Capturing the Invisible. *IEEE Computational Intelligence Magazine, 5,* 28-41.

van der Aalst, W. M. P., Pesic, M., & Song, M. (2010). Beyond Process Mining: From the Past to Present and Future. *Lecture Notes in Computer Science, 6051,* 38-52.

van der Aalst, W. M. P., Rubin, V., Verbeek, H. W. M., van Dongen, B. F., Kindler, E., & Gunther, C. W. (2010). Process mining: a two-step approach to balance between underfitting and overfitting. *Software and Systems Modelling, 9,* 87-111.

van Dongen, B. F., Alves de Medeiros, A. K., Verbeek, H. W. M., Weijters, A. J. J. M., & van der Aalst, W. M. P. (2005). The ProM framework: A new era in process mining tool support. *Lecture Notes in Computer Science, 3536,* 1105-1116.

van Dongen, B. F. & van der Aalst, W. M. P. (2005). A Meta Model for Process Mining Data. In (pp. 309-320).

van Giessel, M. (2004). *Process Mining in SAP R/3 : A Method for Applying Process Mining in SAP R/3.* Master Thesis, Eindhoven University of Technology: Eindhoven.

Vergidis, K., Turner, C. J., & Tiwari, A. (2008). Business process perspectives: Theoretical developments vs. real-world practice. *International Journal of Production Economics, 114,* 91-104.

Weske, M. (2007). *Business Process Management: Concepts, Languages and Architectures*. Berlin Heidelberg: Springer-Verlag.

Yin, R. (2003). Case study research: design and methods. Third edition. Sage publications: California.

# Appendix A – Systematic review procedure

Elements of the systematic review approach of Tranfield et al. (2003) were adopted for conducting the literature review, since:

- A systematic review approach is well-suited for research that aims to inform both policy and practice, which is the aim of this research project.
- It increases transparency of the literature review process, and reduces researcher bias (Tranfield et al. 2003).

Section A.1 describes the systematic review protocol that was used for this study. Section A.2 summarizes the process of selecting publications. Section A.3 gives an overview of the selected publications. Section A.4 describes the research synthesis method that was used to construct the theoretical background chapter.

## A.1 Systematic review protocol

Tranfield et al. (2003) describe the systematic review protocol as: "a plan that helps to protect objectivity by providing explicit descriptions of the steps to be taken. The protocol contains information on the specific questions addressed by the study, the population (or sample) that is the focus of the study, the search strategy for identification of relevant studies, and the criteria for inclusion and exclusion of studies in the review."

### A.1.1 Systematic review questions

Initially, to gain more insight into the research field of business process mining, two rudimentary research questions were formulated:

1. *What is business process mining?*
2. *Which research fields are closely related to business process mining? What are important differences and commonalities?*

From talks with process mining experts, the author's own experience and an initial scan of literature, it quickly became clear that business process mining is closely related to the fields of business process modelling and data mining.

The next question relates to the research question of this master thesis project. Process mining has the potential advantage for organizations to uncover valuable process information from data that is stored in information systems. Still, the actual application of process mining within an organization requires valuable resources. Therefore, the question is how the success of a process mining project can be or is evaluated within an organization?

3. *How can the success of a process mining project within an organization be evaluated?*

Initial talks with experts and initial literature search results yielded few to none publications on process mining success. A possible explanation for this is the relatively young age of the research field (publications since 2001). The majority of the articles seemed to emphasize the technical development of process mining, while a smaller number of process mining case studies were found that describe the application of process mining within an organizational context or with real-life data sets. Still, most publications only conclude something about the applicability of process mining. No publications were found that give a comprehensive description of process mining success evaluation. This is why the following questions were formulated:

4. *How is the success of information systems commonly evaluated?*

5. *How is success evaluated in the related discipline of business process modelling?*
6. *How is success evaluated in the related discipline of data mining?*

Additionally, in line with the second master thesis project research question, it would be interesting to know what factors can contribute to process mining project success in organizations.
It seemed that the majority of process mining publications only mention technical factors that affect the applicability of process mining. Only a few process mining publications mention other types of factors. Still, there exists no comprehensive overview of process mining success factors. That is why it would be interesting to inventarize all the possible success factors that are currently mentioned in scientific literature and other types of scientific publications on business process mining:

7. *Which important success factors for business process mining projects conducted in organizations are described in literature?*

The final question would be how to integrate theoretical findings on success factors and measures into one conceptual model of process mining success:

8. *How can a conceptual model of process mining success be constructed using knowledge from the fields of business process modelling and data mining, and process mining case studies?*

### A.1.2 Keywords, search strategy & population

The next phase of this literature review was to accumulate relevant scientific literature. Based on the previously formulated research questions, keywords were specified to guide the search for scientific literature in the ABI/Inform, INSPEC and Google Scholar databases.

The following search terms were used:

- "process mining"
- "data mining"
- "business process modelling"
- "process mining" AND "success"
- "information systems" AND "success"
- "process modelling" AND "success"
- "data mining" AND "success

Table 4 below provides an overview of the number of relevant publications that were found during this database search. These were publications were deemed relevant if the title or publication summary could possibly provide an answer to the research questions.

| ***Search keyword*** | ***Google Scholar*** | | ***ABI/Inform*** | | ***Inspec*** | | ***Total*** |
|---|---|---|---|---|---|---|---|
| | *# of hits* | *# relevant* | *# of hits* | *# relevant* | *# of hits* | *# relevant* | |
| "process mining" | 2990 | 18 | 21 | 9 | 56 | 5 | 32 |
| "data mining" | 686000 | 4 | 1715 | 0 | 14177 | 0 | 4 |
| "business process modelling" | 6280 | 5 | 113 | 0 | 183 | 0 | 5 |
| "process mining" AND "success" | 793 | 8 | 0 | 0 | 3 | 0 | 8 |
| "information systems" AND "success" | 2380 | 2 | 81 | 0 | 55 | 0 | 2 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| "process modelling" AND "success" | 9320 | 1 | 17 | 1 | 117 | 0 | 2 |
| "data mining" AND "success" | 60800 | 2 | 62 | 1 | 641 | 0 | 3 |
| | | | | | | **Total** | **56** |

Table 9 – Database search results

Additional to the database search, publications of master students and well-known researchers from Eindhoven University of Technology (Mathematics & Computer Science and Information Systems research groups) were searched for process mining publications. Both research groups are considered to be leading worldwide in the field of business process mining. Important resources for finding publications from these groups are: the process mining Wiki website (http://www.processmining.org) and the Eindhoven University of Technology library website (http://w3.tue.nl/en/services/library/). Lastly, the obtained publications were scanned for interesting references through the so called 'snowball' method. Table 5 below presents the results of the literature search process:

| | **Publications** | | | | | |
|---|---|---|---|---|---|---|
| **Search method** | **Practice-Oriented (PO)<br><br>Process mining** | **Technically-Oriented (TO)<br><br>Process mining** | **Data mining (DM)** | **Business process modelling<br><br>(BPM)** | **Information system/ success<br><br>(ISS)** | **Total** |
| Database search | 11 | 30 | 7 | 7 | 2 | **57** |
| TU/e publications | 11 | - | - | - | 1 | **12** |
| Snowball method | - | 2 | 1 | 4 | 2 | **9** |
| **Total** | **22** | **32** | **8** | **11** | **5** | **78** |

Table 10 – Number of publications found per search method and category

From table 2 it can be seen that 78 publications relevant publications were selected. Also, in light of the literature review questions, a distinction was made between process mining publications which are technically-oriented (TO) and focus on describing specific algorithms and technical issues within process mining, and practice-oriented (PO) publications that focus on describing the actual application of process mining on real-life data sets or real-life settings.

### A.1.3 Selection criteria

**Relevance criteria**

During the first stages of the systematic review, 78 potentially relevant articles were obtained. The aim of this section is to provide a set of relevance and quality criteria to reduce the number of publications used for the literature review. Possibly not all publications might be as relevant for providing a quality answer to the literature review questions as was thought at first glance. Also, for reasons of research scope, the number of publications has to be reduced. For an article to get selected, it is first necessary to pass all relevance criteria:

| Criterium | Subject area | | | | |
|---|---|---|---|---|---|
| | **TO** | **PO** | **DM** | **BPM** | **ISS** |
| *R1* | Provides a comprehensive treatise of process mining and/or important general issues with process mining | Describes the real-life application of process mining and mentions important issues/success factors | Describes the basic concept of data mining and/or mentions important success factors and measures | Describes the basic concept of process modelling and/or mentions important success factors and measures | Provides a comprehensive treatise of information system success |
| R2 | Has not been superseded | Not used as a case study in this research project | Has not been superseded | Has not been superseded | Has not been superseded |

Table 11 - Relevance selection criteria

The above criteria provide an opportunity to scrutinize the publications in more detail. It should be noted that 2 publications (Staal, 2010, Rozinat, 2006b) were not selected for the literature review, since they were selected for use in the multiple case study phase. After applying the relevance criteria, 25 publications were removed, leaving 52 publications for inclusion.

**Quality criteria**

A quality assessment was undertaken of the publications the remaining publications. Tranfield et al. (2003) propose a set of quality criteria that can be used to rate a study's internal validity. Ideally, due to the subjective nature of some of these measures, the rating should be done by multiple independent raters to reduce bias. Due to time and scope constraints this was not feasible in this study. Instead objective external criteria were used in an attempt to validate the quality of a study.
However, the number of citations can help to distinguish high quality publications, since it is an indicator of how many scholars have valued the publication and subsequently used the publication in their own work. A high impact factor is an indicator of a high-quality journal, a journal that will probably set high quality standards for the published article. Finally, it is important to obtain recent information on all topics . Either one of these criteria must be adhered to in order for a publication to get included. An exception is made for practice-oriented TUE master thesis publications since they provide unique information on the application process mining. It assumed that the quality of these publications is sufficient, because they are explicitly rated on quality by a first and second assessor from Eindhoven University of Technology.

| Criterium | Subject area | | | | |
|---|---|---|---|---|---|
| | **TO** | **PO** | **DM** | **BPM** | **ISS** |
| *Q1* | # citations >=45 AND/OR journal impact factor >1.2 AND/OR publication date =>2007 | # citations >45 AND/OR journal impact factor >1.2 AND/OR publication date =>2007 OR a TUE Master Thesis | # citations >100 AND/OR journal impact factor >1.2 AND/OR publication date =>2007 | # citations >100 AND/OR journal impact factor >1.2 AND/OR publication date =>2007 | # citations >100 AND/OR journal impact factor >1.2 AND/OR publication date =>2007 |

Table 12 - Quality selection criteria

All 52 publications were included. No publications were removed, confirming the quality of these publications.

## A.2 Literature selection

| # | 1st Author | Category | Year | Journal/book | Impact factor | citations | R 1 | R 2 | Q 1 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Agrawal | TO | 1995 | Proceedings of the International Conference on Data Engineering | not available | 3456 | X | X | X |
| 2 | Aguilar-Saven | BPM | 2004 | International Journal of Production Economics | 2.068 | 197 | X | X | X |
| 3 | Alter | ISS | 2002 | Book | not available | 224 | X | X | X |
| 4 | Alves de Medeiros | TO | 2007 | Lecture Notes in Computer Science | not available | 36 | | X | |
| 5 | Atkinson | DM | 1999 | International Journal of Project Management | not available | 321 | X | X | X |
| 6 | Bandara | BPM | 2005 | European Journal of Information Systems | 1.2 | 61 | X | X | X |
| 7 | Bandara | BPM | 2007 | QUT PhD Thesis | not available | 6 | X | X | X |
| 8 | Bozkaya | PO | 2009 | International Conference On Information, Process and Knowledge Management | not available | 2 | X | X | X |
| 9 | Buijs | TO | 2010 | Master Thesis TUE | not available | 4 | X | X | X |
| 10 | Cardoso | TO | 2006 | International Journal of Business Intelligence and Data Mining | not available | 10 | | X | |
| 11 | Chen | DM | 1996 | IEEE transactions on knowledge and data engineering | 2.285 | 1515 | | X | |
| 12 | Cook | TO | 1998 | ACM Transactions on Software Engineering and Methodology | not available | 477 | X | X | X |
| 14 | Davenport | BPM | 1990 | Sloan Management Review | not available | 2235 | X | X | X |
| 13 | de Boer | PO | 2010 | Master Thesis TUE | not available | 0 | X | X | X |
| 17 | de Medeiros | TO | 2003 | Lecture Notes in Computer Science | not available | 74 | X | X | X |
| 15 | DeLone | ISS | 1992 | Information Systems Research | not available | 3588 | X | X | X |
| 16 | DeLone | ISS | 2003 | Journal of Management Information Systems | 2.098 | 1528 | X | X | X |
| 18 | Driessen | PO | 2006 | Master Thesis TUE | not available | 0 | X | X | X |
| 19 | Duan | TO | 2009 | The Journal of Systems and Software | 1.34 | 2 | | X | |
| 20 | Edgington | PO | 2010 | Decision Support Systems | 1.873 | 1 | | X | |
| 21 | Fayyad | DM | 1996a | IEEE Expert | not available | 386 | X | | |
| 22 | Fayyad | DM | 1996b | AI magazine | 1.018 | 0 | X | | |
| 23 | Fayyad | DM | 1996c | Proceedings Second International Conference Knowledge Discovery and Data Mining | not available | 486 | X | X | X |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 24 | Feelders | DM | 2000 | Information & Management | 2.282 | 99 | X | X | X |
| 25 | Gerke | TO | 2009 | Lecture Notes in Business Information Processing | not available | 2 | X | X | X |
| 29 | Green | BPM | 2000 | Information systems | 1.66 | 209 | X | X | X |
| 26 | Gunther | TO | 2006 | Lecture Notes in Computer Science | not available | 35 | | X | |
| 27 | Gunther | TO | 2007 | Lecture Notes in Computer Science | not available | 50 | X | X | X |
| 28 | Gunther | PO | 2008 | BPM Center Report | not available | 1 | X | X | X |
| 30 | Hammer | BPM | 1993 | Book | not available | 5387 | X | X | X |
| 31 | Hinojosa | PO | 2008 | Master Thesis TUE | not available | 0 | X | X | X |
| 32 | Ho | PO | 2009 | International Journal of Production Research | 0.803 | 0 | | X | |
| 33 | Ingvaldsen | PO | 2008 | BPM workshops 2007 | not available | 10 | X | X | X |
| 34 | Kock | BPM | 2009 | Decision Support Systems | 1.873 | 4 | | X | |
| 35 | Lassche | PO | 2010 | Master Thesis TUE | not available | 0 | X | X | X |
| 36 | Lau | PO | 2009 | International Journal of Production Economics | 2.068 | 1 | | X | |
| 37 | Li | TO | 2010 | Information Systems and E-Business Management | not available | 2 | | X | |
| 38 | Li | TO | 2008 | IEEE International Conference on Service Computing | not available | 12 | | X | |
| 39 | Manilla | DM | 1997 | Proceedings of the 6th International Conference on Database Theory | not available | 201 | X | X | X |
| 40 | Mans | PO | 2009 | Communications in Computer and Information Science | not available | 11 | X | X | X |
| 41 | Maruster | PO | 2009 | Knowledge and information systems | 1.733 | 1 | X | X | X |
| 42 | Melao | BPM | 2001 | Information Systems Journal | not available | 117 | X | X | X |
| 43 | Misev | PO | 2008 | Lecture Notes in Computer Science | not available | 3 | | X | |
| 44 | Mutschler | TO | 2005 | Proceedings International Workshop on Enterprise Modelling and Information Systems Architecture | not available | 9 | | X | |
| 45 | Nemati | DM | 2003 | Industrial Management & Data Systems | 1.535 | 13 | X | X | X |
| 46 | Quaglini | TO | 2010 | Journal of Software Maintenance and Evolution | 1.143 | 0 | | X | |
| 47 | Ramesh | PO | 2005 | Master Thesis TUE | not available | 0 | X | X | X |
| 51 | Rozinat | PO | 2006b | BETA paper working series | not available | 17 | X | X | |
| 48 | Rozinat | TO | 2009a | Information Systems | 1.66 | 18 | X | X | X |
| 49 | Rozinat | TO | 2006a | In Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN | not available | 10 | X | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 50 | Rozinat | TO | 2008 | Lecture Notes in Computer Science | not available | 16 | X | X | X |
| 52 | Rubin | PO | 2007 | Lecture Notes in Computer Science | not available | 22 | | X | |
| 53 | Rusu | PO | 2010 | Master Thesis TUE | not available | 0 | X | X | X |
| 54 | Seddon | ISS | 1998 | Proceedings of the international conference on Information systems | not available | 51 | X | X | X |
| 55 | Segers | PO | 2007 | Master Thesis TUE | not available | 0 | X | X | X |
| 56 | Staal | PO | 2010 | Master Thesis TUE | not available | 0 | X | X | |
| 57 | Tiwari | TO | 2008 | Business Process Management Journal | not available | 15 | X | X | X |
| 58 | Van der Aalst | BPM | 2003 | Data & Knowledge Engineering | not available | 587 | X | X | X |
| 59 | van der Aalst | PO | 2007b | Information Systems | 1.66 | 159 | X | X | X |
| 60 | van der Aalst | TO | 2004 | Computers in Industry | 2.04 | 159 | X | X | X |
| 61 | van der Aalst | TO | 2008 | Journal of Intelligent Information Technologies | not available | 0 | | X | |
| 62 | van der Aalst | TO | 2005a | Requirements Engineering Journal | not available | 80 | X | X | X |
| 63 | van der Aalst | TO | 2005b | Lecture Notes in Computer Science | not available | 51 | X | X | X |
| 64 | van der Aalst | TO | 2007a | Advanced Engineering Informatics | 1.848 | 7 | X | X | X |
| 65 | van der Aalst | TO | 2010a | IEEE Computational Intelligence Magazine | not available | 0 | X | X | X |
| 66 | van der Aalst | TO | 2010b | Software and Systems Modeling | 1.533 | 24 | X | X | X |
| 67 | van der Aalst | TO | 2010c | BPM Center Report | not available | 0 | X | X | X |
| 68 | van der Aalst | ISS | 2009 | Lecture Notes in Computer Science | not available | 3 | X | X | X |
| 69 | van Dongen | TO | 2005a | Proceedings of the CAiSE Workshops | not available | 51 | X | X | X |
| 70 | van Dongen | TO | 2005b | Lecture Notes in Computer Science | not available | 205 | X | X | X |
| 71 | van Giessel | PO | 2004 | Master Thesis TUE | not available | 4 | X | X | X |
| 72 | Vergidis | BPM | 2008 | International Journal of Production Economics | 2.068 | 24 | X | X | X |
| 73 | Weber | TO | 2006 | Lecture Notes in Computer Science | not available | 23 | | | |
| 74 | Weijters | TO | 2006 | BETA paper working series | not available | 46 | X | X | X |
| 75 | Wen | TO | 2009 | Journal of Intelligent Information Systems | 0.98 | 25 | | X | |
| 76 | Weske | BPM | 2007 | Book | not available | 304 | X | X | X |
| 77 | Zhang | PO | 2010 | Master Thesis TUE | not available | 0 | | | |
| 78 | Zhou | TO | 2010 | Decision Support Systems | 1.873 | 0 | | X | |

## A.3 Shortlist

(Selected publications appear in bold face)


**Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. Proceedings of the 11th International Conference on Data Engineering, Taipei: Taiwan.**

**Aguilar-Savén, R. S. (2004). Business process modelling: Review and framework. *International Journal of Production Economics, 90,* 129-149.**

**Alter, S. (2002). Information systems: The foundations of E-business. Fourth edition. Pearson Education: New Jersey.**

Alves de Medeiros, A. K., Pedrinaci, C., van der Aalst, W. M. P., Domingue, J., Song, M., Rozinat, A. et al. (2007). An Outlook on Semantic Business Process Mining and Monitoring. *Lecture Notes in Computer Science, 4806,* 1244-1255.

**Atkinson, R. (1999). Project Management: cost, time, quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17, 6, 337-342.**

**Bandara, W., Gable, G. C., & Rosemann, M. (2005). Factors and measures of business process modelling: model building through a multiple case study. *European Journal of Information Systems, 14,* 347-360.**

**Bandara, W. (2007). *Process Modelling Success Factors and Measures*. PhD thesis Queensland University of Technology, Brisbane.**

**Bozkaya, M., Gabriels, J., & van der Werf, J. M. (2009). Process Diagnostics: a Method Based on Process Mining. In *International Conference on Information, Process, and Knowledge Management*.**

**Buijs, J. (2010). Mapping Data Sources to XES in a generic way. Master thesis Eindhoven University of Technology, Eindhoven.**

Cardoso, J. & Lenic, M. (2006). Web process and workflow path mining using the Multimethod approach. *International Journal of Business Intelligence and Data Mining, 1,* 304-328.

Chen, M., Han, J., & Yu, P. S. (1996). Data Mining : An Overview From a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering, 8*, 866-883.

**Cook, J., & Wolf, A. (1998). Discovering models of software processes from event-based data. *ACM Transactions on Software Engineering and Methodology,* 7, 3, 215-249.**

**De Medeiros, A.K., van der Aalst, W.M.P., & Weijters, A.J.M.M. (2003). Workflow mining: Current status and future directions, Lecture Notes in Computer Science, 2888, 389-406.**

**Davenport, T., & Short, J. (1990). The New Industrial Engineering: Information Technology and Business Process Redesign. Sloan Management Review, 31, 4, 11–27.**

**de Boer, W. (2010). *Data- and Process Mining of Educational Data.* Master Eindhoven University of Technology.**

**DeLone, W., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, 3, 1, 60-95.**

**DeLone, W. H. & McLean, E. R. (2003). The DeLone and McLean Model of IS Success: A Ten-Year Update. *Journal of Management Information Systems, 19,* 9-30.**

**Driessen, S. (2006). *Finding Process Bottlenecks from Logged Process Data using Data Mining.* Eindhoven University of Technology.**

Duan, H., Zeng, Q., Wang, H., Sun, S. X., & Xu, D. (2009). Classification and evaluation of timed running schemas for workflow based on process mining. *Journal of Systems and Software, 82,* 400-410.

Edgington, T. M., Raghu, T. S., & Vinze, A. S. (2010). Using process mining to identify coordination patterns in IT service management. *Decision Support Systems, 49,* 175-186.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From Data Mining to Knowledge Discovery in Databases. *AI Magazine, 17,* 37-54.

Fayyad, U. M. (1996b). Data Mining and Knowledge Discovery : Making Sense out of Data. *IEEE Expert: Intelligent Systems and their Applications, 11,* 20-25.

**Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996c). Knowledge discovery and data mining: Towards a Unifying Framework. In (pp. 82-88).**

**Feelders, A., Daniels, H., & Holsheimer, M. (2000). Methodological and practical aspects of data mining. *Information & Management, 37,* 271-281.**

**Gerke, K., Mendling, J., & Tarmyshov, K. (2009). Case Construction for Mining Supply Chain Processes. *Lecture Notes in Business Information Processing, 21,* 181-192.**

**Green, P., & Rosemann, M. (2000). Integrated process modelling: An ontological evaluation. *Information Systems,* 25, 2, 73-87.**

Gunther, C. W., Rinderle, S., Rinderle, S., & van der Aalst, W. M. P. (2006). Change Mining in Adaptive Process Management Systems. *Lecture Notes in Computer Science, 4275,* 309-326.

**Günther, C. W. & van der Aalst, W. M. P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. *Lecture Notes in Computer Science, 4714,* 328-343.**

**Gunther, C. W., Rozinat, A., & van der Aalst, W. M. P. (2008). *Monitoring Deployed Application Usage with Process Mining*.**

**Hammer, M., & Champy, J. (1993). Reengineering the corporation: a manifesto for business revolution. New York: Harper Business.**

**Hinojosa, E. U. (2008). *Process Mining Applied to the Change Control Board Process.* Eindhoven University of Technology.**

Ho, G. T. S., Lau, H. C. W., Kwok, S. K., Lee, C. K. M., & Ho, W. (2009). Development of a co-operative distributed process mining system for quality assurance. *International Journal of Production Research, 47,* 883-918.

**Ingvaldsen, J. E. & Gulla, J. A. (2007). Preprocessing Support for Large Scale Process Mining of SAP Transactions. In *International Conference on Business process management*.**

Kock, N., Verville, J., Danesh-Pajou, A., & DeLuca, D. (2009). Communication flow orientation in business process modeling and its effect on redesign success: Results from a field study. *Decision Support Systems, 46,* 562-575.

**Lassche, R. (2010). *Care Pathway Analysis and Redesign: A Methodologu.* Master Thesis Eindhoven University of Technology.**

Lau, H. C. W., Ho, G. T. S., & Zhao, H. Y. (2009). Development of a process mining system for supporting knowledge discovery in a supply chain network. *International Journal of Production Economics, 122,* 176-187.

Li, C., Reichert, M., & Wombacher, B. (2008). Mining Process Variants: Goals and Issues. In *IEEE International Conference on Services Computing* (pp. 573-576).

Li, J., Wang, H. J., Zhang, Z., & Zhao, J. L. (2010). A policy-based process mining framework: mining business policy texts for discovering process models. *Information Systems and E-Business Management, 8,* 169-188.

**Manilla, H. (2007). Methods and Problems in Data Mining. In *Database Theory — ICDT '97*, 41-55.**

**Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M. P., & Bakker, P. J. M. (2009). Application of Process Mining in Healthcare - A Case Study in a Dutch Hospital. *Communications in Computer and Information Science, 25,* 425-438.**

**Maruster, L. & van Beest, N. R. T. P. (2009). *Redesigning business processes: a methodology based on simulation and process mining techniques,* 267-297.**

**Melão, N. & Pidd, M. (2000). A conceptual framework for understanding business processes and business process modelling. *Information Systems Journal, 10,* 105-129.**

Misev, A. & Atanassov, E. (2011). Performance Analysis of GRID Middleware Using Process Mining. *Lecture Notes in Computer Science, 5101,* 203-218.

Mutschler, B., Bumiller, J., & Reichert, M. (2005). An Approach to quantify the Costs of Business Process Intelligence. In.

**Nemati, H. R. & Barko, C. D. (2003). Key factors for achieving organizational data-mining success. *Industrial Management & Data Systems, 103,* 282-292.**

Quaglini, S. (2010). Information and communication technology for process management in healthcare: a contribution to change the culture of blame. *Journal of Software Maintenance and Evolution: Research and Practice, 22,* 435-448.

**Ramesh, A. (2006). *Process Mining in Peoplesoft.* Master Thesis Eindhoven University of Technology: Eindhoven.**

Rozinat, A., Mans, R. S., & van der Aalst, W. M. P. (2006a). Mining CPN Models Discovering Process Models with Data from Event Logs. In (pp. 57-76).

Rozinat, A., de Jong, I. S. M., Gunther, C. W., & van der Aalst, W. M. P. (2006b). *Process Mining of Test Processes: A Case Study*.

**Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W. M. P. (2007). Discovering simulation models. *Information Systems, 34,* 305-327.**

**Rozinat, A., Alves de Medeiros, A. K., Gunther, C. W., Weijters, A. J. J. M., & van der Aalst, W. M. P. (2008). The Need for a Process Mining Evaluation Framework in Research and Practice. *Lecture Notes in Computer Science, 4928,* 84-89.**

Rubin, V., Gunther, C. W., van der Aalst, W. M. P., Kindler, E., van Dongen, B. F., & Schafer, W. (2007). Process Mining Framework for Software Processes. *Lecture Notes in Computer Science, 4470,* 169-181.

**Rusu, S. (2010). *Discovery and Analysis of Field Service Engineer Process Using Process Mining.* Eindhoven University of Technology.**

**Seddon, P.B., Staples, S.D., Patnayakuni, R., & Bowtell, M.J (1998). The IS effectiveness matrix: the importance of stakeholder and system in measuring IS success. Proceedings of the international conference on Information systems, Atlanta, USA.**

**Segers, I. E. A. (2007). *Investigating the application of process mining for auditing purposes.* Eindhoven University of Technology.**

Staal, J. (2010). Using process mining and data mining techniques to define and improve standardization in a workflow healthcare environment. Master Thesis Eindhoven University of Technology, Eindhoven.

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

**Van der Aalst, W.M.P, van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A.J.M.M. (2003). Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering, 47*, 2003.**

**van der Aalst, W. M. P. & Weijters, A. J. J. M. (2004). Process mining: a research agenda. *Computers In Industry, 53,* 231-244.**

**van der Aalst, W. M. P. (2005). Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing. *Requirements Engineering, 10,* 198-211.**

**van der Aalst, W. M. P., Alves de Medeiros, A. K., & Weijters, A. J. J. M. (2005). Genetic Process Mining. *Lecture Notes in Computer Science, 3536,* 48-69.**

**van der Aalst, W. M. P. (2007a). Exploring the CSCW spectrum using process mining. *Advanced Engineering Informatics, 21,* 191-199.**

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

van der Aalst, W. M. P. (2008). Mining E-Mail Messages : Uncovering Interaction Patterns and Processes using E-Mail Logs. *International Journal of Intelligent Information Technologies, 4,* 27-45.

**Van der Aalst, W.M.P. (2009). Process-Aware Information Systems: Lessons to be learned from Process Mining. Lecture Notes in Computer Science, 5460, 1-26.**

**van der Aalst, W. M. P. (2010). Process Discovery : Capturing the Invisible. *IEEE Computational Intelligence Magazine, 5,* 28-41.**

**van der Aalst, W. M. P., Pesic, M., & Song, M. (2010). Beyond Process Mining: From the Past to Present and Future. *Lecture Notes in Computer Science, 6051,* 38-52.**

**van der Aalst, W. M. P., Rubin, V., Verbeek, H. W. M., van Dongen, B. F., Kindler, E., & Gunther, C. W. (2010). Process mining: a two-step approach to balance between underfitting and overfitting. *Software and Systems Modelling, 9,* 87-111.**

**van Dongen, B. F., Alves de Medeiros, A. K., Verbeek, H. W. M., Weijters, A. J. J. M., & van der Aalst, W. M. P. (2005). The ProM framework: A new era in process mining tool support. *Lecture Notes in Computer Science, 3536,* 1105-1116.**

**van Dongen, B. F. & van der Aalst, W. M. P. (2005). A Meta Model for Process Mining Data. In (pp. 309-320).**

**van Giessel, M. (2004). *Process Mining in SAP R/3 : A Method for Applying Process Mining in SAP R/3.* Master Thesis, Eindhoven University of Technology: Eindhoven.**

**Vergidis, K., Turner, C. J., & Tiwari, A. (2008). Business process perspectives: Theoretical developments vs. real-world practice. *International Journal of Production Economics, 114,* 91-104.**

Weber, B., Reichert, B., Rinderle, S., & Wild, W. (2006). Towards a Framework for the Agile Mining of Business Processes. *Lecture Notes in Computer Science, 3812,* 191-202.

Weijters, A. J. J. M. (2006). *Process Mining with Heuristics Miner*.

Wen, J., Wang, J., van der Aalst, W. M. P., Huang, B., & Sun, J. (2009). A novel approach for process mining based on event types. *Journal of Intelligent Information Systems, 32,* 163-190.

**Weske, M. (2007). *Business Process Management: Concepts, Languages and Architectures*. Berlin Heidelberg: Springer-Verlag.**

Zhang, Y. (2010). *The Application of Process Mining in Tracking and Tracing Systems.* Master Thesis Eindhoven Universtity Technology.

Zhou, W. & Piramuthu, S. (2010). Framework, strategy and evaluation of health care processes with RFID. *Decision Support Systems, 50,* 222-233.

## A.4 Literature review synthesis

### A.4.1 Synthesis approaches

According to Mulrow (2004) research synthesis is: “the collective term for a family of methods for summarizing, integrating, and, where possible, cumulating the findings of different studies on a topic or research question.” The simplest and most familiar method of research synthesis is the narrative review, although such an approach does not seek to attempt generalization or cumulative knowledge (Tranfield et al., 2003). According to Denyer et al. (2008), “narrative synthesis has the benefit of being able to address a wide range of questions, not only those relating to the effectiveness of a particular intervention. Despite excellent examples of its application, it is criticized because of its potential bias and lack of both transparency and reproducibility.”

Publications pertaining to the sections of business process management, information systems, process mining, business process modelling (success), data mining (success) were synthesized through a narrative review since these sections only aim to provide some background information on these topic areas and do not seek to provide a holistic description of these research areas. This would also lead to an unfeasible increase of the scope of this research project. Furthermore, a narrative review is still well-suited for addressing a wide range of research questions.

## Appendix B – Definition of a-priori model constructs

The definitions of a-priori constructs were partly adapted from Bandara (2007, p. 149), and are listed in table 5 below.

| Process Mining Success Factors |
| --- |
| ***Management support:***<br>Management support refers to the involvement and participation of senior management, and their ongoing commitment and willingness to devote necessary resources and time of senior managers to oversee the process mining efforts. |
| Sub-constructs:<br>o Funding<br>o Provide other types of resources<br>o Decision making<br>o Active participation |
| ***Project Management:***<br>Project management is the management of activities and resources throughout all phases of the process mining project, to obtain the defined project outcomes. |
| Sub-constructs:<br>o Defined objectives (objectives are carefully defined in advance)<br>o Scope-defined (choice of which processes to model and level of detail)<br>o Communication management (systematic communication results to project stakeholders)<br>o People management (inclusion of people with the appropriate skills and knowledge)<br>o Time management<br>o Quality management<br>o Cost management |
| ***Resource availability:***<br>Degree of information available from the project stakeholders for the design, approval and maintenance of the models |
| Sub-constructs:<br>o Support in information gathering<br>o Review and approval of models by process stakeholders<br>o Availability of stakeholders knowledgeable about the process<br>o Available documentation of the process<br>o Data availability |
| ***Process miner expertise:***<br>Process miner expertise describes the experiences of the person conducting the mining, in terms of event log construction, doing process mining analysis and knowledge of the business processes being mined |
| Sub-constructs:<br>o General event log construction skills<br>o Process mining analysis skills<br>o Knowledge of the data structure and relationships |

| |
|---|
| ○ Technical knowledge (knowledge of how to apply process mining within a system)<br>○ Business knowledge (knowledge of the business process being mined) |
| ***Process mining approach:***<br>The way in which the process miner conducts the actual process mining |
| Sub-constructs:<br>○ Way of constructing the event logs<br>○ Iterative approach in creating and analyzing event logs<br>○ Choice of the appropriate process mining algorithms<br>○ Number of different mining perspectives used |
| ***Data & Event log quality***<br>Characteristics of the raw data and subsequently constructed event logs |
| Sub-constructs:<br>○ Data quality (representativeness, noise etc.)<br>○ Conformance to process mining requirements<br>○ Amount of data<br>○ Data integration |
| **Process Mining Mediating factors** |
| ***Importance:***<br>The criticality of the process modelling project to the organization |
| Sub-constructs:<br>○ Importance of the process modelling project<br>○ Relevance of the process modelling project<br>○ Need for the process modelling project |
| ***Complexity:***<br>Many different features of the process modelled |
| Sub-constructs:<br>○ Number of transactions/activities in the process<br>○ Number of dependencies between activities in the process<br>○ Number of stakeholders involved in the process<br>○ Number of inputs and outputs of the process<br>○ Number of alternative options (process paths/process variants) within the process or the structuredness of the process<br>○ New: Number of changes to the process |
| ***Type of IT system:***<br>The type of information system(s) involved in the process under consideration |
| Sub constructs:<br>○ Type (WfMs, ERP, database etc.)<br>○ Structured vs. unstructured<br>○ Data-centred vs. process-centred<br>○ Logging-functionality<br>○ Complexity of the system(s) |

| |
|---|
| **Process Mining Success Measures** |
| ***Model quality:***<br>The extent to which all desirable properties of a model are fulfilled to satisfy the needs of the model users in an effective and efficient way |
| Sub-constructs:<br>o Ease of use of the process models for model users<br>o Realisation of user requirements<br>o Information accuracy (the process model accurately depicted the modelled processes)<br>o Information relevance (information available from the process models was important, relevant and complete)<br>o Understandability of the process models<br>o Conciseness of process models<br>o New: process insights |
| ***Process impacts:***<br>Process Impacts refers to the overall effect of process modelling on the processes modelled (e.g. improvements achieved). |
| Sub-constructs:<br>o Cost-effectiveness (process modelling project resulted in more cost-effective processes)<br>o Staff management (process modelling project resulted in improved understanding of personnel requirements of the modelled process)<br>o Increased product/service quality (The process modelling helped to identify improvements to the quality of products and services resulting from the modelled processes)<br>o Reduced processing time<br>o Overall improved business processes |
| ***Project Efficiency:***<br>The ratio of obtained outcomes over invested resources |
| Sub-constructs:<br>o Efficiency in relation to the invested person days of effort<br>o Efficiency in terms of overall project duration<br>o Efficiency in relation to other resources<br>o The process mining project was effective (achieved its objectives) |

**Table 13 - Definition of a-priori constructs**

## Appendix C – Interview structure

Note that this interview structure has been partly adapted from Bandara (2007, p.434).

### Phase 1 – Context specific information

- State appreciation for cooperating with this interview
- Get permission for an audio recording
- Introduce study
    - Identifying success factors and measures for process mining projects in organizations
- Purpose of interview
    - Get more information on this specific case study at XXXX

**Process modelling project goals**

| Role | Question |
|---|---|
| All | 1. Was the process mining initiative that we are looking at part of a larger initiative? |
| Miner, Leader, Sponsor | 2. How did you go about introducing the process mining initiative within the organisation? |

**Process modelling project organisation**

| Role | Question |
|---|---|
| All | 1. How would you describe the process mining project team?<br>a. Project leader?<br>b. Project sponsor<br>c. Process miners<br>i. Expertise<br>ii. Experience<br>2. What role did you have yourself? |

**Modelling approach**

| Role | Question |
|---|---|
| All | 1. How would you describe the overall *process mining* approach? |
| All | 2. What were “issues” or concerns in the application of process mining? How did you overcome them?<br>3. Was there anything <critical> that had to be there *first,* for you to proceed with the next steps of the process? |
| All | 4. What were things that facilitated the process mining? |
| All | 5. How did you decide which processes to mine? |

**Model applications**

| Role | Question |
|---|---|
| Leader, Sponsor, User | 1. Can you explain for what applications, and how you intend to use these process mining results? |

**Evaluating the process modelling initiative**

| Role | Question |
|---|---|
| Sponsor, Leader, User | 1. Where do you see further applications of process mining at XXXX? |
| Sponsor, | 2. Can you specify the planned next steps of the process mining initiative? |
| All | 3. Have you done any formal 'evaluation' on your process mining initiative? (if yes, ask for procedure and results) |
| Sponsor, Leader, User | 4. Can you specify the benefits that the process modelling initiative set forth for XXXX, which would have otherwise been unattainable? |
| All | 5. What lessons (possible weaknesses) have you learnt from this exercise? *(In other words, what would you do differently, next time?)* |

## Phase 2 – Testing the a-priori model

### Testing the success factors

| Factor | Question |
|---|---|
| Management Support | 1. Did you get the relevant support from the management for the process mining initiative?<br>a. How important was this for you to proceed? |
| Project management | 2. Did you have a "leader" role involved in the project? (get details of this person). What did she/he did or didn't do to assist the process? Do you see this role as critical?<br>3. Did you have specific 'structure' laid out for the modelling team? Did you have specific roles and strategies? How closely did you abide by them?<br>4. Did you have project management strategies? Were they documented? |
| Resource availability | 5. Did you possess all the necessary resources to conduct the process mining project?<br>6. Access to what kind of resources is of high importance? |
| Domain knowledge | 7. Was it important to involve certain people in the process mining project?<br>a. What was their job description/role?<br>b. What did they add?<br>c. How crucial was their input to success?<br>8. How important is it to have ontogical knowledge/understanding of the IT system(s) under consideration?<br>9. How important is it to have a good understanding of the business process under consideration? |
| Process mining expertise | 10. Did you as a miner, or the modellers have sufficient process mining expertise in your view? |

| | |
|---|---|
| Process mining approach | 11. How important is it to adopt an iterative approach while doing process mining?<br>12. How important is it to combine different mining perspectives?<br>13. How important is it to discuss intermediate and final results with clients/project leaders/process owners? |
| Data characteristics | 14. How would you characterize the data quality? How important is this?<br>15. What about the amount of data that you analyzed?<br>16. How representative was the data in your view? |
| Type of system | 17. What type of IT system was involved in process mining? |
| Process characteristics | 18. How variable was the process? Did this have an effect on process mining?<br>19. What do you think of the structure of the processes being modelled? How complex was the process? Did it have an impact? |
| Importance | 20. What do you think of the organisational culture? (their willingness to use process mining). Did it have an impact?<br>21. Was there a clear need for process mining? |

### Testing the Success Measures

| **Factor** | **Question** |
|---|---|
| Overall | 1. Do you think that you achieved a successful process mining initiative?<br>a. What elements would you consider to make this decision? |
| Model quality | 2. Do you think that measuring the quality of the models would be a good way of looking at the overall success of the process mining project?<br>a. What types of things should be asked to measure this? |
| Process impacts | 3. Do you think that measuring the impact that process mining set forth for the process (that was mined) would be a good way of looking at the overall success of the process mining project?<br>a. What types of things should be asked to measure this? |
| Recognisability of results | 4. Do you think that the recognisability of results is a good process mining project success measure? |
| Project efficiency | 5. Do you think that measuring project efficiency would be a good way of looking at the overall success of the process mining project?<br>a. How should one measure project efficiency? |

<Statement of appreciation>
<End of interview>

## Appendix D – Definition of respecified model constructs

Table 6 below presents the definition and sub-constructs of the constructs in the respecified model.

| Process Mining Success Factors |
|---|
| ***Management support:***<br>Management support refers to the involvement and participation of senior management, and their ongoing commitment and willingness to devote necessary resources and time of senior managers to oversee the process mining efforts. |
| Sub-constructs:<br>○ Inform stakeholders on purpose<br>○ Provide access to resources |
| ***Project Management:***<br>Project management is the management of activities and resources throughout all phases of the process mining project, to obtain the defined project outcomes. |
| Sub-constructs:<br>○ Communication management<br>○ Data management<br>○ Objectives defined<br>○ People management<br>○ Scope definition |
| ***Resource availability:***<br>Degree of information available from the project stakeholders for the design, approval and maintenance of the models |
| Sub-constructs:<br>○ Availability of people knowledgeable about the process<br>○ Availability of people knowledgeable about the system<br>○ Available datamodels<br>○ Available documentation<br>○ |
| ***Process miner expertise:***<br>Process miner expertise describes the experiences of the person conducting the mining, in terms of event log construction, doing process mining analysis and knowledge of the business processes being mined |
| Sub-constructs:<br>○ Business process knowledge<br>○ Data analysis expertise<br>○ Event log construction skills<br>○ Tool and algorithm expertise<br>○ Understanding of the data structure |
| ***Process mining approach:*** |

| |
|---|
| The way in which the process miner conducts the actual process mining |
| Sub-constructs:<br>○ Combining mining perspectives<br>○ Choosing the appropriate algorithms<br>○ Iterative way of working |
| ***Data & Event log quality***<br>Characteristics of the raw data and subsequently constructed event logs |
| Sub-constructs:<br>○ Amount of data<br>○ Completeness<br>○ Reliability<br>○ Conformance to process mining requirements |
| **Process Mining Moderating factors** |
| ***Importance:***<br>The criticality of the process modelling project to the organization |
| |
| ***Complexity:***<br>Many different features of the process modelled |
| |
| ***Type of IT system:***<br>The type of information system(s) involved in the process under consideration |
| |
| **Process Mining Success Measures** |
| ***Model quality:***<br>The extent to which all desirable properties of a model are fulfilled to satisfy the needs of the model users in an effective and efficient way |
| |
| ***Process impacts:***<br>Process Impacts refers to the overall effect of process modelling on the processes modelled (e.g. improvements achieved). |
| |
| ***Project Efficiency:***<br>The ratio of obtained outcomes over invested resources |
| |

Table 14 - Definition of respecified model constructs