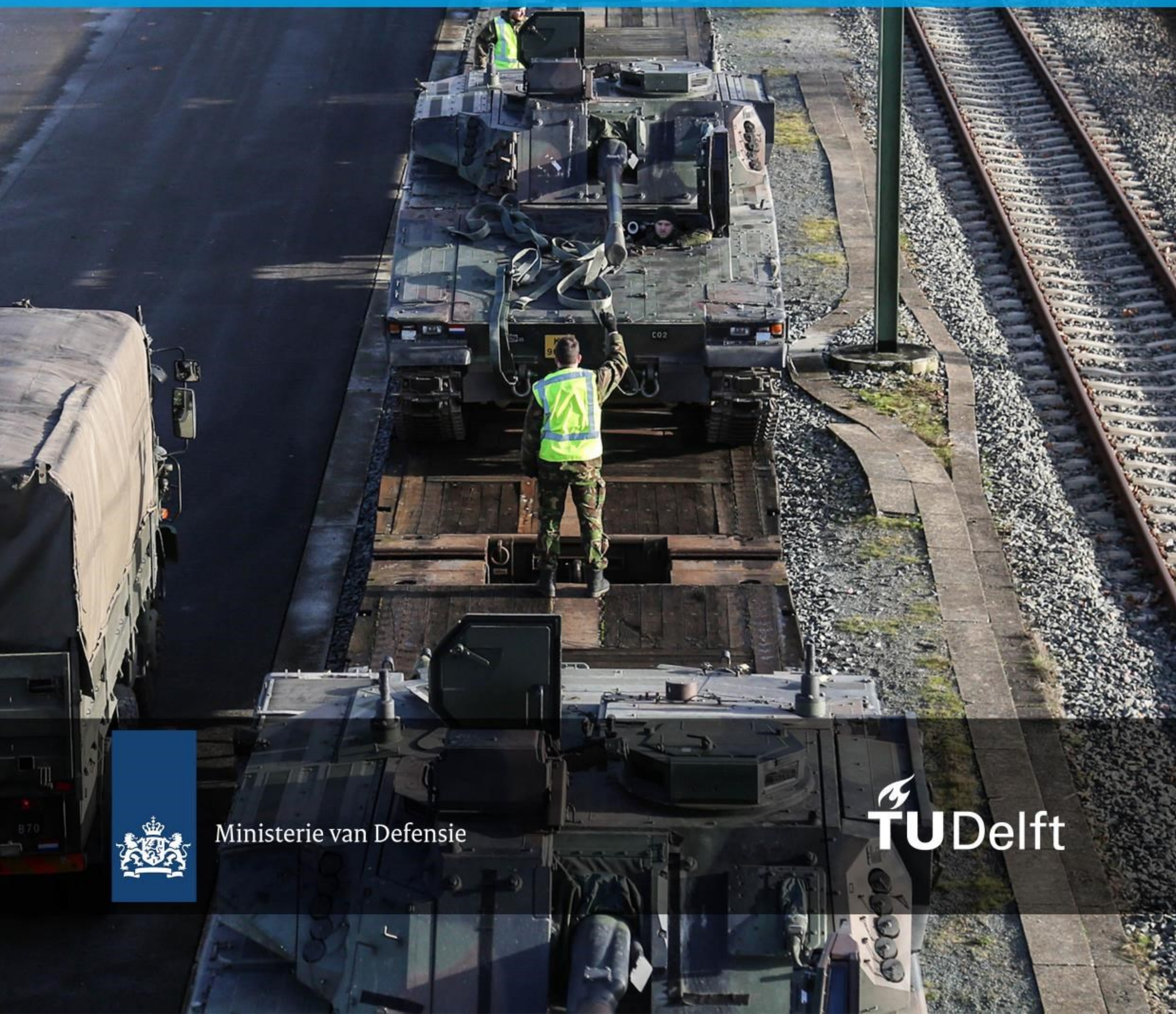
The usability of the Process Mining analysis method to improve processes of the Netherlands Ministry of Defence

Graduation thesis at the Netherlands Ministry of Defence

Rick Driessen
Student number 4025725

13 September 2017

Keywords: Process Mining, Business Process Improvement, Defence, PM^3, Logistics, Big Data



Ministerie van Defensie

TU Delft

Ministerie van Defensie

TU Delft

The usability of the Process Mining analysis method to improve processes of the Netherlands Ministry of Defence

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Management of Technology**

Faculty of Technology, Policy and Management

by

Rick Driessen

Student number: 4025725

To be defended in public on the 27th of September 2017

Graduation committee

Chairperson	: Prof.dr.ir. M.F.W.H.A. Janssen, Section ICT
First Supervisor	: Ir. M.W. Ludema, Section Transport and Logistics
Second Supervisor	: Drs. J. Ubacht, Section ICT
Third Supervisor	: Dr. A.M.H. Teixeira, Section ICT
External Supervisor	: Ing. D.M. de Goede, Ministry of Defence

Ministerie van Defensie

TU Delft

Data is a powerful resource to understand where the problems originated and also help to diffuse emotions and blame. Kanji, 1990

Preface

The thesis in front of you, with the title 'The usability of the Process Mining analysis method to improve processes of the Netherlands Ministry of Defence', is written to fulfil the graduation requirements of the master Management of Technology at Delft University of Technology, the Netherlands. In the period February to September 2017, I studied several aspects of applying process mining at the Netherlands Ministry of Defence.

I deliberately choose to fulfil my thesis at this intriguing organisation. With its flashing recruiting campaigns, you might expect a pumped-up organisation, always ready to move. But instead I only met restrained professionals. It was a unique experience. Not just because of the thesis, but mostly because of the way how my colleagues from the Defence Materiel Organisation welcomed me. Doors were always open (well, sometimes not for security reasons) and help was offered without asking. The various trips to Defence organisations were very interesting, just like the masterclasses, meetings, and presentations.

I want to thank all my colleagues from the DMO and in particular the department Policy Participation. It was a pleasant surprise how quickly I was seen as a full colleague and how much trust I received to pursue my ambition. Thank you, Dennis, for all the discussions we had and the amount of patience it sometimes took to convince me. Thank you, Marcel, Marijn, André, and Jolien for guiding my thought some setbacks but eventually leading me to my thesis defence. My parents, for the continues flow of feedback. And Suzanne, for helping me with the graphics until deep into the night.

With this document also ends my time as a student in Delft. It was an awesome ride and I would not have missed it for the world. So, final thanks for all the friends that made it possible!

Per Scientiam Vires!

Rick Pieter Jan Driessen
September 2017, Delft

Executive summary

Situation

In recent years, the Netherlands Ministry of Defence (MoD) modelled, standardized, and digitally secured many of its organization-broad business processes. Additionally, these processes are now supported and guided by enterprise software (like the Enterprise Resource Planning application SAP). However, MoD still experiences great difficulty in gaining insights in these processes, in particularly its maintenance processes, and steering them productively. Business process improvement methodologies like Lean and Six Sigma are being used, just like various analysis software applications, but a comprehensive methodology is needed to discover, analyse, and enhance formal process models based on factual data from the field.

Complication

The novel business process analyses method Process Mining came to MoD's attention as a promising instrument to tackle this problem. However, MoD lacks experience with process mining as a way to analyse and subsequently improve processes. In addition on how to apply process mining, MoD wants to learn the requirements for starting a process mining project. This leads to the following research question “What kind of processes from the Netherlands Ministry of Defence can benefit from using the process mining analysis method applied as a process improvement methodology?”

Approach

The thesis starts with a literature study on eight business process improvement methodologies (BPIs), viz. Lean, Six Sigma, Lean Six Sigma, Total Quality Management, Lean MRO, Total Productive Maintenance, Business Process Reengineering, and Business Process Improvement. These BPIs are compared with each other in an overview based on nine characteristics. Then, a ninth BPI is studied: process mining. This analysis methodology is used to build a bridge between data mining and model based process analysis (van der Aalst, 2016). Using (often large) observational data sets from enterprise software (like an ERP), the process mining software structures business activities in novel ways that are both understandable and useful to the process owner. These “mined” models can be compared with the theoretical models, if available. And can be used for further analysis and improvement. The thesis elaborates on various key elements of process mining, like event logs, miners, visualisations, and software. However, process mining only gives insights into the process, while a BPI also aims to improve it. To apply process mining and successfully improve business operations, a fitting project methodology is fundamental. The literature on this topic is scattered, still a study is conducted of several methodologies. After that, pros and cons are discussed and process mining is placed in the overview of BPIs.

Besides performing desk research, the author of the thesis fulfils a significant role in the Defence broad pilot Process Mining. With the pilot, MoD aims to gain more hand on experience in using process mining to improve processes. The author is closely involved in managing the pilot as secretary and its analyses as team member. The combination of desk and field research leads to the improvement of the best practise methodology PM^2. And this improved methodology, called PM^3, is further used during the pilot.

The gained knowledge and experience is then used to focus on the main research question. A process mining decision framework is developed according to the waterfall model. This framework can score business process models on several characteristics. The resulting total score determines whether process mining can successfully be applied to improve the process. In its implementation phase, a sample of fourteen of MoD's processes are assessed by the framework to validate its functionality. The sample represents a true reflection of MoD's processes and where tested with the framework as well as by a rational reflection. Ideally, these processes are also verified by running an improvement project for each process. Models with high scores should improve significantly, while models with low scores should not. However, these projects take months or even years to implement, making it impossible to evaluate them as part of this thesis. Alternatively, an evaluation is organised in which an expert panel is asked if it can discover improvement potential in the process models. The outcomes are analysed on whether process mining could be used to find this potential and to actually improve the process. The scores of the framework should be in line with the results of the evaluation.

Results

Next to several discoveries, the study results in two major scientific contributions. First, the development of an integral process mining business improvement methodology: PM^3. This methodology is based on PM^2 by van Eck, Lu, Leemans & van der Aalst (2015), which is used as a guideline at the start of the pilot, but fell short in several areas. PM^3 spends less attention to the data acquisition and processing than its predecessor, since this is (nearly) automated in software used by MoD: ARIS PPM. On the contrary, this methodology focuses attention on the construction and interaction of the team and focuses on continuous improvement. It consists of five steps that correspond to the well-known and proven BPI: Six Sigma. The steps are plan & define, extract & measure, analyse, improve, and control. Based on the several sessions, the process can be raised to a higher level till continuous improvement is possible.

The second key discovery concerns the development of a decision framework that assess a formal process model regarding the usability of process mining to improve the process (as shown on Table 0.1). The framework asks about twelve characteristics that can easily be found in the process model. Five scores are possible, ranging from ++ (2 points), + (1 point), +/- (0 points), - (-1 point), to -- (process mining not suited). The final score is the sum of all points. The higher the score, the more suitable the model for process mining.

Table 0.1: Decision framework

	Characteristic	**Answer (score)**
1	Availability of the minimum required attributes	Available (++) or unavailable (--)
2	Number of roles involved in the process	Many (+), few (-), or none (--)
3	Number of decision points	Many (+), few (+/-), or none (-)
4	Number of activities	Many (+), few (+/-), or none (--)
5	Lead time (average)	Short (+) or long (+/-)
6	Quality of dataset	High (+) or low (-)
7	Methods of importing data	Off-the-shelf adapter (++), adapter can be developed of file-based data import (+/-), or none (--)
8	Possibility to filter on attributes	Yes, profound (++); yes, superficial (+/-); or no (-)
9	Nature of process	Manufacturing or auditing (++), finance and office-administration or maintenance (+), logistics (+/-), or product development (-)
10	Process maturity (years of experience)	Mature (+) or immature (-)
11	Availability of process model	Yes, of high quality (+); yes, of low quality or no (+/-)
12	Mandated is known	Yes (++) or no (-)

On average, the processes scored 9,9 points with a minimum of -4 and a maximum of 15. Because of the restricted scoring, the lowest score possible is -13 and the highest is 18. In just two occasions the decisions chart gave a score deviant of the researcher's verdict. In order to verify the framework, it was successfully presented to a group of experts.

Recommendation for MoD

For MoD, process mining can be a valuable instrument for improving processes. It can convert a gut feeling into clear numbers and models, which can form the start of an improvement project. It is thus recommended to make process mining available to the organisation. The PM^3 and decision framework can be of great value. The decision framework still requires several rounds of field testing before its operation can be determined with certainty. However, even in its current form, the framework will encourage people to think about the requirements for process mining, which by definition is good. When it is fully tested, it will help MoD order its improvement projects on their potential. This reduces costs and sequentially streamline MoD's logistics.

Future research

The thesis gives a good insight in the possibilities of process mining, but it also comes with several opportunities for future research. For example, PM^3 has demonstrated itself during the pilot, but this have not yet led to measurable improvements (only important insights in the processes). Future research is required to verify its true protentional. Second, the decision framework is tested by assessing fourteen of MoD's processes and evaluated by discussion it with ten system experts. Yet again, empirical research is probably the only way to truly verify it, but this was not possible during the thesis. Third, both PM3 and the decision framework focus on MRO within MoD. It is unlikely that a change of domain or organisation makes a significant difference, but the possibility should be taken into account and further investigated. And finally, a significant limitation of the framework is that it only scores on process mining. So, if it presents a very low score, it does not suggest an alternative BPI. This fell outside the scope of the thesis, but is nevertheless an interesting research opportunity.

Abbreviations

Abbreviation	Description
6σ	Six Sigma
AI	Artificial Intelligence
BI	Business Intelligence
BPI	Business process improvement methodology
BPMN	Business Process Model and Notation
CRISP-DM	CRoss-Industry Standard Process for Data Mining
CRM	Customer Relationship Management
CSV	Comma-Separated Values
DEMO	Design & Engineering Methodology for Organizations
DMO	Defence Materiel Organisation
EPC	Event-driven Process Chain
ERP	Enterprise Resource Planning
ETL	Extract, Transform, and Load
IoT	Internet of Things
IT	Information Technology
IV	*Informatie Voorziening* (Information Supply)
KPI	Key Performance Indicator
ML	Machine Learning
MoD	The Netherlands Ministry of Defence
MRO	Maintenance, Repair, and Overhaul
MXML	Macromedia Flex Markup Language
n/a	Not applicable
PM	Process Mining
PPM	Process Performance Manager
QRM	Quick Response Manufacturing
RCM	Reliability Centred Maintenance
RfC	Request for Change
SAP	*Systeme, Anwendungen, und Produkte* (systems, applications, and products)
SCM	Supply Chain Management
SSM	SAP Solution Manager
ToC	Theory of Constraints
TPM	Total Productive Maintenance
UML	Unified Modelling Language
XES	eXtensible Event Stream

List of figures

List of tables

Content

1. Introduction

This chapter starts with the objective and boundary conditions of the thesis project. It then continues with some background information on the Netherlands Ministry of Defence, its tasks, its organisation, and one of its current challenges: the integration of processes across the Ministry's different organisations. The Defence broad pilot Process Mining is discussed briefly and the chapter closes with an outline of the thesis report.

1.1. The Ministry of Defence

In an increasingly dangerous world with a growing number of geopolitical conflicts, the Netherlands Ministry of Defence (MoD) retains a vital role in defending the peace and safety in the Netherlands and abroad. With its motto: "Protecting what we value", MoD fights for freedom where turmoil reigns, and where it is peaceful, MoD keeps the peace. Where disasters occur, MoD offers help. However, after several relative stable decades and a severe financial crisis, MoD has gone through numerous rounds of cutbacks and reorganizations, forcing them to use their remaining resources sparingly.

MoD comprises seven organisational elements and is led by three key officials. The largest organisations are the four armed services: the Air Force (*Koninklijke Luchtmacht*), the Army (*Koninklijke Landmacht*), the Navy (*Koninklijke Marine*), and the Marechaussee (*Koninklijke Marechaussee*). The Defence Materiel Organisation (DMO) supports the armed services by providing products and services. The DMO has three main tasks. It is responsible for the procurement and sales of, among others, weapon systems, clothing, fuel, and ammunition. It has a leading role in innovation projects within Defence (technological and also social). And it is responsible for the logistics services, that also cover maintenance, repair, and overhaul (MRO) activities.

Traditionally, the four armed services were organized as (semi)independent bodies. Nowadays, however, defence organisations are expected to work as a single team and act if possible per universal process models. This path of business process integration led to the fact that the last couple of years, many (business) processes were modelled, standardized, and secured in MoD's process management software 'ARIS Business Process Analysis Platform' (Software AG, n.d.), as shown in Figure A.1.1 (in the appendix). The modelled processes describe, for example, how to plan and perform maintenance of a weapon system, but also how weapon upgrades should be managed, how invoices should be paid, and how transport should be arranged. All employees have access to the process models with their computer account.

Standardizing processes comes with several challenges. First of all, people need to change their (ingrained) work patterns and this often leads to resistance. In some cases, the standardization will actually lead to a reduction of productivity from a user's point of view. It therefore seems fair to assume that not all employees directly adjust their work patterns or are even willing to change them voluntarily. Because of the size and complexity of the organization, it is rather difficult to measure the total adoption and effects of the standardization. This, in turn, makes it almost impossible to intervene in a timely manner if needed. Even though the business process integration intended to streamline logistics and reduce costs, the Netherlands Court of Audits rang the alarm bell on the bad (physical) state MoD was in, proclaiming the need for a more efficient (maintenance) organisation (Algemene Rekenkamer, 2016). This need led, among other things, to the interest of investigating the possibilities of process mining as a business process improvement methodology.

1.2. Defence broad pilot Process Mining

During the Defence broad pilot Process Mining, the possibilities of process mining within MoD are examined and described. The pilot also provides guidance for future process mining projects and is meant to make the organization enthusiastic about process mining. Because of its orientating nature, a possible conclusion could be that process mining has no added value within MoD. In that case, it is described why this is so and under which prerequisite(s) process mining might be useful.

The pilot is initiated by Central Staff and adopted by the DMO with the author of the thesis as secretary and team member. The pilot consists of three phases (these can be seen in Figure 1.1.). First, the preparation phase, where three activities are carried out. The team has to be assembled. During the course of the pilot, a business analyst, an expert in the process mining software, an expert in the digital infrastructure of MoD, and a secretary/team leader (the author of the thesis) form the core process mining group. When the team is formed, it has to be introduced in process mining. Lastly, the process mining software has to be deployed and configured. For the second phase, a total of five (freestanding) cases are carried out. Each case finds its origin in an actual need of different parts of MoD. The fact that the pilot is adopted by the DMO, steers the cases towards the domain of logistics and MRO. For each case, the core process mining group is supported by experts from the domain in question. For the third and last phase, an evaluation is carried out, with a report as final product. It describes the course of the five cases, but more importantly, it analyses the costs and benefits of using process mining. The evaluation will lead to the decision if and how process mining will be implemented within MoD.

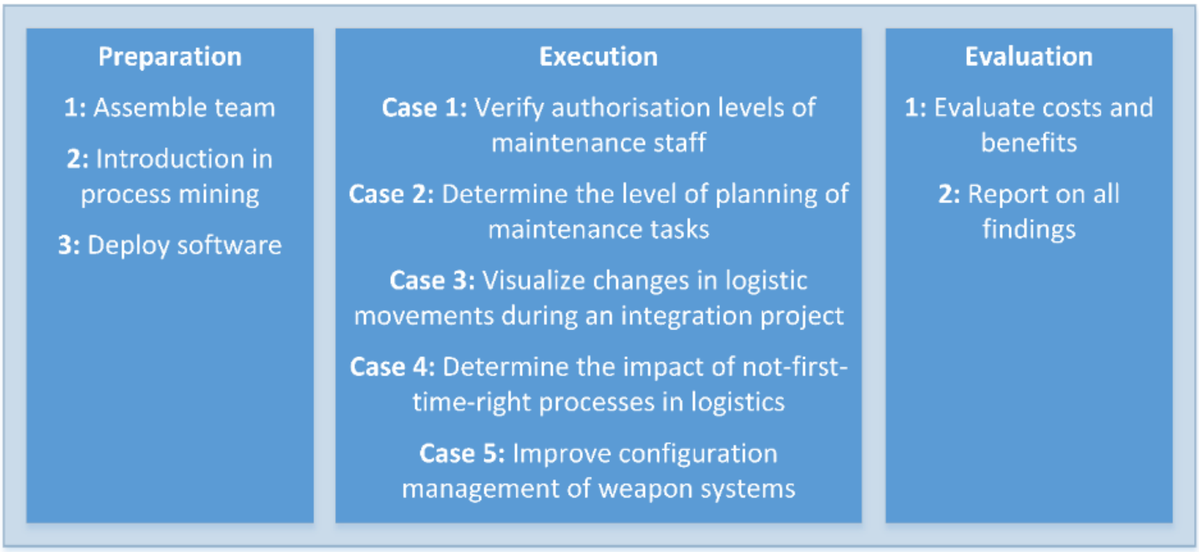


Figure 1.1: Structure of the Defence broad pilot Process Mining

The connection between the pilot and the thesis project can be found in the fact that the author of the thesis is actively involved in all aspects of the pilot. Together with the client, he had to conceive five cases and for each case a group of supporting experts had to be sought. Since the project groups are relatively small (five to eight people) and none are familiar with process mining, the author often had to guide the team through the process.

The pilot and the thesis are separate projects, but experience gained during the pilot is used in the thesis. The pilot was assigned more resources (man-hours and budget) and enjoys a hands-on approach while the thesis is not financially supported and is more literature based. Also, since the DMO and the cases steer towards the domain of logistics and MRO, the thesis focusses on the same domain.

1.3. Report outline

This report describes a complete thesis project. The first chapter acts as an introduction. It starts with some general background information on MoD, then explains its current challenges, and finally a possible solution is introduced, viz. process mining. This solution is examined during the Defence broad pilot Process Mining.

Chapter 2 provides a more specific background on the problem and its origin. It starts with a problem statement and continues with the research objective, questions, design, and methodology.

In Chapter 3, a thorough investigation of business process improvement methodologies is presented. Lean is seen by many as the founding theory on process improvement. That is why it is given a central role in the thesis's history of business process improvement. After that, some definitions are given that are often used in the thesis. Then, eight different improvement methodologies are discussed and compared. In the second part of the chapter, process mining is being discussed. In isolation, process mining can (only) be seen as a business analysis methodology. After elaborating on several key elements of this analysis methodology, process mining is placed in an applied context as a business process improvement methodology and compared with the already discussed traditional methodologies.

In Chapter 4, the process mining methodology 'PM^2' is further examined. Based on experience gained during the Defence broad pilot Process Mining, several improvements are suggested, with the improvement methodology 'PM^3' as the result.

In Chapter 5, a framework that can evaluate process models on their mining suitability is developed. This development is done according to the waterfall model (as described by Royce (1987)). Its first phase describes the requirements. Second, the framework is designed. This starts with the development of a conceptual model of all business processes. This model is used to define what elements a business process has and serves as a starting point for determining what characteristics can be used to establish how well process mining can be applied to a process. After a short introduction on how processes are modelled at MoD, twelve of these characteristics are distinguished and are placed in a decision chart. In the third phase, the decision framework is implemented. A sample of all processes of MoD is taken and placed in the decision chart to verify its function. The last phase concerns the evaluation. Together with other internal experts, the results of the sample are compared with expectations from the field.

The thesis is closed in Chapter 6 with some conclusions and a discussion. Here, research limitations are also discussed and recommendations for future research are given.

To improve the readability of the thesis, the larger and the additional tables, figures, models, and sections are placed in a separate appendix. Figures in the appendix are marked with an A (e.g. A.1.1.). The reader can place the appendix next to the thesis while reading.

2. Problem definition and research design

Chapter 1 discussed the reason of this report, focussing on the Netherlands Ministry of Defence and its he Defence broad pilot Process Mining. This chapter starts with a problem statement where the situation is briefly explained. An organisational and a scientific research objective are formulated and this leads to the main research question, which is divided in six sub-questions. Then the research design and methodology are presented.

2.1. Problem statement

MoD needs to make its maintenance organization more efficient and process mining is an upcoming and potential breakthrough methodology to discover, conform, and enhance business processes. However, MoD has no experience with this, while conducting a process mining project is not straightforward and the literature does not appoint an all-encompassing roadmap. Besides process mining, other methodologies are available. Literature on where process mining is positioned in this landscape is scarce, making it difficult to decide if process mining is the right methodology for a project.

2.2. Research objective

The assignment from MoD, and therewith one of the research objectives of this thesis project, is to discover the added value of process mining when analysing and improving logistic processes of MoD. MoD is aware that several other business process improvement (BPI) methodologies exist, but is mainly interested in how a process mining project can be executed and if it should be added to its standard toolkit.

This project goes further than the assignment from MoD. A scientific and societal objective is also formulated. The thesis will contribute by investigating how and where process mining is positioned among other business process improvement methodologies. And on what kind of processes (both logistical and non-logistical) process mining can be used. Thus, helping scholars and process managers choosing the right methodology for improving business processes. Even though the results are generally applicable, most examples come from the logistical domain.

2.3. Research questions

Based on information from Chapter 1, viz. the problem statement and the research objective, the following main research question was developed:

What kind of processes from the Netherlands Ministry of Defence can benefit from using the process mining analysis method applied as a process improvement methodology?

In order to answer the main research question, seven sub-questions were formulated. The first four will be answered by literature research and the last three by field research.

Sub-question 1: What business process improvement methodologies are described in the literature?
The goal of this question is to get an overview of the most used methods. For example, several well-known methods are Lean, Six Sigma, and TQM. Process mining is relatively novel, but also belongs, when integrated into a project methodology, on this list. The overview will be used to place process mining in the landscape of traditional improvement methodologies.

Sub-question 2: How can process mining be used as a process analysis methodology?

On its own, process mining is an analysis methodology; a tool to create different views on how business processes are executed. The literature on this topic is scattered and a review is needed. This answer will give a clear view on the concept of process mining.

Sub-question 3: How can process mining be used as a business process improvement methodology?
In order to improve processes with process mining, the methodology needs to be integrated in a business process improvement framework. To answer this question, the available literature needs to be collected and reviewed. Is there a strict procedure and does the literature give tips and tricks? Known methodologies are compared.

Sub-question 4: How does the Netherlands Ministry of Defence execute process mining projects in the Defence broad pilot Process Mining?
During the implementation of the Defence broad pilot Process Mining, the author had the possibility to experiment with several smaller process mining projects. This sub-question compares the methodologies from the literature with practical experience. Lessons learned will be used to improve the methodologies from sub-question 3.

Sub-question 5: What are the characteristics of processes at the Netherlands Ministry of Defence?
To understand what kind of processes benefit from process mining, processes from MoD need to be inventoried and characterized. However, in total there are hundreds of processes, so a sample is taken. The processes that are selected are analysed on its main characteristics. A generic model of all business process is developed. This model is evaluated with an actual Defence process.

Sub-question 6: What processes at the Netherlands Ministry of Defence are suitable for process mining projects?
Based on the sampled processes and characteristics, and knowledge gained by answering the previous questions, a decision framework is constructed to test the processes of MoD for their process mining suitability. The input for this framework are the process models already available at MoD.

2.4. Research design

The methodological set-up of this research can be viewed in Figure 2.1.

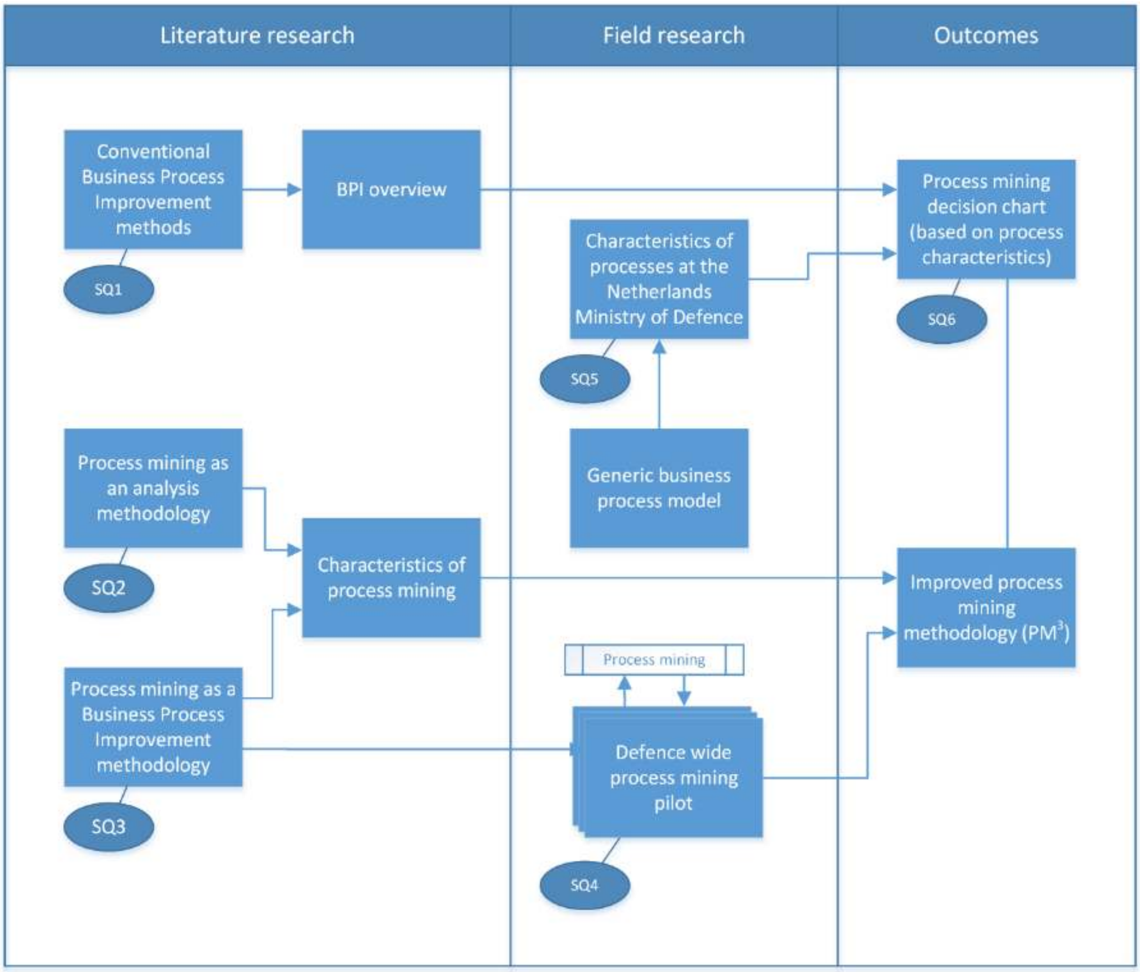


Figure 2.1: Schematic view of the research process

2.5. Research methodology

The first four sub-questions will be answered after an extensive literature research. In total over eighty journal articles, conference proceedings, and books were found, studied, and used. Most used sources are Google Scholar, Scopus, and Processmining.org.

Sub-question 1 will result in an overview on eight popular BPIs based on nine characteristics. Sub-questions 2 and 3 complement this overview by providing theoretical background on process mining. Together these three questions make up Chapter 3.

The knowledge gained during the literature research will be used during the author's role in the Defence broad pilot Process Mining. This results in answering sub-question 4.

Sub-question 5 will also be answered based on experience gained during pilot and previous questions in combination with a generic model of all processes that was developed of the thesis. This model is verified by comparing it with one of MoD's actual business processes.

Sub-question 6 is answered last, since it requires all knowledge gained till this point. It needs to be said that the decision framework leaves room for personal interpretation, since the focus is on several quantitative and qualitative characteristics of the process model. However, a single process can have multiple kinds of problems, solved in different ways. Based on the characteristics of the process model alone, there is still room for discussion about the effects of process mining as an improvement methodology. This gives room for future research.

3. Theory of business process improvement

This chapter describes the current scientific knowledge on the topic of improving business processes. The chapter is divided into two parts. The first provides an overview of traditional business process improvement methodologies. And the second part introduces the concept of process mining. Part of the research is comparing the traditional BPIs with process mining. However, the two cannot simply be compared, since process mining on its own is an analysis methodology that discovers problems and areas with improvement potential, while a complete BPI also aims to address problems and improve the process. To apply process mining and successfully improve business operations, a fitting project methodology is fundamental. This is also discussed in the second part.

3.1. Traditional business process improvement methodologies

Today's global and digital economy demands industrial organizations to have a production process with high efficiency, availability, and reliability. Companies must improve at a faster rate than their competition to gain or keep a strategic advantage (Baluch, Abdullah, & Mohtar, 2012). For the last hundred years, many theories have been developed and deployed to obtain this strategic advantage.

First, a short summary on the history of business process improvements is given. Here, focus is placed on Lean, because it is the origin of all modern BPIs and is still well known. After that, several related concepts will be defined. Following that, a more in-debt analysis of the common business process improvement methodologies will be presented. There is an endless collection of variations in common business process improvement methodologies and only the eight most famous and influential will be discussed. This selection was made based on an online search and the number of published papers per methodology found on Scholar.google.com. Lean MRO and TPM may not meet the criteria of being most famous and influential, but are especially interesting because of their links to MRO (Maintenance, Repair, and Overhaul). Three other methodologies where found, but left out the comparison since they were not popular enough, viz. Theory of Constraints (ToC), Reliability Centred Maintenance (RCM), and Quick Response Manufacturing (QRM). On the eight selected BPIs, a vast amount of literature can be found. However, an overview capable of comparing them is missing, while this is very useful for managers choosing an BPI and for scholars studying them. This first part will resolve this by giving a systematic overview.

3.1.1. History of business process improvement

Many historians will give Henry Ford the honours of being the first pioneer in process improvement. Till 1913, producers followed the American System: an assembly hall filled with general-purpose machines grouped by process. Parts where made separately and brought to central places for sub- and final assembly by master craftsmen (Lean Enterprise Institute, 2006). Krafcik (1988) characterized this as the craftsmen-era: work was not standardized, inventories and buffers were large, workers had a large span of control, and they worked together at a moderate level. The outcome was a highly flexible and buffered production system with a low level of efficiency.

Ford came up with the idea to create flow by lining up special-purpose machines connected by conveyer belts. Workers only performed specific actions and assembled thereby perfectly fitting parts into cars with an unseen efficiency and effectiveness (Lean Enterprise Institute, 2006). The original Ford system (the Pure Fordism) upholds a high level of work standardization, moderate inventories, small buffers, and workers had a narrow span of control and did not work together. It is a rigid but very lean system (Krafcik, 1988).

In 1902, the Toyota Motor Corporation (originally called Toyoda Group) was founded as a producer of automatic looms, to be transformed into a car manufacturer in 1937 (Becker, 1998). Toyota was aware of the position that Japan was in. Since the country lacks natural resources, it has to work more efficiently to compete with other countries (Sugimori, Kusunoki, Cho, & Uchikawa, 1977). Different Toyota managers spend years in the United States to learn from Ford's operations, ending with the birth of the Toyota Production System (TPS) of continuous improvement (Kaizen); the basis of Lean manufacturing (Becker, 1998). Like Fordism, the TPS has a high level of standardization, but in this case in teams. The span of control is somewhat bigger, but inventories are kept to a minimum because of the Just-in-Time (JIT) principal (Krafcik, 1988).

From the early 1980s on, many Western producers started to adopt Toyota's interpretation of Pure Fordism (Krafcik calls this stream Recent Fordism). Differences with the TPS can be found in the large buffers and inventories to reduce risks of downtime, however leading to the increase of costs (Krafcik, 1988). In 1985, Motorola engineer Bill Smith presented a paper that introduced Six Sigma (Harry, 1998). This concept aims to improve production quality to an astonishing high level.

Womack and his International Motor Vehicle Program (IMVP) carried out a $5 million research project to identify the key differences between the lean and mass way of thinking (Womack, Jones, & Roos, 1990). The findings can be summarized as follows: "North Americans and Europeans had assumed and accepted the mass production theory and honed it to perfection. Japan and Toyota had used mass production as a starting point and evolved it further to TPS" (Baluch et al., 2012, p. 853). It was Womack who started using the phrase "Lean Manufacturing" to encourage the use of Lean in all production industries. Today, it is often rephrased to "Lean Thinking" to inspire all industries (even service providers and healthcare) to use Lean.

This brief history lesson shows us the starting point of almost all modern manufacturing models and also of business process improvement.

3.1.2. Definitions

To get a better understanding of the term business process improvement methodology, it is first taken apart. The term "process" is an important concept. Over time, many scholars have given their own interpretation on this concept and many of them were formed based on different perspectives (Davenport, 1992; Hammer & Champy, 1993; Childe, Maull, & Bennett, 1994). Two popular definitions of process are:

"Any activity or group of activities that takes an input, adds value to it, and provides an output to an internal or external customer. Processes use an organization's resources to provide definitive results" (Harrington, 1991, p. 9).

And:

"A lateral or horizontal organizational form, that encapsulates the interdependence of tasks, roles and people, departments and functions required to provide a customer with a product or service" (Earl, 1994, p. 13).

When used for commercial organisations, the term may need some more direction, thus the term "business process" has been conceived. Again, numerous scholars came with different interpretations. Harrington supplements his own definition to the following:

"All service processes and processes that support production processes. A business process consists of a group of logically related tasks that use the resources of the organization to provide defined results in support of the organization's objectives" (Harrington, 1991, p. 9).

Checkland defines a “methodology” as “a collection of problem-solving methods governed by a set of principles and a common philosophy for solving targeted problems” (as cited in Kettinger, Teng, & Guha, 1997, p. 58).

With previous mentioned definitions in mind, business process improvement methodology (BPI) is defined as:

“A collection of problem-solving methods, governed by a set of principles and a common philosophy, with the collective goal to improve another set of logically related tasks that use resources of the organization to provide defined results that support the organization's objectives.”

In this paper, the term “business process improvement methodology” is used as a collective term for all different methods available. However, since 1995, business process improvement is also used as a concept on its own, different from the term explained previously. Harrington, Esseling, and van Nimwegen define “their” business process improvement methodology as:

“A methodology that is designed to bring about step-function improvements in administrative and support processes using approaches such as process benchmarking, process redesign, and process reengineering” (as cited in Adesola & Baines, 2005, p. 39*).*

This concept is first described by Harrington in 1991 as a reaction on business process reengineering, which is a more radical methodology aimed to examine business process from a “clean sheet” perspective and then to reconstruct the complete process. The creator, Hammer (1990), calls it “an all-or-nothing proposition with an uncertain result (p. 105)”. Both methodologies will be further explained in Section 3.1.9.

On top of the different improvement methods lays the concept of business process management. Next to the representation of business processes, it also includes “concepts, methods, and techniques to support the design, administration, configuration, enactment, and analysis of business processes (Weske, 2012, p. 5).” Additional benefits can be achieved with the use of software systems, called “business process management systems”. “[These] generic software systems explicit process representations to coordinate the enactment of business processes” (Weske, 2012, p. 6). For the next section, eight BPIs will be discussed.

3.1.3. Lean

The aim of Lean is to increase productivity by reducing operating costs through the elimination of all waste. In this context, waste is everything that does not add value to the product or service (Womack & Jones, 1996; Monden, 1983, as cited in Baluch et al., 2012, p. 486; Smith & Hawkins, 2004). Thomas, Barton, & Chuke-Okafor use the credo “do more with less” (2009, p. 114). There is a simple test to determine if an activity adds value. Namely, the activity needs to comply with three conditions (Sayer & Williams, 2012):

1. The activity must transform the product/service.
2. The customer must be willing to pay for it.
3. It must be done correctly the first time.

Besides value-adding activities, the Lean methodology points out seven deadly wastes, namely: overproduction, waiting, transport, extra processing, inventory, motion, and defects (Hicks, 2007). These types should be eliminated as quickly as possible.

A more structured method is given by Womack, Jones, and Roos (1990). They distilled the following five principles from the Lean philosophy:

1. Specify the value desired by the customer.

2. Identify the value stream for each product providing that value and challenge all of the wasted steps (generally nine out of ten) currently necessary to provide it.
3. Make the product flow continuously through the remaining value-added steps.
4. Introduce pull between all steps where continuous flow is possible.
5. Manage toward perfection so the number of steps and the amount of time and information needed to serve the customer continually falls.

The pros of Lean are the reduction of costs and time needed for the production, logistic, and maintenance processes. On the other hand, Lean can be seen as a threat by employees. Frustration can arise when more work is needed to be done with less resources. Effective human resource management is then key in this process.

3.1.4. Six Sigma

Where Lean finds its origin in the car industry, Six Sigma (6σ) starts in the world of communication and electronics. In order to challenge the Japanese, the American technology company Motorola raised the bar of its production quality, demanding a success rate of 99.99966%, or 3.4 defects per million opportunities (Sester, 2001). Their approach was called Six Sigma (named after the sixth standard deviations from a normal distribution). At its foundation "Six Sigma recognises that there is a direct correlation between the number of product defects, wasted operating costs, and the level of customer satisfaction" (Harry, 1998, p. 60). Six Sigma can be seen as both a business strategy and a science to reduce costs while improving customer satisfaction (Thomas, et al., 2009). Traditionally, the five-phased DMAIC methodology is applied. These phases are:

1. Define. Who are the customers and what are their priorities? Where lie their problems? Which do we tackle first?
2. Measure. How is the process measured and how is it performing? What is its current state of performance?
3. Analyse. What are the most important causes of performance failure?
4. Improve. How do we remove the causes of poor performance?
5. Control. How can we embed and maintain the improvements made?

Douglas Ferguson (2007) defines Lean as a philosophy while describing Six Sigma as a dynamic, problem-solving program. Six Sigma is a data-driven approach. Every step provides data to measure status quo and improvements. The Sigma Rating reflects the level of quality at each quality aspect (Al-Aomar, Aljeneibi, & Almazroui, 2016). Improvements are based on Design of Experiments (DoE). DoE is a statistical approach to optimize processes by changing certain parameters and measure the change in results. Its power lies in reducing the number of experiments, while providing statistically reliable results (Pawlak, Rosienkiewicz, & Chlebus, 2017).

One of the key advantages of Six Sigma is its statistical base, making it possible to analyse the whole cycle. A serious difficulty is the fact that it is so tightly defined. Work is carried out in teams and all members need to be committed to Six Sigma and trained in the process.

3.1.5. Lean Six Sigma

The concepts of Lean can be adapted to comply with the need of different industries and situations. One of these bifurcations is the combination of Lean with Six Sigma. Thomas et al. (2009) describe how this integrated approach can create synergy. They combine DMAIC with components from the Lean toolbox and place them in a 10-step approach:

1. Define. What is the problem? Does it exist?
2. Measure. How is the process measured? How is it performing?
3. Analyse. What are the most important causes of defects?
4. Improve. How do we remove the causes of the defects?

5. Control. How can we maintain the improvements?
6. Implement 5S technique.
7. Application of value stream mapping (VSM).
8. Redesign to remove waste and improve value stream.
9. Redesign the manufacturing system to achieve single unit flow (SUF).
10. Apply total productive maintenance (TPM) to support manufacturing functions.

The goal of this fused Lean Six Sigma is to maximize performance by achieving the fastest rate of improvement in customer satisfaction, cost, quality, process speed, and flexibility as possible (de Jong, Wouter, & Van Blokland, 2016). While Lean is a philosophy carried out by the whole organisation, Six Sigma is a team-based program. Lean Six Sigma aims to combine the best of two worlds (Thomas, et al., 2009). The entire organisation is involved in the waste-elimination and also, a culture of learning and continuous improvement is created. However, it is not expected that the whole organisation gets the same profound training than that the Six Sigma teams are getting, since these trainings are quite expensive. The theory around Lean Six Sigma exists for less than two decades and the literature gives few downsides of its application.

3.1.6. Total Quality Management

Total quality management (TQM) can be defined as the goal to accomplish total quality by involving everyone's daily commitment (Kanji, 1990). With total quality Kanji means the achievement of satisfying customers' requirements continually at low costs. TQM does not have strict specifications or steps (Holmes, 1992). It can be seen as a toolkit with a specific goal and the underlying philosophy of continuous improvement, focusing on processes, measuring performance, and involving and empowering employees and several techniques (as quality training, statistical process control, just in time) (Gershon, 2010). At the core of continuous improvement lays the Deming wheel. This wheel has four stages: plan (define the process and set measurable objectives), do (execute the process and collect information), check (analyse the information), and act (execute corrective actions using TQM techniques and assess future plans). After the act stage, the process is either standardized or the new targets are defined and the cycle continues (Kanji, 1990).

When comparing TQM with Six Sigma, Gershon (2010) comes to the conclusion that Six Sigma includes all tools and philosophies of TQM. Gershon continuous that Six Sigma could be seen as the improved successor of TQM. TQM only succeeds when management participates and when group work is required enough while not having the prescriptive methodology for its implementation like Six Sigma does.

3.1.7. Lean MRO

The concept of maintenance, repair, and overhaul (MRO) can be defined as "all actions which have the objective of retaining or restoring an item during its life cycle in which it can perform its required function" (de Jong et al., 2016, p. 36). MRO is not a method, but an industry or part of a process. In order to execute MRO operations efficiently and exactly, researchers have developed many technologies and methods (Zhang, Liu, Jiang, & Chen, 2015). Besides all technical (maintenance) actions, MRO also includes corresponding administrative, managerial, and supervision actions. This sets MRO apart from the manufacturing industry. MRO is made up of both service-orientated as well as production-orientated functions (Al-kaabi, Potter, & Naim, 2007). The term MRO receives particular interest in the aviation industry, because of their high level of regulation and certification of processes, assets, and components.

Zhang et al. (2015) come up with another variant of Lean: Lean MRO. They define Lean MRO as "a systemic waste minimized method for the MRO stakeholder to take a closed-loop product lifecycle information management approach" (pp. 040908-3). This variant is needed, since the production process of an MRO organisation is different from that of a regular manufacturing enterprise. Next to different actions (e.g. fault detection, product disassembly), the process involves different stakeholders at different stages of the product life-cycle. It is important that each MRO stakeholder provides sufficient information. Lean MRO handles the same philosophy as "regular" Lean, but with the distinction that it needs to address both production and service-orientated activities. According to the limited available literature on this topic, successful implementation is scarce (Ayeni, Ball, & Baines, 2016).

3.1.8. Total Productive Maintenance

In his book "TPM Development Program: Implementing Total Productive Maintenance", Nakajima (1989) explains the philosophy of Total Productive Maintenance. Baluch et al. (2012) summarise it as follows: "TPM is an innovative approach to maintenance that optimises equipment effectiveness, eliminates breakdowns, and promotes autonomous maintenance by operators through day-to-day activities involving the total workforce" (p. 851). Bakri, Rahim, Yusof, & Ahmad (2012) add that "TPM is the productive maintenance carried out by all employees through small group activities and can be viewed as equipment maintenance performed on a company-broad basis" (p. 486). Riss, James, & Thorsteinsson (1997) use the same explanation, but add the need to "develop a preventive maintenance programme for the life-cycle of the equipment" (p. 354). 3.1.8. Total Productive Maintenance (TPM) can be seen as a philosophy where the operators of a piece of machine treat it as if it were their own and not blindly rely on the fact that the maintenance crew will fix it when the operators brake it. When operators take good care of their assets, overall equipment effectiveness (OEE) will increase and cost will decrease. From a management point of view, it is important to know who creates value in the organisation and give them the necessary support. This can range from tools for a mechanic, to training and the ability to come up with and realize improvements.

The main advantage of TPM is the reduction in delays and downtime by reducing breakdowns and equipment failures. However, TPM is difficult to implement since it requires the cultivation of ownership and awareness (Baluch et al., 2012). Next to that, the training of personnel can lead to high initial costs.

3.1.9. Business Process Reengineering and Business Process Improvement

There is a lot of ambiguity about the many business improvement methods and in particular the concepts of Business Process Reengineering and Improvement. Many management websites and even academic literature use the terms as they seem fit, neglecting the conceived structure.

Business process reengineering (sometimes called business process redesign) tend to radically redesign business processes. Processes are modelled to help the organization understand how these are acting. Modern information technology is used to achieve dramatic performance improvements. Breaking away from conventional wisdom and the constraints to enable a new process instead of automating an existing one (Hammer, 1990).

Business process improvement is a less drastic and disruptive, and more incremental approach. Harrington (1991) describes in his book that this approach tries to simplify the process by reducing the number of process sequences.

3.1.10. Overview

Currently, eight different BPIs have been discussed. However, this is only the tip of the iceberg. Scholars and management consultants frequently come up with new BPIs and combine them till distinction is long gone. For example, next to Lean Six Sigma and Lean MRO the internet shows us that the variations Lean Startup, Lean Manufacturing, Lean Management, Lean Thinking, Lean Enterprise, and Lean Maintenance also are being used. These variations could be completely new methodologies, slightly adjusted methodologies, or identical to Lean as described in this chapter.

As a guideline, an overview is made of the BPIs discussed in this chapter and is shown in Table A.3.1 (of the appendix). This overview is based on the work of Andersson, Eriksson, & Torstensson (2006), who described three BPIs (Lean, Six Sigma, TQM) on nine properties. Their structure is general enough to explain more BPI's. Based on all information in this chapter and previous mentioned sources, the overview is supplemented with the five discussed BPIs. (The last column concerns process mining and will be discussed in the next section.)

3.1.11. Conclusion

In this paragraph, the term Business Process Improvement methodology (BPI) is defined as a collection of problem-solving methods, governed by a set of principles and a common philosophy, with the collective goal to improve another set of logically related tasks that use resources of the organization to provide defined results that support the organization's objectives. Almost all BPIs find their origin in Lean and while throughout the years many different BPIs arose, the boundaries between these BPIs remain vague. Different BPIs have been combined to form new BPIs and users often interchange the terms. As a science, this makes it difficult to investigate their characteristics. An overview is presented to order eight widely used BPIs on nine properties. This overview contributes to the knowledge on BPIs. The thesis will continue discussing process mining (first as an analysis methodology and later as a BPI). In Section 3.2.6, a supplement concerning process mining is given to the overview in Table A.3.1.

3.2. Process mining: a novel business process improvement methodology

From an abstract viewpoint business processes are comparable to the chains of an industrial production process (e.g. Henry Ford's production lines). They all follow a predetermined logical workflow and use resources of the organization to achieve a predefined business outcome (Bose, van der Aalst, Žliobaitė, & Pechenizkiy, 2014). For an organization (service or production orientated) to become effective and efficient, it is important to streamline these processes. However, because of globalization, digitalization, and the growth and diversification of the workforce of multinational corporations, business processes are getting more complex and keeping an overview is difficult. With the relatively novel methodology of process mining, a new instrument becomes available to discover, analyse, and enhance formal process models of an information system (van der Aalst, 2016). The research and advisory firm Gartner positions Automated Business Process Discovery (as it calls process mining) on the Peak of Inflated Expectations, predicting it will reach the Plateau of Productivity in two to five years, confirming its potential (Gartner, Inc., 2016).

The second part of this chapter starts with addressing process mining as an analysis methodology. After an introduction, several key elements of process mining are discussed, viz. event logs, miners, visualisations, and software. Analysis by itself do not lead to progress, that is why a comprehensive project methodology is essential. Three process mining methodologies are discussed. After that, case studies that use process mining are examined. Consideration is given to the nature of the industry it is used in and the results of the case studies. A set of practical pros and cons is given and thereafter, the overview of traditional BPIs in Section 3.1.10 is supplemented with the process mining BPI: PM^2. Before the conclusion, a prediction on the future of process mining is given. This part contributes to the knowledge of process mining, by giving a clear explanation on the topic. MoD requires this to train its own employees and to assess the usefulness of process mining. Placing a process mining improvement methodology next to the traditional BPIs is not yet found in the literature. This makes Table A3.1 a valuable contribution.

3.2.1. Introduction to process mining as an analysis methodology

Process mining aims at building a bridge between data mining and model based process analysis (van der Aalst, 2016). Data mining is commonly defined as: "the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner" (Hand, Mannila, & Smyth, 2001, p. 1). Data mining uses statistics, databases, and algorithms to create its value (just like process mining). And is (also) relatively novel because of the recent emergence of cheap computing power and digital storage space. Model based process analysis aims at optimizing decision-making regarding which activities need to be executed and in which order. Models are used to perform the analysis, capture the results and communicate.

A first work published on process mining comes from Cook & Wolf (1995). In their work, they describe the semi-automatic generation of formal process models from process execution data to improve software development. Three years later, Argawal, Gunopulos, & Leymann (1998) and Datta (1998) independently published a paper. The first article uses logs from an IBM Flowmark installation to construct process models, the second explains that having an AS-IS model is a prerequisite for applying Business Process Reengineering and connects the BPI with the automated creation of models by applying algorithms. In a since, this combination was the first attempt to use process mining as a BPI. In the years following, research was done on various topics relating to process mining (Herbst, 2000; Weijters & van der Aalst, 2003; Hand et al., 2001). It was in the year 2004 that process mining got its current form. Van der Aalst & Weijters (2004) challenged the academic world to help solving the most urgent scientific questions by publishing a research agenda. The academic world responded and the amount of publications grew steadily.

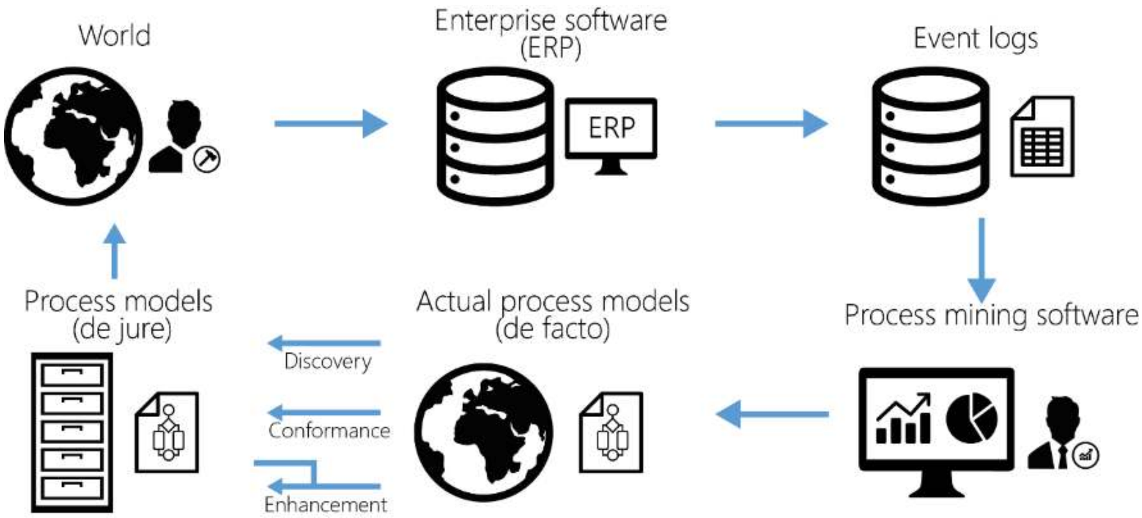


Figure 3.1: Positioning of the three main types of process mining (based on van der Aalst, 2016, p. 32)

An overview of the process mining landscape is presented in Figure 3.1. Here it is shown that mining processes is a cyclic practice. In many cases, the cycle starts with a **process model**. Large organisations, just like MoD, secure their business processes in formal process models. These models describe the activities that employees need to perform and how software supports them herein. This is further discussed in Section 5.2. An example of part of a real process model of MoD is given in Figure A1.1. The model is designed in ARIS (just like all process models of MoD) and party describes the counting of supply and inventory. For the following section, a fictional example is used. This process model, also designed in ARIS, can be seen in Figure A3.1 and describes the checking, picking, and shipping of an item.

Business activities are performed in the **real world** and are guided and registered by **enterprise software**. MoD uses the ERP (Enterprise Resource Planning) software package: SAP, which is used by many organisations worldwide (SAP SE, n.d.). These activities include the receiving, reporting, examining, enhancing, and dispatching of (digital) resources. Information about these activities is recorded in a dataset, usually called an **event log** or audit trail. The event log contains information in the form of cases and activities (van der Aalst, et al., 2007). The **case** (also called process instance) is the “thing” being handled. This can be the ordering of an item at an online shop, repairing a specific vehicle, the processing of an invoice, etc. Each case has a unique **case ID**. This ID is used to link the different **activities** (also named tasks, operations, actions, or work items) to a single case. Activities are accompanied by a **performer ID** (also known as an actor ID) and a **timestamp**. The case ID, activity, performer ID, and timestamp are attributes. Many more attributes are possible, like the cost of the activity, overseeing manager, item to be shipped, etc. Table 3.1 shows a brief example of an event log (based on the model as seen in Figure A3.1). Here, Adam and Bob are responsible for checking the availability of an item. If the item is not available and cannot be ordered, they can look for an alternative. Carol is responsible for ordering the item at a different warehouse if it is not in stock and Dave is in charge of shipment.

Table 3.1: An example of an event log

Case ID	Activity	Performer	Timestamp
Case 1	Check availability	Adam	2-3-2017 9:01
Case 2	Check availability	Bob	2-3-2017 9:03
Case 2	Look for alternative item	Adam	2-3-2017 9:05
Case 3	Check availability	Bob	2-3-2017 9:08
Case 2	Check availability	Adam	2-3-2017 9:10
Case 2	Order item	Carol	2-3-2017 10:03
Case 3	Collect and ship item	Dave	2-3-2017 11:02
Case 2	Collect and ship item	Dave	2-3-2017 11:08
Case 1	Collect and ship item	Dave	2-3-2017 11:14

Process mining software is then used to (re)construct a (visual) model based on the observed behaviour in the event log. This is done by the process of mining, the execution of a set of computerized algorithms. There are many ways to express the event log in a (visual) model. Common models use the modelling languages Petri Net and BPMN. Figure 3.2 shows the outcome of mining the event log of Table 3.1 in Petri Net, Figure 3.3 shows the same model in BPMN. More models are discussed in Section 3.2.2.

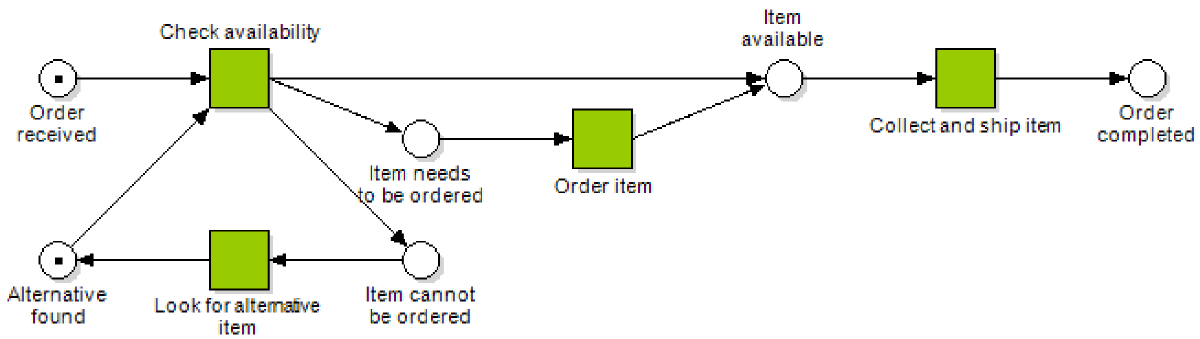


Figure 3.2: Visualization of the example case in Petri Net

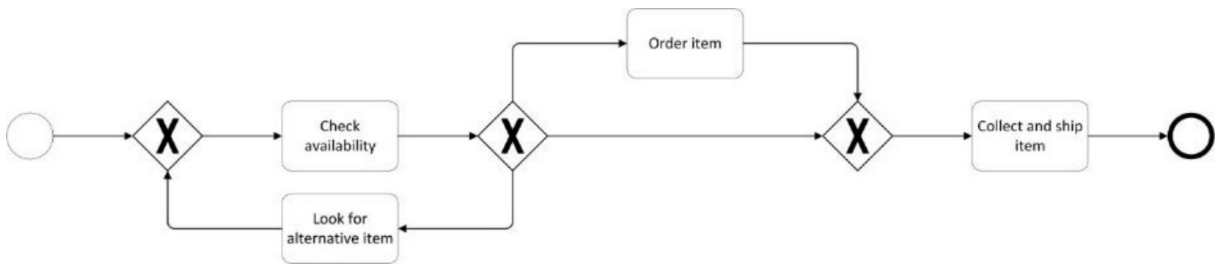


Figure 3.3: Visualization of the example case in BPMN

The example shows a simple process and it can be assumed that the activities performed perfectly “fit” the theoretical process. In larger processes, deviations are more common. For example, Carol may have ordered an item while it already was available. Process mining can handle the splitting and converging of activities, as well as loops. The strength of process mining lays in its ability to handle large datasets, while producing clear **models as they actual occur**. The example in Table 3.1 contains three cases and nine activities, but real-life examples may include hundreds of thousands of cases.

The mining of processes does not automatically lead to better processes. It visualises them, acting as a starting point for further analysis and recommendations. As shown in Figure 3.1, process mining can be used for three purposes: the discovery, conformance checking, and enhancement of models. When the organisation has not formally modelled its processes already, process mining will lead to the discovery of models. Bottlenecks and unexpected pathways can be found and adjusted if necessary. For MoD, this is seldom required, since it has modelled most of its processes already. When theoretical models are present, the conformance can be checked, measuring the alignment between reality and model. Deviations from the theoretical model do not have to be a problem and can prove to work even better (in that case the theoretical model can be adjusted towards the practical model). Finally, there is the enhancement of models, where these models are used to extend or improve the process model (van der Aalst, 2016). Process mining is more than just an analysing methodology, but it is not (yet) a matured full-scale business process improvement methodology. A certain procedure (as described in Section 3.1) is needed to come to improvements. Current process mining methodologies will mature and perhaps be combined with traditional business process improvement methodologies.

An analysis in process mining is characterized by a significant amount of liberty of the researcher. No standard exercise exists that can be applied on all event logs and process models. Decisions need to be made about the scope of the data, the choice of miner and its configuration, the mode of visualization, and the interpretation of the outcome. Each miner (or mining algorithm) and modelling language has its own pros and cons. And the researcher can make his decisions based on personal taste, experience, and literature (the last may not be adequate). Van der Aalst (2016) discerns four quality criteria to be taken into account during the actual mining, namely: fitness (do all cases from the event log fit in the discovered model?), precision (is there any unrelated behaviour modelled?), generalization (does the model show overfitting the log?), and simplicity (how simple is the discovered model?).

3.2.2. Key elements of process mining as an analysis methodology

After briefly explaining the general idea of process mining, this section then focusses on several key elements of process mining, viz. event logs, minders, visualisations, and software.

Event logs

Data (in the form of event logs) are the fuel for a process mining project and will make the difference between success and failure. The concepts of process mining are not software or platform bound and till November 2016 there was no standard structure for formulating event logs.

Several software developers developed adapters to extract data from the different ERP (Enterprise Resource Planning), SCM (Supply Chain Management), CRM (Customer Relationship Management), and other BI (Business Intelligent) software packages to create comprehensible logs for their process mining software. For example, before processes can be mined in ARIS PPM (the process mining software used by MoD), an adapter needs to be created and customized to link the data from the ERP package to the ARIS process miner. Customization takes place based on the used software package and the processes that need to be analysed (Scheer, Abolhassan, Jost, & Kirchmer, 2004). This partly is a manual process and developers do not support all enterprise software packages by default. For example, Celonis supports software packages like SAP, OTRS, Oracle, Microsoft Dynamics, and Salesforce (Celonis SE, n.d.). An advantage of using an adapter is that data can be used for online (real time) analyses.

Alternatively, event logs can be extracted manually from the enterprise software package and imported into the process miner. (More of this is described in Section A.3.1 of the appendix.) Since November 2016 there is a standard to structure event logs. The XES standard was adopted by the *IEEE Task Force on Process Mining* and finally approved by the *IEEE Standards Organization* as IEEE Std 1849-2016 (IEEE Computational Intelligence Society, 2016). XES is the preferred format for some process miners like ProM. Other formats used are MXML and CSV. Depending on the mining software, these formats can either be imported directly or need to be converted beforehand with software like XESame (van der Aalst, 2016; van Dongen & van der Aalst, 2005).

Bose, Mans, & van der Aalst (2013) remark that in the future "process mining will not be limited by the availability of data, but by the quality of event data" (p. 21). Stressing that data quality has a significant impact on the success of a process mining project. Bose et al. identified ten categories of data-related problems, encouraging the development of systematic logging approaches and repair techniques. Some of these are mentioned in the book Process Mining Techniques in Business Environments (Burattin, 2015). Burattin also discusses the implementation of mining data streams (real-time, continuous, ordered sequence of items). Suriadi, Andrews, Hofstede, & Wynn (2017) discuss several commonly encountered problems and solutions concerning event log quality.

Even with high quality data, mining results can give a distorted picture. The effects of *concept drift* occur when during the period of data acquisition the actual process changes. For example, a dataset of two years is analysed. During the first year process A is followed. At the end of that year, management changed the procedures and for the next year process B is followed. Than the mined process could look like the "average" of A and B (from now on called C). In reality C never occurred. It can be difficult to pinpoint the moment when processes are changed, and apart from that, process mining is a data hungry practise, meaning that mining results depend on the amount of data entered. Limiting the dataset to the last day, thus ensuring its actuality, will not give enough data to mine.

Even though van der Aalst et al. (2012, p. 14) claim that "understanding concept drift is of prime importance for the management of processes", concept drift has been discussed in only a few papers. Bose, van der Aalst, Žliobaitė, & Pechenizkiy (2011) categorized four classes of concept drift: sudden, recurring, gradual, and incremental. They only developed sets of techniques to identify sudden drift, their approach led to the development of a ProM plug-in to identify sudden drift. Weber, Bordbar, & Tiňo (2011) proof that, with an optimal amount of data, statistic methods can detect changes. The following years, different researchers developed various methods to tackle parts of the problem. Carmona & Gavaldà (2012) present a mechanism for online detecting and managing concept drift, based on abstract interpretation and sequential sampling. These methods could only detect abrupt changes. Fei, Liqun, GuangYun, & Xiaolei (2013) developed an algorithm based on context detection. Luengo & Sepulveda (2012) and Hompes, Buijs, van der Aalst, Dixit, & Buusman (2017) use clustering to deal with concept drift. Three years after their original paper, Bose, van der Aalst, Žliobaitė, & Pechenizkiy (2014) published a revision, concluding that there are still many challenges to be addressed.

Miners

At the heart of a processing mining project lays the process mining algorithm. This algorithm, often embedded in the process mining software, translates the data from event logs into readable models. There are several algorithms available, each having its own properties concerning the form of input, conversion of data, and form of output. One must pick the right algorithm for a dataset for the right goal and right way of visualisation. Much has been written about the mining algorithms.

The first miner developed was the Alpha-algorithm (van der Aalst, Weijters, & Maruster, 2004). This algorithm is based on eight simple mathematical definitions and visualises its models in the Petri Net modelling language. Because of its simplicity, it is popular among scholars, but it is unpractical in real-life, because of its difficulty in handling noise, infrequent/incomplete behaviour, and complex routing constructs (van der Aalst, 2016). A second miner, the Heuristic miner (previously called Little Thumb), is better equipped to handle complex routing and it can abstract exceptional behaviour and noise, making it suitable for actual logs (Weijters & van der Aalst, 2003). The Fuzzy miner focusses on unstructured behaviour and large event logs. Its output is configurable to reach a desired level of abstraction, but can only be visualized in a fuzzy model (Günther & van der Aalst, 2007). Another interesting algorithm is the Multi-phase miner. It was inspired by the Event-driven Process Chains (EPCs, used by ARIS) and is useful for relatively simple processes (van Dongen & van der Aalst, 2004).

Visualisations

Depending on the goal of the analysis and on the analyst's personal taste, several ways of process visualisation can be used. Most common are the visualisations of a process model in Petri Net (seen in Figure 3.2) and BPMN (seen in Figure 3.3). However, not all software uses standardized modelling languages. Disco, for example, uses a self-developed language. An example of a process model mined with Disco is shown in Figure A.3.2. Next to the "classic" visualisations, focus can be placed on different aspects of the process. For example, the actors involved in the process. This is done with a social network, as shown in Figure A.3.3. A social network analysis shows the handover of work and can, for example, be used to determine the importance of an actor in the process (do all instances of a process go past a single, and thus indispensable, person?). Another visualisation is the dotted chart (shown in Figure A.3.4). The chart consists out of a number of horizontal lines. Every line represents a single process instance (case) and its activities are visualised by dots on that line. The type of activity is distinguished by the colour of the dot. The x-axis represents the time. With this visualisation, it is simple to see what activity was carried out at what moment. Many more modes of visualisation exist.

Software

Several process mining software packages are available. Choice can be based on a specific needed set of functionalities, supported data formats, but also costs. Finding the right software can be difficult, since no comprehensive comparison exists. Several scholars dedicated their Master Thesis to this subject, but the wide range of software vendors and the pace of new software being offered and updates being released, made their results almost immediately outdated (Ailenei, 2011; Verstraete, 2014; Kebede, 2015). Table A3.2 shows a list of process mining software. The list is compiled after extensive online research and contains most of the currently supported software. To give an idea of its similarities and differences, three leading packages will be discussed. The websites of the software developers have been used as the source of the comparison.

Overall, three earnings models can be distinguished. First the common licencing structure where an organisation can buy the software for a certain period or indefinitely. The software can be sold in combination with or without support for implementation or analysis. This is the case with, for example, ARIS PPM, Celonis Process Mining, and Disco. Another earnings model is offering process mining as a service (PMaaS). This is provided by Icris and Coney. And lastly, several open source tools are available. Most famous example is ProM, but also Apromore is popular.

Based on the number of academic publications on the topic, open source process mining platform ProM seems to be the most popular (as shown in Figure A.3.5 and Figure A.3.6). ProM is an extensible framework that runs on Java and obtains its functionality by a wide variety of plug-ins. Because it is an independent platform and is developed by process mining "godfather" Will van der Aalst, it is popular among scholars performing applied research. There are over 1500 free plug-ins available, each with different functionalities and options (van der Aalst, 2016). For example, the use of different miners (heuristic, alpha-algorithm, and fuzzy), sorts of output (Petri Net and BPMN), and types of analysis (process discovery, dotted chart, and social networks). However, the academic character makes it difficult to use. Manuals and instructions are missing and support can only be found in its community of volunteers. It seems that only few commercial organizations use this software and even if they do so, they mostly use it to learn the concept of process mining before buying more user-friendly software.

A more user-friendly process mining software package is Disco (as shown in Figure A.3.7 and Figure A.3.8). Developed by Fluxicon (led by two former PhD candidates of Will van der Aalst), Disco lacks some functionalities when compared to ProM, but distinguishes itself with a fast, well-documented, and clear interface. Very limited knowledge on process mining is required to perform an analysis. But the lack of real-time connections to databases makes it less useful for large companies. It seems logical that Fluxicon focusses on small and medium-sized enterprises.

A third software package, used by MoD, is ARIS Process Performance Manager (as shown in Appendix Figure A.3.9 and Figure A.3.10). This package contains three components: the administrator's section, the business analyst's section, and the dashboard. The section for the administrator is used to load the data. This can be a single file, but also a connection to a database. A good example is SAP. With the right connector, the SAP databases can be periodically and automatically loaded into the ARIS databases. Initializing this connector will cost some effort, but can be a good investment. The business analyst's section of the software is used for in-depth analysis. With the use of filters, selections of the dataset can be made. Analysts can use several techniques and models to answer their process related questions. The steps of this process (the query) can be saved, so when the data are refreshed, the analysis can be updated. The last section, the dashboard, is meant for tracking the organisation's process performance. The analyst can develop certain queries. For example, the average lead time of preventive maintenance of a specific weapon system, and this can be loaded into the process-centric dashboard as a key performance indicator (KPI). The dashboard periodically collects recent data from the ERP and the process owner can follow the progression. Based on the average process and on its excesses, the process owner can decide to intervene.

Kebede (2015) developed a model to compare ProM, Disco, and Celonis on fifteen characteristics. The model was updated to the lasted software versions and Celonis was replaced for ARIS PPM. This can be seen in Table A3.3; an explanation is provided in Section A.3.2 of the appendix.

3.2.3. Process mining to improve business processes

Section 3.1 discussed several BPIs and Section 3.2.1 introduces process mining as a business analysis methodology. At this point it is not possible to compare these two, since process mining only discovers problems and areas with improvement potential, while a BPI also aims to address problems and improve the process. To apply process mining and successfully improve business operations, a fitting project methodology is fundamental.

The first methodology that addresses process mining was developed by Bozkaya, Gabriels, & van der Werf (2009). Their Process Diagnostics Method (PDM) consist of six steps: log preparation, log inspection, control flow analysis, performance analysis, role analysis, and transferring the results. Each step is accompanied by understandable instructions and aims to give a broad process overview within a brief period without any prior and domain specific knowledge required. Rebuge & Ferreira (2012) claim that existing process mining methodologies do not perform well in the complex and ad hoc healthcare environment. They continued to build on the work of Bozkaya et al. (2009) and added a sequence clustering analysis between the earlier proposed log inspection and control flow analysis. They successfully analysed the careflows of emergency patients in a Portuguese hospital.

Van der Aalst (2011) designed a new methodology for “lasagne processes” (structured processes), containing five stages: plan and justify; extract; create control-flow model and connect event log; create integrated process model; and operational support. This L* life-cycle model can also be applied on “spaghetti processes” (unstructured processes), but only after another sub-process of simplification. Since van der Aalst enjoys significant respect regarding the topic, the L* life-cycle model is frequently mentioned in the literature.

To better address lasagne processes, van Eck, Leemans, & van der Aalst (2015) propose PM², as presented in Figure 3.4 and discussed in Chapter 4. Their six-step methodology (consisting of planning, extraction, data processing, mining & analysis, evaluation, and process improvement & support) is a more project based approach, deliberately starting with setting a good research question. Unfortunately, the process improvement & support step seems to require more elaboration. The authors suggest that other BPIs can be used to carry out the actual implementation of improvements, but the integration of PM² with other BPIs is hardly mentioned. Still, every author on the topic emphasises the fact that no methodology is perfect and that they act as high-level guidelines instead of a clear step-by-step cookbook.

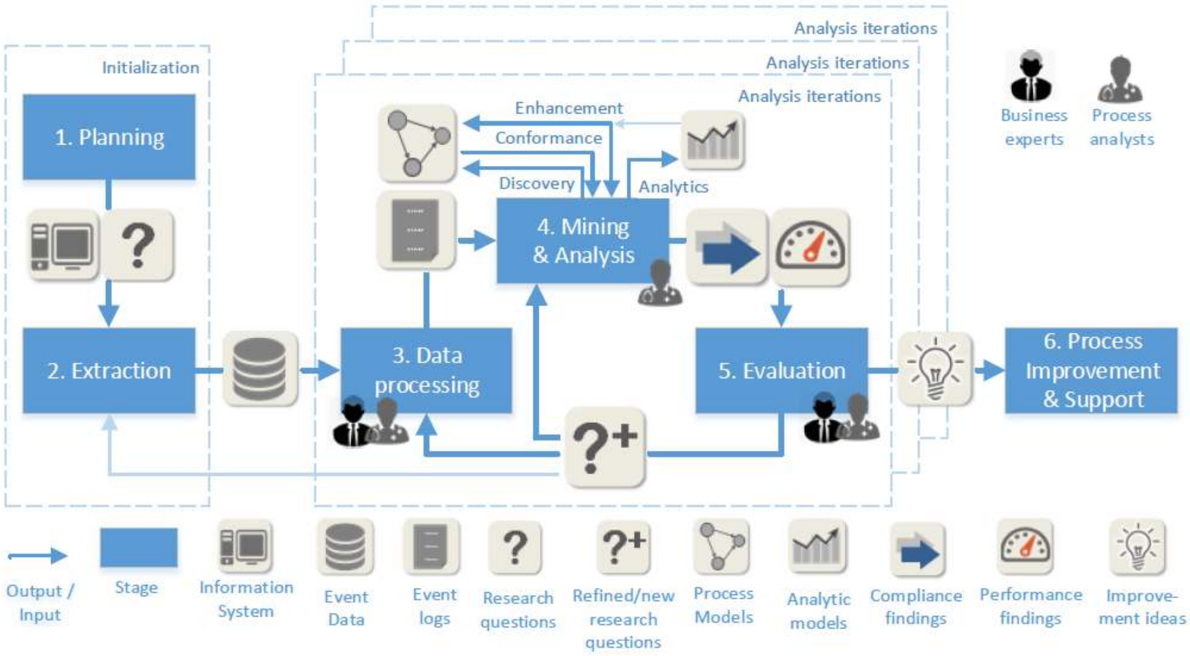


Figure 3.4: The PM² model (van Eck et al., 2015)

3.2.4. Case studies

Management consultant HSPI recently published an overview of 113 process mining projects (Gonella, 2016). Figure 3.5 shows a breakdown of projects categorised by industry. It is interesting to see that process mining is mostly used by service oriented organisations (including healthcare) and not so much in the manufacturing industry. This observation is also seen in other studies described in the literature. A reason can be that data produced in the service industry is of higher quality and granularity. It is also possible that processes in service industry have a relatively short lead time, where construction and maintenance processes take longer. A large timeframe is than needed to ensure having enough complete processes to mine. Why process mining is less popular in the manufacturing industry is remarkable. Mass production, for example, is characterized by its strict process models. Perhaps this industry did not discover the potential of process mining yet. More possible reasons will be discussed in Section 5.2.3.

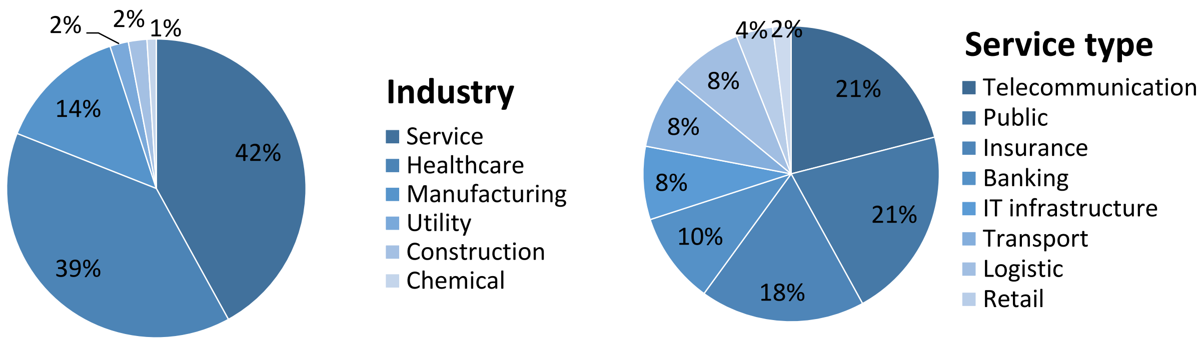


Figure 3.5: 113 process mining projects categorised by industry (Gonella, 2016)

The large attention of process mining in healthcare is also recognized by Rojas, Munoz-Gama, Sepúlveda, & Capurro (2016). In their literature review they distinguish 74 case studies of process mining in healthcare. The paper does not go into detail regarding the results obtained in each case study, but it gives a profound overview of techniques, trends, and challenges for process mining in healthcare.

Several case studies on process mining can be found in the literature. Strikingly, very few are able to express the effects of their improvements in numbers. They all agree on the great potential that process mining brings and they also explain their recommendations that derived from their analysis, but the results are seldom put in figures as saved money, shortened lead time, or improved customer satisfaction. This is remarkable, because process mining is particular useful for doing such an analysis. First, process mining is used to measure the status quo and suggest improvements. After being implemented, the same analysis can be done again to determine its effects. Also, most literature focuses on the analytic side of process mining and neglects the process mining project methodology and the implementation of improvements at a managerial level. Business analysis is a different field of study than change management, but for better results, the two depend on each other. Here lays an opportunity for future research.

3.2.5. Known pros and cons of process mining

Three examined case studies (Kooij & Rozinat, 2016; de Weerdt, Schupp, Vanderloock, & Baesens, 2013; Manuel, 2012), together with merely all papers discussed in this chapter, are predominantly positive about process mining. They do, however, share the opinion that process mining is still a young discipline and further development is desirable.

Perhaps one of its foremost pros, is that process mining can uncover understandable models of complex situations. None of the traditional methods have this capability. These models can even be visualizing different aspects of the situation (e.g. models organized on time, showing social networks, etc.) and filter on different attributes (e.g. a single weapon system or complete organisation). A side note that has to be made, is that spaghetti processes (unstructured processes) can still be a challenge to mine. The analyst may need to simplify the model to make it understandable.

Another widely agreed positive aspect of using process mining as a business improvement methodology, is that it provides a fact driven view on a human controlled process. Recommendations that come from other BPIs are often based on assumptions and gut feelings. TQM, for example, focusses on improving customer satisfaction. It is possible to conduct a customer survey measuring the effects of a process change by TQM, but it is impossible to predict the extent of these effects in advance. With process mining it is possible to simulate and calculate effects beforehand. Because of this, changes may be easier accepted by workers.

Process mining can be carried out offline and online. Offline means that the input data are an isolated (dated) dataset and online means that there is a direct connection between the process mining software and the data source (ERP). The advantage of offline process mining, is that the dataset can easily be shared and the analysts is not bound to heavy computers. Online mining, on the other hand, gives a direct representation of the situation, but requires a better digital infrastructure. With online mining, the analysts can set up a dashboard to monitor processes in real-time.

One of the major downsides of process mining is that the analysis, and thus the BPI as a whole, is completely dependent on the availability and quality of input data. "The quality of a process mining study can only be as good as the quality of the input data" (de Weerdt, Schupp, Vanderloock, & Baesens, 2013, p. 65). Data can be missing, imprecise, incorrect, or irrelevant. Therefore, the organisation needs to have capabilities to log its activities; a digital infrastructure is required, workers need be trained to enter the data correctly, and management need to be able to explain the necessity of logging to its workers. And even then, a gap in data can make the whole set useless. Data also needs high precision. Time stamps, for example, may need to be at the second, while not all ERPs support this. Even with the best training and intentions, small mistakes are easily made and can have fast effects. While reporting on his/her activity, a worker can enter the month in the entry field of the day and vice versa, or his/her personal ID in the entry field where he should enter the costs. A single entry can then effect the whole analysis. Lastly, monitoring irrelevant data will distract the analyst, making it more difficult to come to conclusions.

A multidisciplinary team is another necessity. The team should have expertise in the process, the way the events are logged, the network, the process mining software, and how the data should be analysed and interpreted. This expertise is seldom found in a single person.

The process is also highly iterative. Depending on the complexity of the project, business analysts and process analysts need to meet several times. This does not have be a problem, but it can make progress rather slow and unpredictable. Where improvement techniques like TQM and 5S show almost direct results, process mining may lead to a dead end.

Another of its requirements, is the availability of process mining software. Most case studies mentioned here used the open source platform ProM or the commercial available Disco (which is free for academic purposes). While ProM is great for academic research and Disco for smaller organisations, larger organisations likely need a more sophisticated package that supports connections to the organisations' databases. ARIS Process Performance Manager and Celonis Process Mining are good alternatives, but come with expensive contracts. The software needs to be installed and configured. This is often done by (expensive) external consultants. And own personal needs to be educated in the use of the software.

3.2.6. Process mining compared to traditional BPIs

A comparison of several traditional BPIs has already been given in Section 3.1. Based on the results of the research on process mining, the overview in Section 3.1.10 can be supplemented with the information in Table 3.2. In order to compare process mining with the traditional BPIs, it is hypothesized that PM^2 is a mature BPI.

Table 3.2: Process mining as supplement to the overview of common BPIs

Concept	**Process mining (PM²)**
Origin	The rise of big data and accessibility of computing power
First mentioned	2015
Theory	Align de facto with de jure process models
Process view	Discovering, conforming, and enhancing business processes
Involvement	Multidisciplinary team
Methodologies	Plan, extract, process data, mine and analyse event data, evaluation, and process improvement & support in a team
Primary effects	Gain quantitative and factual knowledge about processes
Secondary effects	Improvements can be monitored and verified
Criticism	Demanding high quality data and structured processes

Almost all discussed BPIs have a statistical or analytical background, but process mining excels in its ability to automatically convert data into organized information. However, compared with the more mature BPIs, process mining lacks in its capabilities to implement improvements. The DMAIC cycle of Six Sigma, for example, shows a more structured approach for this step.

3.2.7. Future developments

In its short history, process mining has made an impressive development. Every year, more process mining software is developed, papers on the topic are published, courses in process mining are given, and case studies are conducted (Gonella, 2016). The future of process mining thus seems bright, but does it also offer new possibilities?

Since process mining is a data driven activity, and with data storage becoming cheaper and cheaper and initiatives like the Internet of Things (IoT) boosting data production, new possibilities do arise. Mechanics can, for example, enter their activity data in the ERP with wireless tablets, giving real time analysis possibilities. Combining this with artificial intelligence (AI) and machine learning (ML) can offer even bigger opportunities (Rinke, 2017). AI can find deviating process instances and even suggest improvements without human intervention. And ML can support decision-making processes. A case study shows positive results of using ML in combination with process mining during the design process of an electric torch (Es-Soufi, Yahia, & Roucoules, 2016).

Less futuristic but not less promising, is the combination of process mining and simulation models (Rozinat, Mans, Song, & van der Aalst, 2009). Process mining provides data on all activities of a process: its throughput, lead times, its delays, etc. These data can be used for building an accurate model in simulation software. By reasoning, but also by trial-and-error, elements in the simulation can be changed till the model cannot be improved anymore. The changes can then be applied in the real world. Integrating process mining software with simulation software can create significant opportunities.

3.2.8. Conclusion

This section started with Process mining as a promising business analysis methodology that uses event data to model processes on how they actually happened. There are many ways to model and visualize these processes. This methodology of choice depends on the data, goal of the process mining, and available software. After that, the aspects of process mining as a comprehensive BPI were discussed. The section started with several project methodologies. Here two of them deserve extra attention: the L* life-cycle model (because of its popularity) and PM² (because of its structure). But, like all methodologies available, these methodologies act more as high-level guidelines instead of a clear step-by-step cookbook. Chapter 4 will further elaborate on the improvement of the PM² methodology. Several case studies are discussed as well. Interestingly, many of them originate from service industry and healthcare and only few case studies mention quantifiable results. Why process mining is being used more frequent in some industries than in others will be briefly discussed in Section 5.2.3, but more research is needed. Just like on the measurable effects of process mining. Several pros and cons are mentioned, together with a supplement on the overview of common BPIs (Section 3.1.10). Process mining may not be as mature as Six Sigma and is missing a structured approach to implement improvements, but it shows potential to become a good addition to the BPI-toolbox. Especially because of its talent to uncover understandable models of complex situations. None of the traditional methods have this capability.

4. Improving the PM^2 process mining methodology

Section 3.1 discussed and compared several traditional BPIs. In Section 3.2, first focus was placed on process mining as an analysis methodology. On its own, process mining can assess processes and even suggest improvements, but a framework around it is necessary to guide the course and implementation of these improvements. Thus, the section continued on process mining as a BPI. Several project methodologies were found in the literature and summarized. PM^2 seems to be the most complete methodology currently available (van Eck et al., 2015). PM^2, for instance, has been successfully used to improve the purchasing process of the computer manufacturer IBM. However, experiences gained during the Defence broad pilot Process Mining suggest that it would be better to deviate from this methodology on at least some points. PM^2 is strongly focussed on the extraction and processing of data, while the software used by MoD (ARIS PPM) does most of this automatically. On the other hand, PM^2 hardly discusses the team dynamics, coming up with good research questions, and the implementation of improvements. Which are all vital components of applying process mining. This chapter begins with a thorough explanation of PM^2. After that, experiences gained during the pilot are used to address shortcomings and improvement potential. These are then used as the starting point of an improved methodology: PM^3. Investigating and designing a methodology will also help in determining what is required to start a process mining project. It indirectly helps with developing the decision framework as described in sub-question 6.

4.1. PM^2: a best practice methodology

The PM^2 methodology can focus on two objectives: improving performance of a business process (based on process discovery and enhancement), and checking process compliance (based on conformance checking). The PM^2 methodology follows six steps, as shown in Figure 4.1. (The complete methodology is shown in Figure 3.4.)

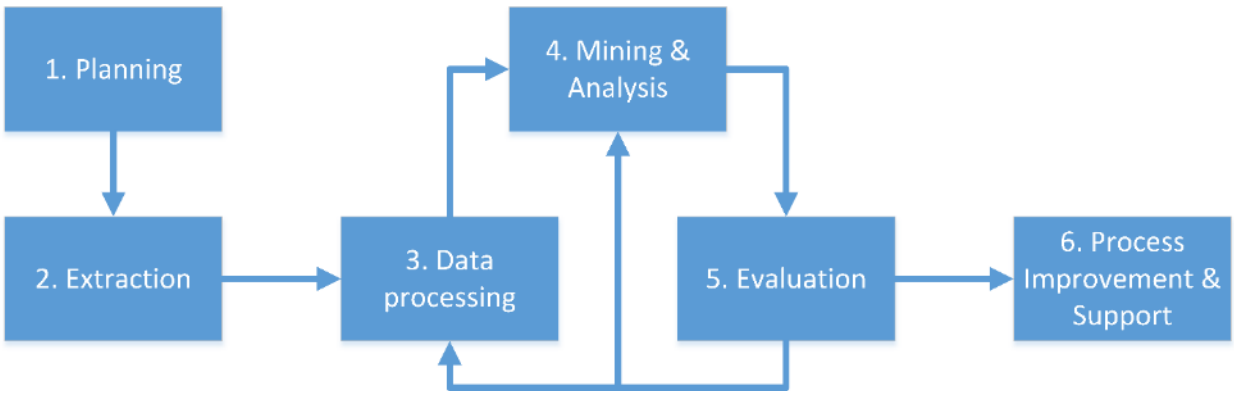


Figure 4.1: The steps of PM^2

4.1.1. Planning

During the planning phase, three activities are distinguished: selecting the business process, identifying the research questions, and composing the project team. It does not matter if the project goals are specified beforehand. This only changes the order of the activities. When **selecting the business process**, it must be certain that there is high quality event data available and that the organisation is able to implement any recommendations resulting from PM^2. The second activity is the **identification of the research questions**. These should be formulated in such a way that the event data can answer them. The more concrete the research questions are, the better the results are, but, if needed, it is possible to refine the questions in a later phase. The third activity is **composing a project team**. This team needs experts from different backgrounds, including: business owners (or process owners), business experts (or process experts), system experts (or IT experts), and process analysts (or business analysts).

4.1.2. Extraction

During the second phase, the event data are extracted. Three activities are identified. First the **scope of the data that is required needs to be determined**. Several matters should be considered. For example: the granularity of the data (does it include all related events), the time span of the collected events, the needed data attributes (like process ID, timestamp, etc.), and the correlation between data. The second activity is the actual **extraction**. This is done by the system and business experts. The last step can be executed simultaneously with the previous one. It concerns the **collection and transferring of process knowledge**. In addition, to the explicit event data, tacit process knowledge is valuable as well. This includes, for example, the experiences with the process from actors involved, predefined KPIs, and known bottlenecks. The business expert and process analyst can collect this information through interviews and reviewing written process documentation. The tacit knowledge helps the project team to understand the process and conceive solutions later on.

4.1.3. Data processing

The third phase focuses on the processing of the data collected. This phase is separated from the extraction activities in the previous phase, since data processing is an iterative process, while the extraction is usually only done once. Four activities are identified. The first is the **creation of the different views (event logs)**. Depending on the research questions, several attributes need to be chosen and added to the event logs. For example, the case ID, timestamp, activity name, activity costs, and resources needed can be added as attributes. These can be used for looking into the costs of resources. During the second activity, the **events are aggregated** to help reducing the complexity and improve the structure of the mining results. With aggregation, event data with matching properties can be merged. When two attributes that only differed on an unnecessary level of detail, it can help to combine them into a more general form to reduce complexity. For example, the two activities “enter person’s name” and “enter person’s ID” both belong to “enter person’s personal data”. The third activity concerns the **enrichment of the event logs** with various additional attributes. This can be done with the help of data from the log itself, or by adding external data. For example, when start date and finish date are known, the duration can be calculated and added as a new attribute. The last activity of this phase includes the **filtering of the logs**. Several techniques are possible to order the dataset and exclude unneeded elements. PM^2 distinguishes three types: slice and dice (remove events or traces based on the recorded values for a specific attribute), variance based filtering (cluster events), and compliance based filtering (remove traces or events that do not fit a given process model).

4.1.4. Mining & Analysis

In this phase, the actual process mining is done. Based on the research questions, a selection of four activities can be executed. The first three are already discussed by Van der Aalst (2016) as standard process mining activities. First of the four is the process discovery. Based on an event log, the software selected **generates a process model**. What modelling language and what mining technique are needed, depends on the research questions. The second activity includes **conformance checking**. Discovered models can be compared with the already documented process models. Another activity is the **enhancement** of process models. An existing process model is extended or improved, based on recently gained knowledge. These adjustments do not have to affect the organisation’s performance directly, but they do improve the usability of the models. For example, a step where workers need to enter a serial number of a vehicle in the ERP, is added to the model. This does not improve the maintenance process directly, but makes it easier to trace maintenance activities. The last activity of this phase is **process analytics**. Other analysis techniques, like data mining, visual analytics, and statistical analytics, can be used to help answering the research questions.

4.1.5. Evaluation

During the evaluation phase, the results from earlier phases are elaborated into improvement ideas. The research questions and goals are already known, just like the mined process models and their performance. Now these need to be combined to find improvement ideas or new research questions. There are two activities involved: **diagnose**, and **verify & validate**. Diagnose is the first activity. During this step, the findings from previous actions are correctly interpreted. Discussions need to ensure that the whole team understands the mined process model. It also helps distinguishing and interpreting expected and unexpected results. Lastly, the research questions can be refined if another round of iterations is needed. The second activity concerns the **verification & validation** of the (sometimes unexpected) findings. First, the findings are compared with the original process data. And second, the findings are compared with the claims of process stakeholders. This is done to help identifying possible root causes and help designing and refining process improvements.

4.1.6. Process improvement & Support

During the final phase, process improvements are implemented. Up to here, process mining delivered fact-based input for process improvement efforts. However, the PM² methodology stops after recommendations are delivered to the client. Implementing the improvements requires a methodology on its own and this can be done according to other BPIs like Six Sigma and Business Process Reengineering. However, when the improvements are implemented, their effects can be determined by mining the process again, but now with new data. Structured and matured processes qualify for operational support while being active. The goals are to detect problematic running cases and predict their future, or suggest recommendations. This is done on a continuous base with live event data and concerns a challenging form of process mining.

4.2. Shortcomings of PM² and potential points of improvement

When looking at the PM² methodology, the emphasis on data stands out. Two out of the six steps solely concern the extraction and processing of data. This is effected by the used tooling. PM2 was built around the academic orientated ProM, which requires more data orientated activities than its businesses orientated alternatives. ARIS PPM, for example, supports a direct connection with the data source and (nearly) atomically retrieves and structures the data. This indicates that PM² is tool dependant.

During the pilot at MoD, a lot of attention was given to the team dynamics. MoD is a large organisation and its digital infrastructure (including SAP) is complicated. On the one hand, this requires well scoped research questions. However, PM² start identifying goals and translating them into research questions even before the project team is composed. Even though it is possible to redefine the question during the project, it seems more logical to first compose the team and then compile the questions. And on the other hand, having a complicated infrastructure requires more knowledge and thus a larger team.

Coming up with a good research question should not be underestimated. When a specific type of process instance needs to be found in a large dataset, it is possible that a series of scenarios needs to be compiled that, when combined, function as a single filter. For example, the question “do maintainers plan their activities in advance?” should be rewritten to something like “what sequence of status changes in the process indicates that maintenance is well-planned and how often does this sequence occur in relation to its total?”. This question should be subdivided in the questions “what weapon systems require maintenance?”, “how often is planning actually required?” (some activities can be done ad hoc), and “how often are there parts ordered before the maintenance request is generated?”. The combination of these three sub-questions should only show the process instances that are actually planned. Coming up with these scenarios is a difficult, iterative process, and often requires a multidisciplinary team. PM² gives little attention to this complex job.

Lastly, van Eck et al. (2015) claim that in the last phase, "the results of a process mining project then form the fact-based input of such proves improvement efforts" (p. 286). An implementation strategy is not mentioned. Because process mining can address shortcomings in a process, but also monitor the implementation of improvements, it is possible to place process mining in a cycle of continuous improvement. A good example is the DMAIC cycle of Six Sigma.

Based on the discussed shortcomings, PM² can be improved. The functional requirements of the improved methodology include:

- General applicable; even though the improved methodology is designed for MoD, it should be usable by every organisation.
- Tooling independent; the methodology should be practical, so focus is placed on businesses orientated tools like ARIS PPM and Celonis.
- A complete and integral BPI; it should not be "just an analytics tool". With the integration of Six Sigma, it can be compared with other BPIs like Lean and TPM.
- Project freedom; no project is the same, they all require different analysis and different views. A step-by-step cookbook seems useful, but is in the end not practical.

4.3. PM³: an improved methodology

In the previous section, several shortcomings of the PM² methodology were addressed and improvement potential was pointed out. These are used here as the starting point for an improved methodology: PM³. The structure of the method will be based on a combination of PM² (as a proven process mining methodology) and Six Sigma (as a proven BPI). For the improved methodology, five phases are distinguished, following the DMAIC cycle from Six Sigma. These are shown in Figure 4.2 and discussed in the upcoming section.

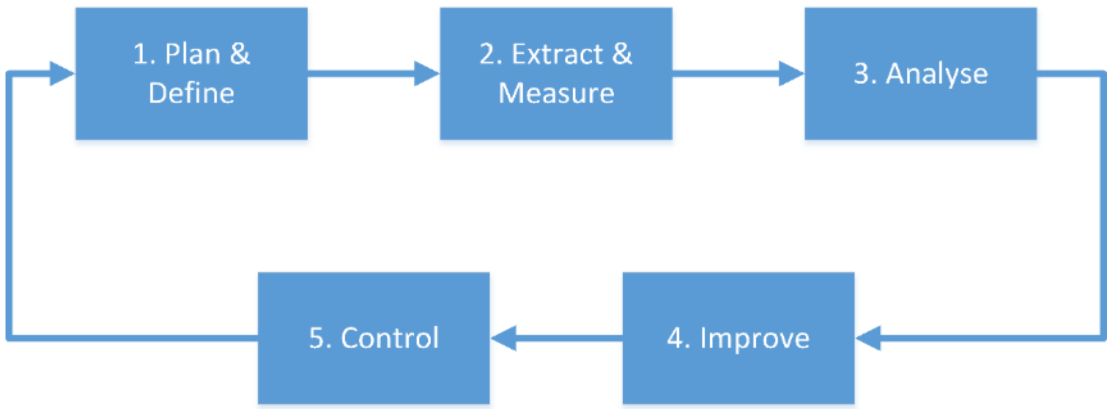


Figure 4.2: The steps of PM³

4.3.1. Plan & Define

The methodology starts with the **client** who has a problem and expects that process mining could support him to find a solution. The decision framework (discussed in Chapter 5) can assess the problem on whether process mining is a suitable BPI. If so, the **assignee** can accept the assignment to solve this problem. In a small team, the assignment should be elaborated till a **goal and clear scope** are determined, stating what needs to be improved. Based on that document, the rest of the team should be composed.

Compared with PM², more emphasis is led on the **composition of the project team**. The digital infrastructure of MoD is immense and complex. Hundreds of different processes are carried out around the clock, on various locations, by thousands of people, and are logged by various systems. In this organisation, expertise is very focused and thus a larger project team may be required. The team should at least have four persons. First, a **process owner** to explain the problem. This can also be the client. Its involvement is strongest in the beginning and at the end of the project. Second, a **process expert** (at MoD this is called the process model holder), who has insight knowledge of the process and is able to explain the current situation in the field. Third, a **system expert** (or IT expert) that understands the information flows within the organisation and has an understanding of its business intelligence systems (ERP, SCM, CRM, etc.). And lastly, a **process analyst** (or business analyst) with knowledge of process mining. In many cases one person per role is not sufficient and a team grows to eight people or even more. The team should be led by **team leader** with knowledge of process mining and the ability to bring them all together. A logical option would be the process analysts. This can also be the person chosen by the client.

The whole team needs to be aware what process mining is, what it can do, and what not. This can be accomplished by organising a **kick-off** and giving a presentation on the topic. At the one hand, this helps to manage expectations. People can be overenthusiastic and drop out when they find out their expectations are not realistic, or they can think process mining is just another hype and they decide to "sit this project out". At the other hand, a joint presentation gives a better understanding of what is expected from them and gets the whole team on the same page. During the kick-off, the team elaborates the goal of the project to several research questions. This may be the most difficult part of the entire project. The questions should lead to root causes of the problem. An example of a bad question would be "do maintainers plan their activities in advance?" and a good question would be "what sequence of status changes in the process indicates that maintenance is well-planned and how often does this sequence occur?". The better the questions, the better the final results. To test the questions, it wise to already discuss possible outcomes of the analysis. At the end of the session, the system expert should know what data are needed and what series of queries is need to distil the answer out of that data.

4.3.2. Extract & Measure

After the kick-off, the system expert starts **extracting the required data**. For a large organisation like MoD, it is normal to use a sophisticated software packages for its business administration. These packages often have interfaces to connect their database to other applications. At the other side, commercial process mining tools like ARIS PPM and Celonis also have interfaces. In some cases, an adapter needs to be developed to connect them, but even if the adapter is already available, it requires significant effort to initialize the connections. This is however, a one-off activity. More of this is described in Section A.3.1. During this phase, the process expert supports the system expert with knowledge where what data can be found. Together they create a draft analysis of the situation.

The findings of the process and system experts are presented during an **acceptance test**. This is a meeting with the complete project team. The raw results are presented and discussed. This discussion may lead to new or redefined research questions, an adjusted project scope, the need for extracting alternative data, or the need of involvement of other specialists. While the process expert shows the results, the team discusses how the results should be presented (what views are required). At the end of the session, the team decides if the project continues with its original questions, with redefined questions, or that it need a new acceptance test after more research is done. This phase also serves as a moment for the process and system experts to asks questions to the team.

4.3.3. Analyse

The analysis phase starts with the process and system experts **further elaborating on the results of the acceptance test**. When all requests are implemented, the **analysis session** can be held. This is again a team effort. The latest results; the performance of the studied processes, are evaluated in order to answer the research questions. The first goal is to statically support the processes' improvement potential by identifying bottlenecks. This part is led by the business analyst. The bottlenecks will lead to root causes. And when the extent of its impact is known as well, improvement suggestions can be made. What these suggestions are and how they are made up is very case depended. It is important to create an active environment where the team has the right knowledge and competencies. Brainstorming can be alternated by zooming in on the process to validate ideas and insights. At the end of the session, the team should agree upon its findings and its future improvement actions.

If required, the system expert can also start **developing a dashboard**. Many software packages (like ARIS PPM and Celonis) support the dashboard to give non-analyst the possibility to monitor the process. First, it can show how the improvements catch on and steer the activities if necessary. Second, when the process mining project has ended, the dashboard can still be used to monitor the process and intervene when the process displays abnormal results.

If there is serious doubt on the quality of the dataset, an action can be to **improve the process of datalogging**. This sets the business process improvement back to its start, but can be a necessary evil. After new and better data are collected, the process can be restarted.

4.3.4. Improve

During the improve phase, the suggestions are translated into **practical measures**. In contrast to the previous phases, not the complete team is required to participate. The improvement of the process can have many forms: adding, removing, or relocating activities, creating extra fail safes, relocating personal or adjusting their indicated required competences, or even reducing lead time by investing in better equipment. During this step, it is important to avoid discussion on facts and figures. The data and analysis prove the need and effects of the modification of the process.

4.3.5. Control

An important aspect of DMAIC is that it is continuous. The control phase serves to sustain the benefits from the improvement and to become a starting point for further improvement. A **closing session** is held to evaluate the project and celebrate its successes. It is important to secure the knowledge in the organisation. Therefore, the project documents should be archived correctly. Comment on the process mining methodology should be spread among other system experts and process analysts that are involved in this methodology. If this is a large group, it is advised to accommodate this in a steering group.

With the dashboard or by periodic repeating the analysis by the process expert, the effects of the **improvement can be monitored** and adjusted if necessary. If necessary, this process of control can be secured in the organisation by making it business process as well.

4.4. Conclusion

In this chapter, the PM^2 methodology was explained in great detail. As a reaction on disparities learned while carrying out the Defence broad pilot Process Mining, this methodology was revised. The basis is formed by PM^2 in combination with Six Sigma and the outcome is PM^3, a methodology that embraces the philosophy of continuous improvement. A model of the methodology is depicted in Figure 4.3. PM^3 is a practical by-product of the thesis project and can be used by others for applying process mining, but it also provides support in answering the main research question. In this chapter, several characteristics of process mining came to light, which will later be used to determine the suitability of process mining on MoD's processes.

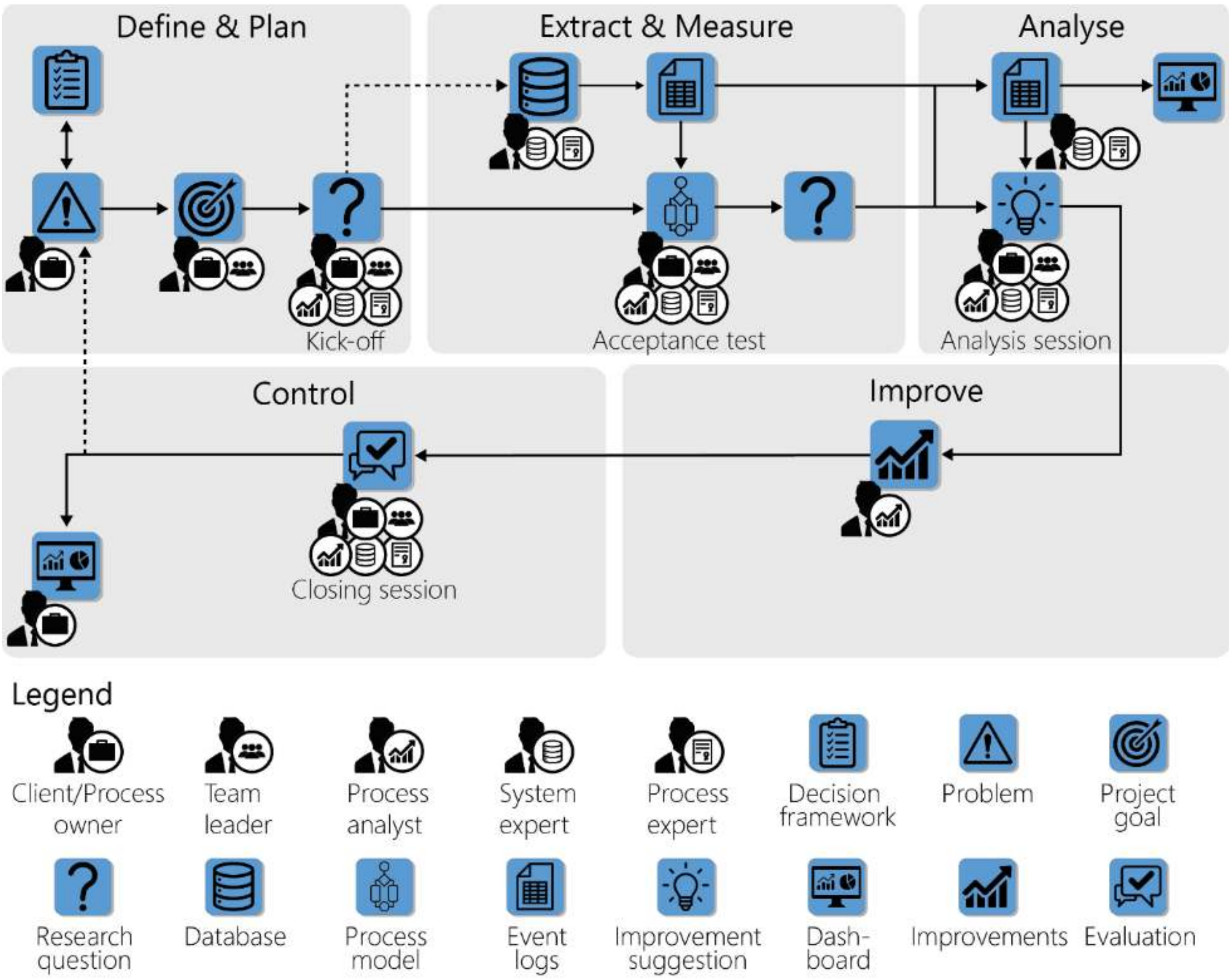


Figure 4.3: The PM³ model

5. Developing a process mining decision framework

After having described process mining and how to conduct a process mining project, the thesis continues with focussing on its main research question. The goal of this chapter is to come up with a list of characteristics that can be deducted from a business process model, to answer the question whether process mining can successfully be applied. This is unknown terrain. The literature does not give a clear list describing the central characteristics of a business process. Nor does it suggest what should be taken into account when determining whether a process is suitable for process mining. In order to develop the list of characteristics, or perhaps better 'the decision framework', the waterfall model is followed. This model is described by Royce (1987) and has proven itself in the software industry as a sequential (non-iterative) design process. Activities steadily succeed each other through a set of phases. This software development model discerns various variants and to adapt it to the development of the decision framework, the following phases are selected: requirements, design, implementation, and evaluation. The waterfall model is shown in Figure 5.1.

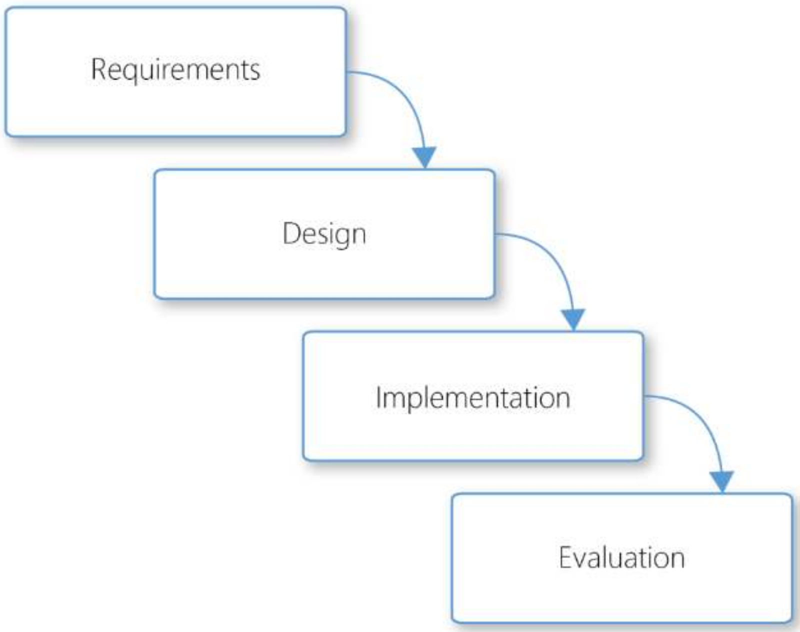


Figure 5.1: The waterfall model used for the decision framework

5.1. Requirements

The first phase involves drafting the design requirements. The goal of the framework is to anticipate on the question if process mining is the right methodology for a project. Meaning, whether process mining will add enough value when analysing and improving processes. For the framework, the usability of other BPIs will be disregarded.

The framework should:

- Use a process model as leading source of information (and look to the "real" world only if the model lacks information).
- Be modelling language and organisation independent. The framework is developed for MoD with its ARIS process model library. However, it should be useable by all types of process model languages and organisation.
- Be usable without experience or complex manual. The goal is to supply a framework that is user-friendly and can be used at the start of a process mining project, even before a team is compiled.

- Be usable in solitary. It should not be a team effort, but help from experts is allowed.
- Support the user by giving a go/no-go answer and/or a numerical score.
- Be adjustable based on experience. The framework is a first attempt and should be improved during its use.

5.2. Design

This phase consists of three parts. First, a conceptual business process model will be developed to gain a uniform and abstract viewpoint about processes and its characteristics. Then, enterprise modelling is explored in more detail. Both in general and how it is done at MoD. This gives a more detailed viewpoint. Both result in the third part, the selection of the characteristics needed to construct a decision framework. The last selection shows the framework in its entirety.

5.2.1. Conceptual business process model

The goal is to develop a list of characteristics that determines whether a specific process of MoD will benefit from process mining. A uniform and abstract viewpoint is needed to develop these characteristics. That is why first, the generic business process is examined. A conceptual business process model shows what characteristics, or attributes, are involved in a business process.

The literature describes many models of business processes. For example, Jäger, Schleicher, & Westfechtel (1999) who developed a model for software processes and Sheth (1973) who developed a model of a generic purchasing process. But to comply with the needs of MoD, a generic model is required that fits all of its processes. Eriksson & Penker (1998) describe a more generic model, but only on a high level of abstraction (e.g. input and output are not further defined). Finally, List & Korherr (2006) developed a conceptual business process model with sufficient depth, however, they divided it into several perspectives and did not use a specific modelling language. For this study, a single generic model from a business process context perspective is needed. That is why a new model is developed.

The conceptual business process model will be developed according to the Unified Modelling Language (UML) (Object Management Group, 2015). This is a commonly used general-purpose modelling language with its roots in the field of software engineering. It intends to standardize a way to visualize the design of a (software) system, but can be used to visualize other models as well.

UML discerns six different structures. These views describe the system from different perspectives. For this model, only the class diagram is used, which shows the system at a level of classes and interfaces. The other structures describe the system on activity level and lose their generic value. In software engineering, a class is a (generic) template for an object with optional properties. However, for the modelling of this generic business process all objects are unique and are not reused like in software engineering. Classes are represented as boxes and can contain three compartments. The first being compulsory, namely the name of the class. The middle and bottom compartments are non-compulsory and respectively contain attributes and operations that the class can execute. For this model, only the name is used.

The class diagram visualises the interfaces between classes in the form of association relationships. Several types of relationships are defined to generalize these connections in a logical way. Table 5.1 shows four types of relationships. The first, the inheritance relationship, visualises an "is a" relationship. All characteristics of the higher class are also applicable to the lower class(es). Next to that, the lower class(es) can have specific characteristics. For example, a printed novel has an inheritance relationship to a book. The second relationship is the aggregation. This is a variant of the "has a" association relationship. It represents a part-whole or part-of relationship and can occur when a class is a container or collection of other classes, but the contained classes can exist when the container is destroyed. For example, a library has books. The books can exist without a library and the library can exist without books (when it only has magazines). The third relationship is the composition. This also represents the “has a” association, but then as a real-world whole-part relationship, which is stronger than the aggregation. The composite is responsible for the creation and destruction of the component parts. In other words, the lower class cannot survive without the higher positioned class. For example, a library has members, but when the library stops, its members will stop being member as well. The last relationship is the dependency. Also, being called a supplier-client relationship, the dependency is used to show when a class requires, needs, or depends on other classes for specification or implementation. For example, the library depends on its suppliers.

Each association relationship can also be supported by a multiplicity. The multiplicity indicates the number of instances that are connected to a class. If this is 1, it means there is exactly one. The library, for example, has exactly one head office. If it is 0..1, it means that there is no or exactly one instance. For example, a member can have zero or one member card(s) connected to his/her account (it is zero when it is a new account or the card recently got lost). When 0..*, there are zero or more instances. The library can have zero to infinite books in its catalogue. Other numbers are also possible.

Table 5.1: Association relationships for UML

◇———	◆———	◁———	<- - - -	1	0..1	0..*
Aggregation	Composition	Inheritance	Dependency	Exactly one instance	No instances, or one instance	Zero or more instances

Developing a conceptual business process model

Based on the rules of UML and the design of a generic meta-model of a business process by List & Korherr (2006), an adapted conceptual meta-model of all business processes is designed. This is shown in Figure 5.2.

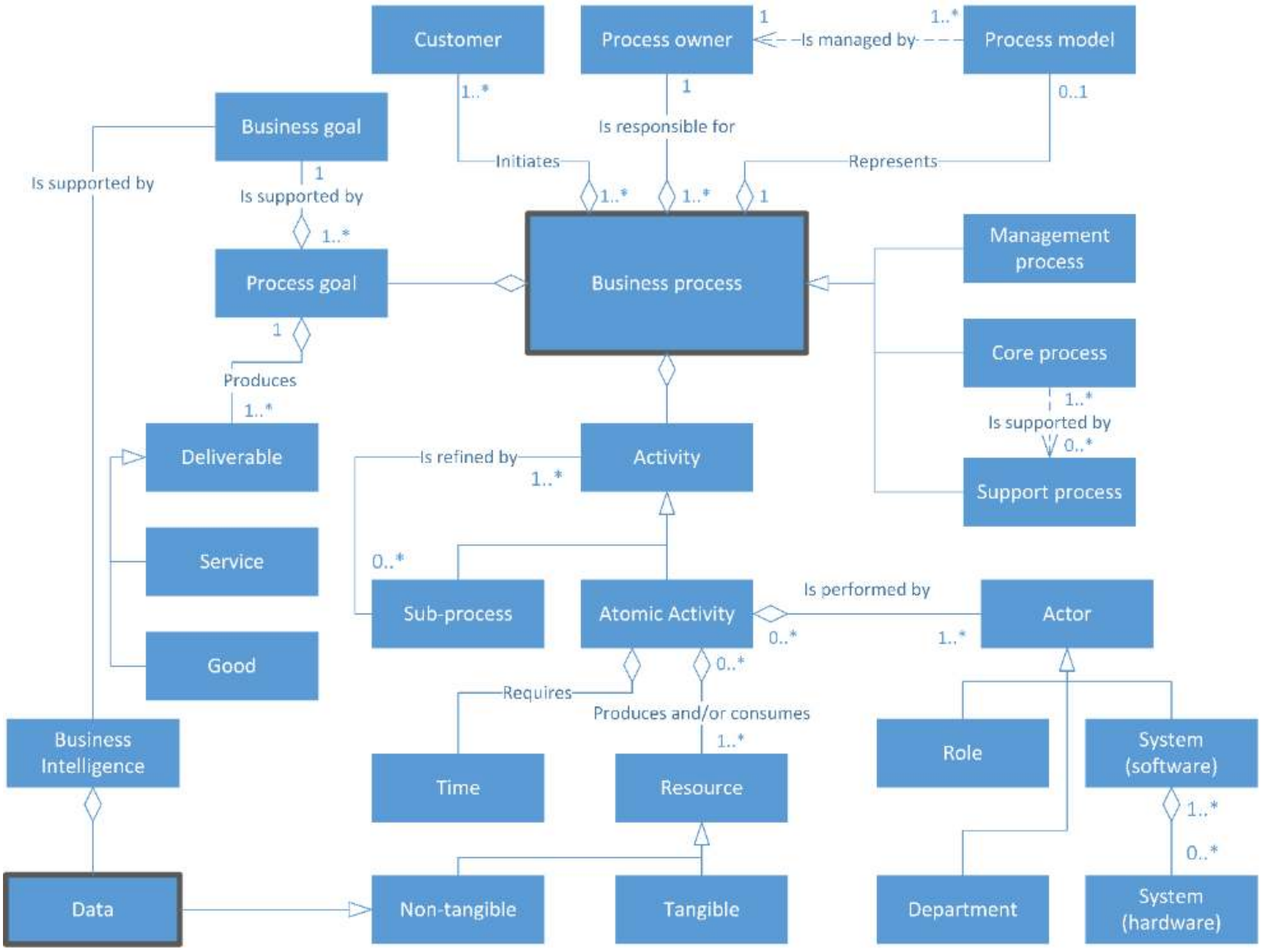


Figure 5.2: UML class diagram of a generic business process

The business process is placed in the centre of the model. The process can be subdivided into core processes, supporting processes, and management processes. Core processes generate products and/or services. Support processes do not directly add value to the organisation, but are necessary to assure the functioning of the core processes. Management processes plan, organize, communicate, monitor, and/or control the activates of the organization. All characteristics of the business process are automatically transferred to the three subordinate processes. A business process has one or more customers and process owners, and it can have no or one formal process model. This last one is particularly interesting for process mining, where the difference between the formal (*de jure*) and actual model (*de facto*) is investigated. Next to that, the business process has a process goal. This can be subdivided into a business (or organisational) goal and direct goals in the form of deliverables (goods or services). Lastly, the business process consists of activities. These activities can be a series of sub-processes on their own. After possible rounds of alliteration, the atomic activity cannot be further broken down to a finer level. The atomic activity is performed by actors. These can be one or more predefined roles in an organisation, or a (complete) department. A system, in the form of software, can also be an actor. For example, when the ERP sends an automatic reminder for annual maintenance. Software can also control hardware, that in turn performs physical activities. The atomic activity requires time. In many situations, activities also need resources. These can be tangible (for example machines, money, and raw materials) or non-tangible (for example data, knowledge, and services). A special non-tangible resource is (digital) data. Next to its role in creating value in the form of services and products, it can support the business process at a higher level. It can lead to business intelligence to place the business process in its organisational environment. This makes it easier to steer the goals of different business processes towards the organisational goal. The data can also be used to assess and improve the process.

Evaluating the conceptual business process model

To evaluate the model, it is compared with one of MoD's actual business processes: Count stock and inventory (P.2.3.06). The model can be seen in Figure A.5.1 and the outcome in Table 5.2.

Table 5.2: The process Count stock and inventory evaluated with the generic business process

Business process	**Count stock and inventory**
- Management process	The process is part of managing the supply chain of goods by monitoring (counting) its throughput. Thus, it is a management process.
- Core process	The process does not directly add value to the organisation and thus is not a core process.
- Support process	The process does support the supply chain process. Next to a management process, the process can be classified as a support process.
Customer	Yes, head of inventory accountant.
Process owner	Yes, currently Lieutenant Colonel Verzijl.
Process model	Yes, ARIS model P2.03.06.
Process goal	The purpose of this process is to perform the count and to record the results.
Business goal	The counting process is part of the whole of activities which ensures that supplies are delivered in the correct quantities and with the desired quality at the agreed time and place in the agreed appearance in packaging and conditioning.
Deliverable	
- Service	Counting document
- Good	n/a
Activity	
- Sub-process	Not at this level.
- Atomic Activity	Six atomic activities are found. For example: Count inventory (number 010).
Time	Short: the process can take from hours to a few days.
Actor	
- Role	Warehouse worker inventory management and Warehouse worker warehouse management
- Department	n/a
- System (software)	SAP
- System (hardware)	n/a
Resource	
- Tangible	Input: counting lists Output: counting document
- Non-tangible	Input: several data
Data	Among outers: warehouse number, warehouse location, inventory document, article document, etc.
Business Intelligence	Progress and performance are being monitored.

The process fits the model. However, it needs to be said that the model is meant as an abstract starting point for this research and does not include all attributes of a business process. Further research could complete and improve this model, but for now, the depth and level of detail are sufficient.

5.2.2. Enterprise modelling

In the previous section, the attributes that belong to the generic business process were discussed. This section goes one level deeper. Here, a closer look at the modelling of complete processes is given. First, the concept of enterprise modelling is being discussed. After that, the modelling techniques used by MoD are described.

Enterprise modelling in general

The obtainment of a comprehensive picture of an enterprise falls within the field of study of *enterprise modelling*. The enterprise model can be seen as an abstract representation, description, and definition of an organization that incorporates all the knowledge regarding an organization, including its resources, products, and the way the organization communicates (Aversano & Tortorella, 2015). Enterprise modelling finds its roots in the information technology planning. Here, models represent the way the organizational structure supports the use of IT and the way both can be integrated. Later, enterprise modelling proved useful for capturing and improving business strategy and business processes as well. The goal is to provide a complete picture of the entire organization. This helps to refine processes to meet organizational goals, counter external threats, and eliminate internal weaknesses. Next to that, the enterprise model can help improve management techniques, develop internal procedures, and secure the organization's long term commitments. The role of an enterprise model is to achieve model-driven enterprise design, analysis, and operation.

In a way, enterprise modelling shows many similarities with business process reengineering (as mentioned in Paragraph 3.1.9). In both cases improvement is based on the modelling of processes. However, business process reengineering focusses on the improvement of individual processes and not the organization as a whole. System theory teaches us that "global performance is not the sum of local business performances but indeed the optimization of all business processes" (Doumeingts & Ducq, 2001, p. 146).This suggest the effectivity of enterprise modelling. Also, modelling an enterprise does not directly lead to improvement. So, on its own, it is not a BPI, but it can be integrated in one to increase results.

Several enterprise modelling techniques are in existence. First of three examples is the Design & Engineering Methodology for Organizations (DEMO), an enterprise modelling methodology for transaction modelling of business processes (Dietz, 2001). An example of this model can be seen in Figure A.5.2. Second, Dynamic Enterprise Modelling (DEM), which is designed as a fast, flexible, and integrated business oriented approach (Baan, 1996). An example of this model can be seen in Figure A.5.3. And last, Integrated Enterprise Modelling (IEM), which is an object-oriented modelling approach that differentiates by reusable libraries of object classes and business (Mertins & Jochem, 2000). These different techniques share the same goal of disclosing the organizational model, but intent to do so with modelling different "views". The focus can be laid on an information view or on a function view and the modelling language can be different as well. For most modelling techniques, a single view is incapable of modelling the entire enterprise. Thus, different views are aligned and connected to create a comprehensive blue print.

Enterprise modelling at MoD

For the modelling of its integrated business processes, MoD uses the software ARIS from Software AG. ARIS is an "Architectural framework for Integrated Information Systems" (Scheer, 1994, p. 607). As a software system, the ARIS Business Architect (formerly ARIS Toolset) is capable of analysing, modelling, and navigating through business processes. The way MoD implemented it, ARIS supports business processes over the complete Business Process Management (BPM) lifecycle from which various views can be considered. It can be seen as more than just a modelling tool, it is moreover a centralized database in which the variety of relations can be visualized, depending on the target audience.

At the basis of this method lays the ARIS-House of Business Engineering (ARIS-HOBE) (Scheer & Nüttgens, 2000). This framework, as shown in Figure 5.3, is used for managing business processes (from organizational engineering to IT implementation and including continuous adaptive improvement).

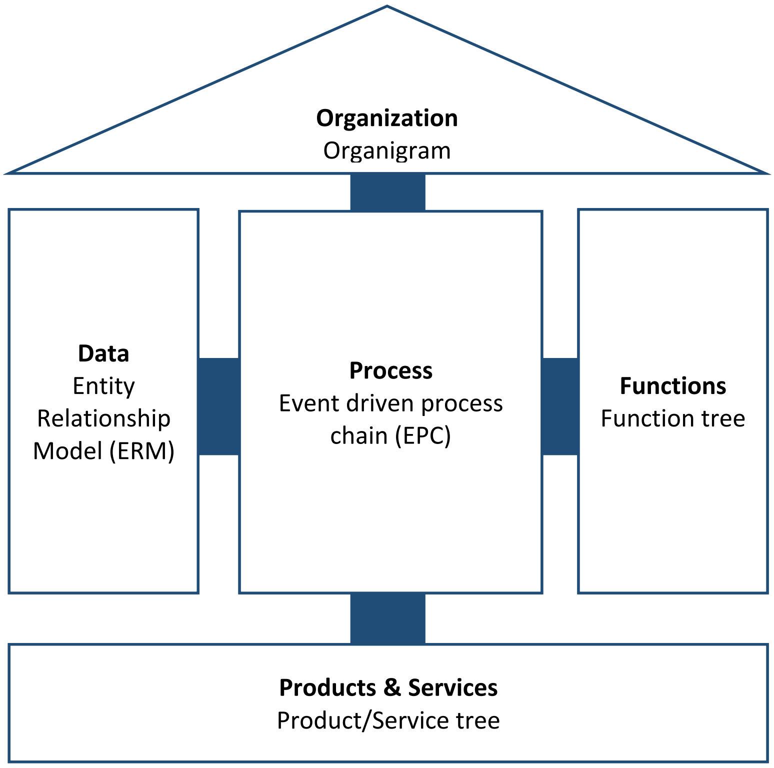


Figure 5.3: ARIS House of Business Engineering

According to this framework, each process can be divided into five different views. At its centre lays the process view, which interconnects the five views. A process is set in motion by one or multiple actor(s). This is shown in the organization view. The same processes require and/or generate data. The question what data are needed or produced is answered with the data view. The functions view shows the composition of the different actions involved. In other words: what is done in this process? Finally, the process produces output. And this is shown in the products and services view. If all elements can be modelled, then the business process is adequately described.

A simplified example of an enterprise model in ARIS is seen in Figure 5.4. The organization view tells us that the library has a front office, and at the front office works a desk clerk. According to the data view, the library has one (or multiple) member(s). This datatype has a membership number and name as attributes. The library also has book(s) in its IT system. A book has a ISBN and name as attributes. This simplified model focusses on the checkout of a book. This function (process) can be subdivided into three (atomic) functions (activities), namely: checking the member's account, register the transaction, and print the ticket as a proof. In the products and services view it is shown that the library has two products: book(s) and magazine(s). All this information is combined in a clear process model, as shown in the centre. This process starts with the request for a book and ends with the completion of the transaction. In real live, these processes can exceed the hundreds or even thousands of blocks. Section 5.1 in the appendix shows an overview of the most used symbols and operators in ARIS (version 9).

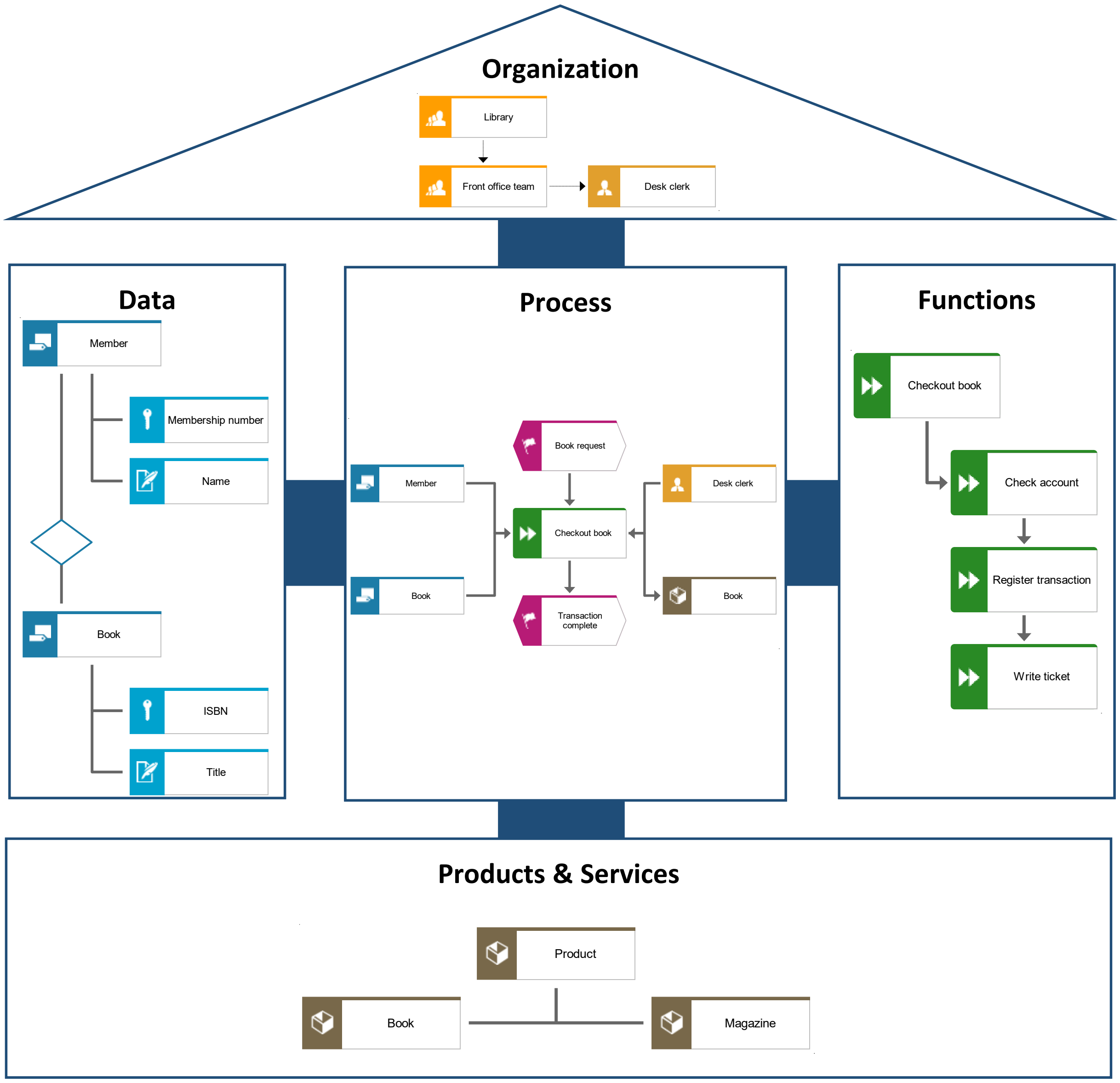


Figure 5.4: An example of the ARIS-HOBE of a library checkout

At MoD, several types of views (flowcharts) are being used in ARIS. For the process view these are: IDEF0-diagram (in the future), Value Added Chain Diagram (VACD), Event-Driven Process Chain (EPC), Function Allocation Diagram (FAD), and Business Controls Diagram (BCD). EPC is central to ARIS and connects all other views, as well as describing the dynamics of the business process. For the organizational view, this is the organizational chart (or organigram). For the data view, this is the Entity Relationship Model (ERM). For the function view, this is the function tree. And for the products and services view, the Product/Service tree is being used. The ARIS Toolset also supports the Business Process Model and Notation (BPMN) in order to model business processes, but this is only used sporadically within MoD.

5.2.3. Selecting characteristics to measure processes mining suitability

After gaining practical experience with process mining, constructing a conceptual business process model, and learning how MoD models its processes, all knowledge needed to construct a decision framework is present. This framework uses several quantitative and qualitative characteristics to distinguish the different processes on their suitability of process mining. These characteristics are chosen based on previously mentioned knowledge in combination with intensive consultation with process experts.

Table 5.3 shows the list of the characteristics used for the decision framework. This section will clarify each characteristic and give an explanation of its origin, meaning, and impact. The characteristics are ordered on their content, and not on level of importance.

Table 5.3: List of characteristics

1	Availability of the minimum required attributes
2	Number of roles involved in the process
3	Number of decision points
4	Number of activities
5	Lead time (average)
6	Quality of dataset
7	Methods of importing data
8	Possibility to filter on attributes
9	Nature of process
10	Process maturity (years of experience)
11	Availability of process model
12	Mandated is known

For each characteristic, the effects are quantified with the use of plusses and minuses. The meaning of these symbols is explained in Table 5.4. Each symbol represents a value, ++ being two and -- being minus two. When the chart is filled in, the final score can be deduced by adding the individual scores. The scoring has been established based on trial-and-error and brainstorming with several process experts. In Section 5.3, the scoring was tested by applying it on actual process models of MoD. Its result was used to improve the initial scores. Only the final scores have been included in this report.

Table 5.4: Explanation of symbols used to quantify the effects of the characteristics

Symbol	**Value**	**Meaning**
++	2	The characteristic contributes positively to the use of process mining.
+	1	The characteristic contributes moderately positively to the use of process mining.
+/-	0	The effect of the characteristic can be positive or negative. This depends on other attributes and the goal of the project.
-	-1	The characteristic contributes negatively to the use of process mining.
--	-2	The characteristics indicates that process mining is not possible.

Characteristic 1: Availability of the minimum required attributes

To mine a process, some attributes in the event log are mandatory. Even though, from a scientific perspective it is still being debated what is and what is not mandatory to mine a process model, from a practical perspective three attributes are necessary to obtain useful results. First is the case ID to distinguish different activities of the same process and pair them to a single instance. Second is the activity. The different process steps or status changes should be named to add detail to the analysis. And lastly, a timestamp is needed to order the activities from a single instance and order the different instances in the log. Theoretically, a number that tracks the order of activities and instances is sufficient to deduce some conclusions. It is for example still possible to compare the number of actives each actor carries out. But adding a proper timestamp gives the possibility to calculate how long the actor needs to carry out these activities.

Whether the minimum required attributes are available cannot be deducted from the process model. The process owner and system expert can be addressed for this information.

Table 5.5: Effect on process mining (Availability of the minimum required attributes):

Required attributes (case ID, activity, and timestamp) are available		
	Effect	++
	Explanation	Process mining could be possible.
Required attributes (case ID, activity, and timestamp) are unavailable		
	Effect	--
	Explanation	Process mining is not possible.

Characteristic 2: Number of roles involved in the process

Even though the performer-ID is not a mandatory attribute to create a model by process mining, the knowledge of who is performing what activity will give extra analysis possibilities. As discovered in the generic business process (Figure 5.2), the performer can be a role fulfilled by a person (administrator, planner, worker, quality controller, etc.), but also a system (automated computer activity, robot, etc.) or a department. Lots of knowledge can be deduced from the involved performers. At process level, a social network analysis can show who interacts with who. And at activity level, the productivity of performers can be compared. If for example one person seems to fulfil a critical role in the process and nobody is able to replace that person if needed, one can think of training someone else to take over if that person falls ill. Or if two persons have the same tasks, but person A does it much quicker than person B, it is wise to investigate the reason of this difference. Another social aspect that can be studied with process mining is the segregation of duties, functions, and responsibilities (the handover of work). Several of MoD's processes have control moments built in. For example, when a mechanic finishes his/her task, his/her superior needs to check and approve the result (the four-eyes principle). When applying the correct filters in process mining, violations of this principle can easily be found.

Based on these process mining possibilities, it can be said that having more performers involved in a process, more and stronger analyses are possible and thus more process improvement opportunities arise. The number of involved (unique) roles is often described in the process model, the number of (unique) performers is not. This is because one role can be fulfilled by multiple persons (normally more than one mechanic is involved in a maintenance task, while the process only knows one role: the mechanic). For this reason, this characteristic only takes the number of roles involved into account. It is assumed that during the analysis, the roles can be split apart in different actors (based on performer ID).

A possible downside of having many different roles (and actors), is that process improvements require a larger group to change the way of carrying out its activities.

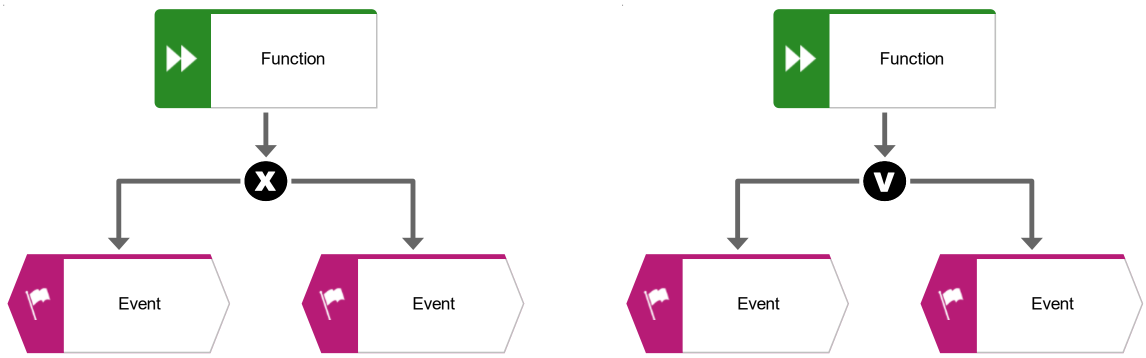
The number of roles involved can easily be counted in a process model. When a process has more than four unique roles, it is determined as many. Less than four unique roles, means few roles are involved.

Table 5.6: Effect on process mining (Number of roles involved in the process)

Many roles involved (n ≥ 5)		
	Effect	+
	Explanation	Having more roles will result in more improvement potential. Social networks can be analysed and individual productivity can be compared.
Few roles involved (n < 5)		
	Effect	-
	Explanation	Less analysis possibilities, but changes are easier to implement.
No roles involved		
	Effect	--
	Explanation	Every process needs actors, so mining a process with no actors is not possible.

Characteristic 3: Number of decision points

The process modelling language used by ARIS, EPC, uses three types of junctions to connect different events with activities, namely: AND, OR (technically an exclusive OR, or XOR gate), and AND/OR (technically an OR gate). Other modelling languages uses different names, like logical gates in computing and gateways in BPMN. Their functions are broadly similar. Several scenarios are illustrated in Section A.5.1 and two of them are shown in Figure 5.5. In these two specific scenarios, carrying out its function requires actors making a decision.



OR-junction
one of the paths is active, but not both

AND/OR-junction
one or both of the paths are active

Figure 5.5: EPC junctions

These decision points determine the course the process is taking. Having more of these decision points gives the possibility of more unique instances. Because with every extra OR- or AND/OR-junction connected to two events, the number of possible unique instances doubles. This does not necessarily have to lead to an increase in instances, but practise shows it is common.

On its own, the number of unique instances is not a good or bad thing, it is merely an observation. However, it is an extra challenge for the analyst. Experiences at MoD tells us that having around 70% to 90% unique process instances is normal. In other words, if a dataset has 1000 cases, the dataset can have 800 unique process instances. This is typical for spaghetti processes. Displaying such a process in a clear model is difficult. That is why the analyst should aggregate and filter during the data processing phase and filter during the mining phase.

On the bright side, the more decision points the process model has, the more possibilities for improvement there are. The decisions are often made by human actors and are sensitive to error. With process mining, the compliance of business rules can easily be checked. Deviating process instances can be taken apart and checked in isolation.

The number of decision points can easily be seen on the process model. When a process has five or more decision points, it is determined as many. Less than five decision points, means few roles are involved.

Table 5.7: Effect on process mining (Number of decision points)

Many decision points ($n \geq 5$)		
	Effect	+
	Explanation	More potential to improve processes and check compliance of business rules.
Few decision points ($n < 5$)		
	Effect	+/-
	Explanation	Process mining becomes easier.
No decision points		
	Effect	-
	Explanation	Having no gateways makes the process rather simple. Process mining still can be used, but only to the extent to visualize lead times.

Characteristic 4: Number of activities

Just like with the number of actors involved and the number of decision points, the more activities the process has, the larger its improvement potential is. In the generic business process (Figure 5.2), these are the atomic activities. Slow, problematic, or resource-intensive activities stand out during process mining. These activities can be improved in isolation or the process as a whole can be improved by altering the activities location in the process.

The number of activities can easily be seen on the process model. Having five or more activities is seen as many. Having less than five is seen as few.

Table 5.8: Effect on process mining (Number of activities)

Many activities ($n \geq 5$)		
	Effect	+
	Explanation	More potential to improve processes.
Few activities ($n < 5$)		
	Effect	+/-
	Explanation	Process mining becomes easier.
No activities		
	Effect	--
	Explanation	Every process needs activities, so mining a process with no activities is not possible.

Characteristic 5: Lead time (average)

Since time is an important resource, improving processes often involves shortening its lead time. When mined, the lead time is thus an important measuring instrument. Overall, processes with a long lead time have more potential to be improved than processes with a short lead time. From a global perspective, the average lead time can be computed by process mining and can be compared with a predefined KPI. When the KPI is not met and when the standard deviation between the different instances is not significant, the process needs to be studied. A common solution is to improve the model and/or assign more resources (like manpower) to the process. From an instance perspective, a deviant lead time can show a flaw in the process handling. Thus, when the lead time does not meet the KPI but this is only seen in a set of instances, that particular set needs to be examined more closely and compared with the better performing instances. Possible improvements would be to train the actors involved.

When the lead time of a process is very long, the chance that the (de jure) process model is altered during the timeframe of monitoring increases. With that, the risk for concept drift increases as well. As described in Paragraph 3.2.2, dealing with concept drift is difficult and can lead to unreliable process mining results.

Another difficulty of mining a process with a long lead time, is the fact that the timeframe studied needs to surpass the length of the complete process. When the process takes on average seven days, mining in a dataset with four days of data cannot lead to conclusions. Thus, longer lead time requires larger datasets. For organisations that already log their processes for a longer period (like MoD), this is not a problem. But using more extensive datasets will again increase the chance of concept drift.

The lead time is not visualized in the process model, but often can be estimated. The process owner should have an accurate idea of the lead time. The turning point has been set on ninety days. Meaning that a lead time of less than ninety days is short and a lead time of ninety or more days is long.

Table 5.9: Effect on process mining (Lead time (average))

Short lead time (t < 90 days)		
	Effect	+
	Explanation	Less time extensive datasets are needed and less chance on concept drift.
Long lead time (t ≥ 90 days)		
	Effect	+/-
	Explanation	More potential to improve processes.

Characteristic 6: Quality of dataset

The first characteristic already determined that event logs are mandatory and that they at least have three specific attributes: case ID, activity name, and timestamp. (Characteristic 8 will discuss the importance of having more attributes in order to filter within the dataset.) But next to the availability of data, the quality of it also matters. Questions need to be asked on whether the timestamp consists of a date with a time or just a date? And does the time include seconds or even milliseconds? This all determines whether it is possible to order the different activities correctly (else two activities with the same date cannot be ordered). Next to the timestamps, the other data should be correct as well. It can be asked whether the data are entered manually (with an extra chance on mistakes) or if it is entered automatically? And if no misspellings have been made. This leads to errors and filters pick up wrong activities. The quality of the dataset is thus important for process mining.

Processes with a low-quality dataset will lead to incorrect or at least less useful results. In computer science, this is referred as the *garbage in, garbage out-principle*. It is however, difficult to quantify the quality of a dataset. For this analysis, the following principles are maintained to quantify the dataset as high quality: the data reflects the real world with a correctness of over 80%. This can be estimated by taking a sample and can be confirmed by comparing it with the data from the ERP. Incorrect entries often stand out clearly. When an entry is missing, it is labelled as incorrect. The second demand is that the timestamps have an accuracy of seconds.

Coming to this conclusion requires deeper knowledge of the dataset and cannot be deduced from the process model. If this knowledge is not available, the characteristic can be left out of the decision.

Table 5.10: Effect on process mining (Quality of dataset)

Dataset is of high quality (c ≥ 80%)		
	Effect	+
	Explanation	Correctness of over 80% and timestamps include seconds. This will lead to reliable results.
Dataset is of low quality (c < 80%)		
	Effect	-
	Explanation	Results can be incorrect and activities cannot be ordered.

Characteristic 7: Methods of importing data

The event data need to reach the process mining software. Software packages like SAP, SAP Solution Manager, and HP Service Center all have their own ways to save and export event data. At the other end, process mining software like ARIS PPM, Disco, and ProM all have their own methods to import these data. Most desirable is the availability of an adapter that directly connects both packages, ensuring periodic and automated data import. Such an adapter is often a stand-alone application delivered by the process mining software developer. If an adapter like this is not directly available, it may be possible to develop one. This however is accompanied by high costs. Less desirable, but another possibility is to import event logs as a single file. Most process mining software support CSV, XLSX, and XES files. Exporting the event logs from the ERP and importing it in the process mining software is however a time-consuming activity. Since this is all big data, retyping a dataset from the source in the process mining software is not possible. So, if there is no exporting and importing of the event logs possible, process mining is not possible as well.

Import possibilities are not shown in the process model. Therefore, one must look to the software at the source (often the ERP) and the process mining software (for example ARIS PPM). The possibilities are featured on the software vendor's website.

Table 5.11: Effect on process mining (Methods of importing data)

Adapter available off-the-shelf		
	Effect	++
	Explanation	Importing data can go periodically and automatically.
Adapter not available off-the-shelf, but it can be developed		
	Effect	+/-
	Explanation	Can come with (non-recurring) high costs.
Only file-based data import possible		
	Effect	+/-
	Explanation	Only for simple and stand-alone projects.
Data import not possible		
	Effect	--
	Explanation	Process mining is not possible.

Characteristic 8: Possibility to filter on attributes

A single process can be used for creating value in different situations and with different objects. For example, the business process of preventive maintenance is the same for every weapon system. Thus, the same process model is used for maintenance scheduling of a standard truck and that of a fighter jet. However, on an operational level, different activities are executed. For example, on a truck the V-string may need to be examined, while on a fighter jet mechanics look at the guided missile system. It can be assumed that the desired lead time for maintenance of a standard truck is shorter than that of a complex fighter jet. When applying process mining for this example, the dataset can contain both weapon systems, while the analyst is only interested in the fighter jet. To filter on this data, each process instance should be accompanied with the attribute "weapon system”. A process instance can have many different attributes, like whether it is a system of the army, air force, or navy, the price of the weapon system, the year of purchase, etc. Having more of these attributes creates more possibilities for the analyst. The attributes do not affect the complexity of the analysis, since the filters can easily be switched off.

Still, detailed attributes are not always necessary. Having the possibility to look at a process from an aggregated level can be useful at dashboards. As long as the average results stay within the KPI (and the standard deviation is small), the process is “healthy”. Without the filters, a larger dataset can be monitored at ones.

Normally, the number of attributes cannot be deducted from the process model alone. The process owner or IT-manager can be asked for this information. Determining what is a “profound” number of attributes and what is “superficial” is difficult. This is something the analyst should decide for himself.

Table 5.12: Effect on process mining (Possibility to filter on attributes)

Profound filtering on attributes possible		
	Effect	++
	Explanation	Filtering creates possibilities to zoom on sets of instances with similar properties.
Superficial filtering on attributes possible		
	Effect	+/-
	Explanation	Even little filtering possibilities can contribute to good process mining results.
Filtering on attributes not possible		
	Effect	-
	Explanation	The process can still be monitored on an aggregated level, but it is not possible to intervene when needed.

Characteristic 9: Nature of process

The nature of a business process is a good indication of its structure, its improvement potential, and thus its process mining possibilities. Six different functional areas are distinguished. This is loosely based on the work of Van der Aalst (2016), who divided functional arias in spaghetti and lasagne structures. His analysis is supplemented with personal experience. For this analysis, typical situations are assumed.

Product development
Developing products and services are often unstructured activities. It is based on informal collaboration, creativity, is unpredictable, has a low frequency, and is not really data oriented. Streamlining this with process mining seems illogical and difficult.

Manufacturing
Processes that produce goods are often continuously repeated. When matured, these processes are efficient and well monitored. The availability of high-quality data and relatively clear process models give excellent conditions for process mining. Improvement potential may be limited, but even small changes can lead to great results.

Maintenance
Maintenance is often a combination of manufacturing and logistical activities. Products are repaired and maintained according to standard business processes, but often on different locations and needing different equipment and parts. This makes the process partly unpredictable and

unstructured. But, since organisations standardize the planning of maintenance, the process is often well logged. Process mining can be of great value, but its results can be unstructured.

Logistics
The movement of goods is often well documented and carried out on a structural way. However, organisations differ in their maturity. In the future, technology will play a greater role. Goods can be tagged with RFID–tags or even GPS-beacons. The geotags of the travelled route combined with the administrative actions logged by the ERP, give a solid base for process mining.

Finance and office-administration
Financial processes are typically structured and well recorded. Several processes are also repeated frequently (for example paying wages or the procurement and payment of goods). Small improvements are found easily and can lead to big savings. Several administrative tasks are supported by IT systems. An example is the process of reporting an incident. This can be a broken computer or the need for other office supplies. It often involves multiple actors and is also well recorded, making it suitable for process mining.

Auditing
Auditing is the procedure of examining financial books, accounts, or other processes on their correctness and compliance. The auditor follows a predefined path, looking for irregularities in the records. When these records illustrate a process and the minimum required attributes are available, process mining is especially suited to support the auditor.

Normally, the process model does not mention its nature. However, models are often classified under a specific functional area and if not, it should not be too difficult to determine its nature.

Table 5.13: Effect on process mining (Nature of process)

Manufacturing		
	Effect	++
	Explanation	Routinely processes with high-quality data.
Auditing		
	Effect	++
	Explanation	High potential when all preconditions are met.
Finance and office-administration		
	Effect	+
	Explanation	Various routinely processes.
Maintenance		
	Effect	+
	Explanation	Unstructured, but often with high-quality data.
Logistics		
	Effect	+/-
	Explanation	High potential, but only when there is enough data.
Product development		
	Effect	-
	Explanation	Unpredictable and little data available.

Characteristic 10: Process maturity (years of experience)
For process mining, it is important to understand whether the process is novel or mature. In the literature, many business process maturity models are discussed that assess this question (Röglinger, Pöppelbuß, & Becker, 2012). These maturity models are used to identify the different phases a process undergoes from its implementation till the moment it is capable of continues

improvement through quantitative monitoring. The models help to guide the process to improve to its highest possible state.

Immature processes are less ingrained in the organisation. Event logs contain more errors, since people are less experienced with carrying out and logging their activities. Also, less event logs are available for analysis. For lengthy processes, this can form a problem. Improving a process that is not yet fully deployed can be risky. Analysis may show that activities do not comply with the agreed models, but when the organisation is still getting used to these models, this is may come as no surprise. However, process mining may be useful in the case that a new model is being deployed and a process owner is curious to its development. In that case, it can be followed how well the new model catches on. Which actors follow the model? Who needs help? When is the deployed complete? This can all be tracked in real-time.

On the other hand, mature processes gain more from process conformance and enhancement. These processes can benefit from the complete process mining spectrum. By developing dashboards and generating business reports, any outliers stand out and are quickly discovered.

Since applying business process maturity models is a science on its own, it is too complicated for the decision chart. That is why for now, the process age is being used as maturity benchmark. The age is defined as the number of years has passed since the process was implemented.

Process libraries like ARIS indicate when the process was modelled. This is a good indication of its maturity. Another option is to question the process owner. Processes older than three years are classified as old and thus mature. Processes younger than three years are immature.

Table 5.14: Effect on process mining (Process maturity (years of experience))

Mature process (age ≥ 3 years)		
	Effect	+
	Explanation	More and better data available and a more experienced organisation.
Immature process (age < 3 years)		
	Effect	+/-
	Explanation	Not sure if the process is still being integrated, but process mining can follow this progression.

Characteristic 11: Availability of process model

The theory distinguishes three options in process mining: discovery, conformance checking, and enhancement (van der Aalst, 2016). Discovery is the only one where no process model is needed in advance, since developing one is its goal. As an improvement methodology, this option offers less direct improvement potential, but it may be useful to understand the organisation.

At organisations like MoD, process discovery will seldom be required, since processes are already modelled one way or another. When a model is available, it can have various formats. This can be in the form of standard format like an EPC or a BPML, but this can also be a textual description. Standard formats with high granularity make it easier to analyse event logs, since the analyst knows where to look for. This gives the process more potential.

Table 5.15: Effect on process mining (Availability of process model)

Process model is available in a high granularity format		
	Effect	+
	Explanation	The model supports the analysis.
Process model is available in a low granularity format		
	Effect	+/-
	Explanation	The model supports the analysis partly.
Process model is unavailable		
	Effect	+/-
	Explanation	A model needs to be discovered first.

Characteristic 12: Mandated is known

When investigating a business process, the ultimate goal should always be to improve it. However, changing a process often requires approval from higher management. The process thus needs a clear process owner and support from mandated management. In a large organisation like MoD, missing this clarity can become frustrating or even problematic. To prevent (or at least anticipate on) this, it is recommended that the person or department mandated to implement possible recommendations is known before process mining takes place. Often, this is the process owner, but when improvements effects multiple departments, this can be a more senior position. In some cases, no one is formally mandated to change the process model.

At MoD, the process owner is mentioned at the process model. However, whether the mandated is known to the analyst is off course a more personal affair. If not, a more thorough search need to be conducted. Since this characteristic is based on personal experience, during the field study, all processes got the score: ++.

Table 5.16: Effect on process mining (Mandated is known)

Mandated is known		
	Effect	++
	Explanation	The mandated is needed to implement recommendations.
Mandated is unknown		
	Effect	-
	Explanation	When the mandated is unknown or undefined, recommendations cannot be implemented.

5.2.4. Process mining decision framework

Combining all characteristics found leads to the decision framework as seen in Table A.5.1. Because of the way how the framework is built, each characteristic has less than five possible scores. For example, the minimum required attributes are available (score ++) or are not available (score --). There is no middle ground. In those cases, the table shows “n/a” (not applicable). Because of this, the lowest score possible is -13 and the highest possible is +18 (and not -24 and +24).

To use the framework, the analyst should have the process model at his or her disposal. For characteristics 2 (number of roles), 3 (number of decision points), 4 (number of activities), 9 (nature of the process), 11 (availability of model), and 12 (mandated is known), the analyst can easily deduct the scores from the process model. However, it is possible that the analyst is not familiar with the process and is unable to give a score on all characteristics. Other experts can be of assistance. For example, characteristics 1 (minimum required attributes available), 6 (quality of dataset), 7 (import method), and 8 (filtering possibilities) can be easily scored with the help of a system expert and 5 (lead time) and 10 (process maturity) with the help of a process expert.

The decision framework can be filled in from top to bottom, but this order is not strictly necessary. Each characteristic receives a score and when the whole table is filled in, the final score is the total of the individual scores. During the implementation phase (in Section 5.3), it was found that the processes used for implementation and testing score 10 points on average. Based on that benchmark, total scores below 9 points are given the verdict of having a low process mining potential, scores of 9, 10, and 11 a medium, and 12 and up a high potential.

5.3. Implementation

After designing the decision framework, it should be supplied with real process models. This implementation has two goals. First, the model can be tested if it meets its requirements. Second, with the use of real processes, some of the parameters in the framework can be adjusted. For example, the number of activities that classify as "many" (characteristic 4). Or from what final score should the framework advice the use of process mining. To improve the readability of this document, the adjustments emerging from the implementation phase are already applied in the framework. For example, in the previous section the connection between scores and process mining potential was already mentioned, while it technically is a product of this section. The implementation phase starts with obtaining a set of process models, uses the framework to determine its process mining potential, and compares these quantitative results with the researcher's expectations.

5.3.1. Sampling Defence processes

The ARIS process model library of MoD contains 335 integrated business processes. A list is seen in Section A.5.2. MoD has even more processes, but these are not integrated and/or modelled in ARIS and thus are disregarded. To test the decision framework, it is impossible to look into the complete list, thus a sample has to be taken.

The sample should give a good reflection of all integrated processes carried out at MoD. As stated, there are more processes than the ones modelled in the ARIS library, but to not overcomplicate the study, only ARIS models are included. This means that there already is a process model available and the mandated is described. This effects the scores by the decision chart positively and thus the results. Although this is not a problem, it should be kept in mind when reading the results. Models in the ARIS library are already sorted into twelve categories. If possible, all categories are represented in the sample. However, for five (smaller) categories, this was not possible because there was no model available. From the remaining seven categories, fourteen processes were chosen. Based on personal experience, processes were chosen from different sectors and with different characteristics (attention was given to domain, origin, length in time, kind of activities, data intensity, etc.). These processes are listed in Table 5.17.Later analysis showed that the sample was diverse with respect to all characteristics (expect the availability of the model and the knowledge of the mandated).

Table 5.17: Processes chosen for testing the decision framework

Code	Name	Description
P.0.2.02	Manage changes (RfC)	The purpose of the change management process is to coordinate Request for Change (RfCs) in the information infrastructure in managed ways.
P.0.2.03	Manage incidents	Orderly describes the activities in order to handle an incident within the IT management organization of Defence. Incidents concern matters that are disturbing. These are things that in principle should have worked. And do not make any changes to the ARIS Design.
P.0.3.05	Archive data	When there is requested, process and performance data from SAP supported processes can be collected and archived on external databases.
P.1.1.05	Create order	The purpose of the process is to conclude single agreements (orders) with external suppliers for the delivery of items, services and/or payroll operations.
P.1.1.16	Simple purchasing	The efficient simple purchase of goods or services of limited financial size.
P.2.3.06	Count stock and inventory	The purpose of this process is to perform the count and to record the results.
P.2.4.02	Pick articles	The purpose of this process is to move articles from a storage location to the location where the items are collected for further processing.
P.3.1.01	Set up and define organisational vision	The process of drawing up and establishing an organisational vision aims to formulate and determine policy objectives for the period T to T+4, based on the government agreement, the annual adjustments thereon and the priorities specified by the government officials.
P.3.3.03	Pay invoices	The invoice processing process ensures the timely and correct payment of the invoices issued for authorization and ensures the transfer of payment files to the home bank.
P.4.3.02	Generate preventive maintenance need	Preventive maintenance is maintenance that is performed to reduce the risk of malfunction. At the time of preventive maintenance, there are (still) no symptoms indicating the need to perform maintenance.
P.4.3.11	Technical work preparation	The purpose of this process is to create and technically prepare work orders.
P.4.3.30	Carry out (maintenance) work	The process describes the execution of (maintenance) work on which tasks are completed on the basis of an order until the order is completed.
P.4.9.10	Monitor Reliability	The purpose of this process is to monitor the Key Performance Indicators: Reliability.
P.5.2.02	Carry out transport planning	Linking freight units to available transportation resources (vehicle resources, trailers, schedules, etc.) so that transport orders (freight orders and freight bookings) can be performed by an internal or external carrier.

5.3.2. Using the decision framework on MoD processes

In this section, each process will be subjected to the criteria in the decision chart. First, the process is described briefly. After that, the decision chart is filled in and the score is calculated. And finally, an interpretation of the score is given. Starting point is always the model in ARIS. Sometimes, answers also come from personal experience and/or after questioning process experts. If this is the case, this is indicated in the interpretation. To increase the readability of this report, half of the investigated processes are moved to the appendix (Section A.5.3 to A.5.9). The processes described in this document are chosen because they give an accurate reflection of the test results. However, results of the other investigated processes are included in the conclusion.

On average, the fourteen processes scored 9,9 points with a minimum of -4 and a maximum of 15. The minimum possible is -13 and the maximum possible is 18.

Manage changes (RfC) (P.0.2.02)

The process coordinates the approval or disapproval of a Requests for Change (RfC). When someone requires an adjustment or a new functionality in SAP, the requestor can place a formal request. Multiple teams will test the request on several requirements before a decision is made. This process is supported by the SAP Solution Manager (SSM), a software package that supports the implementation and maintenance of SAP. Two of its most important modules concern change management and incident management. The data collected is mostly textual (reports, minutes, analysis, etc.) which is difficult to mine, however, these files carry metadata that can be used to follow the process. The process model is shown in Figure A.5.11 and the decision chart in Table 5.18.

Table 5.18: Manage changes (RfC) (P.0.2.02)

	Characteristic	**Result**	**Score**
1	Availability of the minimum required attributes	Yes	++
2	Number of roles involved in the process	13	+
3	Number of decision points	9	+
4	Number of activities	11	+
5	Lead time (average)	Months/years	+/-
6	Quality of dataset	High	+
7	Methods of importing data	Adapter needed or file-based	+/-
8	Possibility to filter on attributes	Yes, superficial	+/-
9	Nature of process	Office-administration	+
10	Process maturity (years of experience)	2011	+
11	Availability of process model	Yes, of high quality (EPC)	+
12	Mandated is known	Yes	++
	Result		**11**

It is a personnel intensive process and has a high number of roles. Process mining could be used to analyse the involvement of each role. A social network analysis and dotted chart could show if the process is delayed by a specific role. Adjusting the decision points may also lead to improvements. Explicitly in this process, decision points often lead to a go/no-go decision. The process should work as a funnel, removing as much as fruitless requests as fast as possible, leaving more resources for the more doubtful requests. Relocating the decision points can optimize the process. The lead time is relatively long, which negatively influences the score. The tool is only used by well-schooled personnel, which ensures a high-quality dataset. The fact that MoD does not have the means to export the data from SSM and import it in ARIS PPM, results in a lower score. The lack of filter possibilities also lowers the score. In conclusion, the score indicates that this process only moderately benefits from process mining. However, when MoD develops a direct connection between SSM and ARIS PPM, the score increases fast. Another recommendation can be the classification of RfCs on their goal. A RfC can for example intent to add an entry field in a form, make a certain action mandatory, or set up guidelines. With a systematic classification, the filter possibilities increase.

Create order (P.1.1.05)

When an order needs to be placed at an external supplier, this process should be followed. It is completely supported by SAP, but there are interfaces with three other software packages: "OMSKPU", "DAAS", and "NEMSIS". These packages will only be addressed at the end of the process and have no role of the event logging. The process model is shown in Figure A.5.12 and the decision chart in Table 5.19.

Table 5.19: Create order (P.1.1.05)

	Characteristic	**Result**	**Score**
1	Availability of the minimum required attributes	Yes	++
2	Number of roles involved in the process	16	+
3	Number of decision points	12	+
4	Number of activities	13	+
5	Lead time (average)	Hours	+
6	Quality of dataset	High	+
7	Methods of importing data	Off-the-shelf adapter available	++
8	Possibility to filter on attributes	Yes, profound	++
9	Nature of process	Office-administration	+
10	Process maturity (years of experience)	2016	+/-
11	Availability of process model	Yes, of high quality (EPC)	+
12	Mandated is known	Yes	++
	Result		**15**

With 15 points, this process scores high. The number of roles, decision points, and activities suggests significant improvement possibilities. The process is used by highly educated personnel who provide a good dataset. Reasons of concern, are the many possible ending (interfaces to other processes and other software packages) this model has. The number of possible endings is not included in the decision chart, but may negatively influence the process mining suitability. More possible endings can increase the number of unique process instances, making it more difficult to create a clear view with process mining. Here may lay an improvement possibility of the decision chart and this should be studied further.

Count stock and inventory (P.2.3.06)

Periodically, warehouse workers need to count their stock and inventory. The counts will be performed physically and the results are then entered from the relevant counting document. The process model is shown in Figure A.5.1 and the decision chart in Table 5.20.

Table 5.20: Count stock and inventory (P.2.3.06)

	Characteristic	Result	Score
1	Availability of the minimum required attributes	Yes	++
2	Number of roles involved in the process	3	-
3	Number of decision points	3	+/-
4	Number of activities	6	+/-
5	Lead time (average)	Hours	+
6	Quality of dataset	Low	-
7	Methods of importing data	Off-the-shelf adapter available	++
8	Possibility to filter on attributes	Yes, superficial	+/-
9	Nature of process	Office-administration	+
10	Process maturity (years of experience)	2006	+
11	Availability of process model	Yes, of high quality (EPC)	+
12	Mandated is known	Yes	++
	Result		**8**

The process revolves around two activates: count the stock and enter the data. A process simple as that does not benefit from process mining to improve it. This is also shown in the score. The minimum required attributes are available, but each case has four or less activities recorded (four being the longest flow possible). The model is so straightforward, little can be improved. Mistakes are easily made, reducing the quality of the dataset and filtering is not done based on weapon system or the like. The low score seems reasonable.

Set up and define organisational vision (P.3.1.01)

Every five years, MoD is required to draw up and establish a policy vision, which guides the Defence-wide planning and budgeting process. Every year, this vision needs to be adjusted to the governmental budget and the present conflicts in the worlds. This (brief) model shows the steps of the yearly adjustments. It is not supported by SAP, nor are there any roles defined or decisions to be made. The process model is shown in Figure A.5.13 and the decision chart in Table 5.21.

Table 5.21: Set up and define organisational vision (P.3.1.01)

	Characteristic	Result	Score
1	Availability of the minimum required attributes	No	--
2	Number of roles involved in the process	0	--
3	Number of decision points	0	-
4	Number of activities	4	+/-
5	Lead time (average)	Weeks/months	+/-
6	Quality of dataset	Low	-
7	Methods of importing data	None	-
8	Possibility to filter on attributes	No	-
9	Nature of process	Office-administration	+
10	Process maturity (years of experience)	2007	+
11	Availability of process model	Yes, of low quality (EPC)	+/-
12	Mandated is known	Yes	++
	Result		**-4**

The low score may come of no surprise. The model only consists of four activities combined with seven events. This process is modelled to show the relationship between objectives, activities, and resources in the overall management structure and has no intention to be supported by software. Even though the nature of this process is classified as office-administration, it may also have resemblances with the product development classification. It is a creative process with a low frequency and process mining cannot contribute to its improvement. The analysist has some liberty in filling in the decision chart and thus, this kind of discussion can arise. It is not seen as problematic.

Generate preventive maintenance need (P.4.3.02)

This process describes the creation of a maintenance requirement based on fixed time intervals or gauges reaching a predefined value (for example every 10.000 kilometre). The process includes four different activities, but its shortest flow only involves one. The process model is shown in Figure A.5.14 and the decision chart in Table 5.22.

Table 5.22: Generate preventive maintenance need (P.4.3.02)

	Characteristic	**Result**	**Score**
1	Availability of the minimum required attributes	Yes	++
2	Number of roles involved in the process	4	-
3	Number of decision points	3	+/-
4	Number of activities	4	+/-
5	Lead time (average)	Days	+
6	Quality of dataset	High	+
7	Methods of importing data	Off-the-shelf adapter available	++
8	Possibility to filter on attributes	Yes, profound	++
9	Nature of process	Office-administration	+
10	Process maturity (years of experience)	2006	+
11	Availability of process model	Yes, of high quality (EPC)	+
12	Mandated is known	Yes	++
	Result		**12**

Based on its simplicity, a score of 12 seems relatively high. The data may be of high quality and the process meet all other requirements, radically improving the process seems difficult, since there is almost nothing to change. Here, the decision chart seems to fall short.

Carry out (maintenance) work (P.4.3.30)

The process describes the execution of maintenance work. The process starts with an order and is completed when the order is given the status "execution ready". The process is generic and includes preventative, corrective and other maintenance, and recovery of recoverable spare parts. It is a complex model, fully supported by SAP, and has only four roles; the maintainer being the most important one. The process model is shown in Figure A.5.15 and the decision chart in Table 5.23.

Table 5.23: Carry out (maintenance) work (P.4.3.30)

	Characteristic	Result	Score
1	Availability of the minimum required attributes	Yes	++
2	Number of roles involved in the process	2	-
3	Number of decision points	25	+
4	Number of activities	32	+
5	Lead time (average)	Days/weeks	+
6	Quality of dataset	Low	-
7	Methods of importing data	Off-the-shelf adapter available	++
8	Possibility to filter on attributes	Yes, profound	++
9	Nature of process	Maintenance	+
10	Process maturity (years of experience)	2009	+
11	Availability of process model	Yes, of high quality (EPC)	+
12	Mandated is known	Yes	++
	Result		**12**

This process scores high points on almost every characteristic, except the number of roles and the data quality. The limited number of roles also limits the possibilities to carry out social analysis. Perhaps it is possible to compare productivity of different maintainers, but partly because of the complexity of the processes, it would be comparing apples and oranges. Experiences show that the datasets can be of low quality. Activities are reported at the end of the day, instead of the moment when they were actually finished. Also, small typing errors where often made. Using tablets in the field for directly logging activities may improve data quality. The possibility that cases can be filtered on their weapon system is very useful.

Monitor reliability (P.4.9.10)

Based on data from SAP, the maintenance manager can carry out an analysis on the reliability of a weapon system. The model guides the manager through the process. In this case, SAP serves as a data source and is not used to log the process of analysing. The outcome of the analysis is reported to SAP and may lead to actions that are supported by SAP. The process model is shown in Figure A.5.16 and the decision chart in Table 5.24.

Table 5.24: Monitor reliability (P.4.9.10)

	Characteristic	Result	Score
1	Availability of the minimum required attributes	Yes	--
2	Number of roles involved in the process	1	-
3	Number of decision points	8	+
4	Number of activities	13	+
5	Lead time (average)	Hours	+
6	Quality of dataset	High	+
7	Methods of importing data	Off-the-shelf adapter available	++
8	Possibility to filter on attributes	Yes, profound	++
9	Nature of process	Auditing	++
10	Process maturity (years of experience)	2014	+
11	Availability of process model	Yes, of high quality (EPC)	+
12	Mandated is known	Yes	++
	Result		**11**

This process may cause a confusing situation. Even though process mining can help the maintenance manager significant with his/her analysis, the process of monitoring reliability cannot be improved by process mining. This is because when the analysis is carried out, no event logs are recorded. Process mining can be beneficial, but then as a part of the analysis. This is not a fault in the decision chart. One needs to be aware of the difference, since on other aspects, the process scores high.

5.3.3. Results of the sample

An overview of the results of the study are placed in Table 5.25. Next to the scores from the decision chart, a verdict is given based on this score and on personal experience with the process. Since on average, the processes scored 10 points, this number is used as a benchmark. Scores below 9 then have a low potential, scores of 9, 10, and 11 a medium, and 12 and up a high potential. The second verdict is based on personal experience; its underpinning can be found in the text below each decision chart in Section 5.3.2.

Table 5.25: Results of the sample

Code	Name	Score	Potential based on score	Potential based on personal experience
P.0.2.02	Manage changes (RfC)	11	Medium	Medium
P.0.2.03	Manage incidents	12	High	High
P.0.3.05	Archive data	2	Low	Low
P.1.1.05	Create order	15	High	High
P.1.1.16	Simple purchasing	10	Medium	Medium
P.2.3.06	Count stock and inventory	8	Low	Low
P.2.4.02	Pick articles	13	High	High
P.3.1.01	Set up and define organisational vision	-4	Low	Low
P.3.3.03	Pay invoices	14	High	High
P.4.3.02	Generate preventive maintenance need	12	High	Medium
P.4.3.11	Technical work preparation	14	High	High
P.4.3.30	Carry out (maintenance) work	12	High	High
P.4.9.10	Monitor Reliability	11	Medium	High
P.5.2.02	Carry out transport planning	9	Medium	Medium

In just two occasions the decisions chart gave a score deviant of the researcher's verdict. This proves the usability of the chart. There were however, signs for improvement. Three new possible characteristics were discovered. First, the number of possible endings. These are the last events or process interfaces in the EPC. More endings, can give more unique process instances, giving unstructured process mining results. However, the number of possible endings was only on the high side. Second, the number of interfaces to other software can be taken into account. When multiple software packages are used during the process, it is difficult (or even impossible) to combine the event data of all software into a single log. In the sample, this was found in the process of carrying out transport planning (P.5.2.02). It is difficult to generalise this phenomenon since each interface can use different software that does or does not need to log events. More research is needed to decide upon influence. Third and last possible characteristic is the presence of loops. Again, loops can give unstructured process mining results. However, it rarely occurred. And thus, it will not be taken into account. Something that does require future research, are the categories of the natures of processes. For this, more cases are necessary.

5.4. Evaluation

During the implementation phase, the framework was tested in a real MoD environment and its results validate that the framework was developed according to the predefined requirements. Ideally, the framework should now be verified on its usage to determine whether its results are also correct. The best way to do so, is by starting process mining improvement projects with the models of the sample. Models with high scores should improve significantly, while models with low scores should not. However, these projects take months or even years to implement, making it impossible to evaluate them as part of this study. To cover for this shortcoming and to still try and proof its correctness, a method of evaluation is established. The method is explained in the first section. In the second section it is described how it was carried out, and in the last section the results are presented.

5.4.1. Method of evaluation

The evaluation of the framework is based on a semi-structured discussion with a panel of ten system experts (all from the DMO department Policy Participation). The framework should be able to predict the effects of a process mining project based on a process model. The panel has no experience with process mining projects, but it does have expert knowledge of the models on a theoretical and a practical level. The author, on the other hand, does have experience with process mining, but is less familiar with the implementation of the model in practice and its improvement potential. Knowledge and experience of both the panel and the author are combined to challenge the scores of the framework.

The evaluation is structured as follows:

1. The panel and researcher take place in a circle in a normal room.
2. The researcher describes the meaning of the discussion and explains that it consists of several rounds.
3. During each round, a process model is presented to the experts.
4. Each model is briefly explained by the researcher and he asks if the experts see improvement potential and if so, if they could indicate where this potential arises and elaborate on their reasons.
5. The panel discusses the model in an open atmosphere while arriving at a list of recommendations.
6. After the last model, the group is thanked for their anticipation and the researcher starts analysing the outcomes.
7. The researcher arranges the recommendations and tests if they are achievable by process mining.
8. If according to the panel a model shows significant improvement potential and if according to the process mining expert this potential is implementable by process mining, the framework should give a high score. If no potential is found or it not achievable with process mining, the framework should score low.

Because of time constrains, a set of four process models is chosen (as seen in Table 5.26). These models are already examined by the framework in Section 5.3.2. A model with a high, medium, and low score are chosen, also a model of which the score of the framework did not match the personal experience. They are discussed in the same (random) order as shown in the table.

Table 5.26: Process models used in the evaluation

Code	Name	Score	Potential based on score	Potential based on personal experience
P.0.2.02	Manage changes (RfC)	11	Medium	Medium
P.2.3.06	Count stock and inventory	8	Low	Low
P.4.3.02	Generate preventive maintenance need	12	High	Medium
P.1.1.05	Create order	15	High	High

5.4.2. Review of the brainstorm session

For each process, an summary of the discussion is given. The discussion is then analysed to estimate whether the ideas of the panel are achievable by process mining.

Manage changes (RfC) (P.0.2.02)

A first remark was made regarding the fact that the model was printed on two A3 sheets and yet, it was often still too small for reading. After guiding the panel through the model, an expert that works with this particular process noticed several differences between the formal model and the way the process is carried out. After that, a discussion started on the position of some of the activities. The question was whether rearranging some of them would speed up the process. The discussion did not lead to consensus on what would be the best arrangement, however, the panel did agree that there was improvement potential present. Some activities could not only be rearranged, several could also be combined and in some instances, adding quick gateway activities could prevent unnecessary activities later on. A question arose about the five different types of events at the start of the process (e.g. BV issue unsolvable, RFC necessary, etc.). The question was whether a different starting event would also lead to a different course of the process. The rest of the panel thought that the starting event was not logged and thus did not had any effect on the process. For this process, no KPIs are defined. It is thus not known if a case is carried out within the expected time. Another remark came from personal experience of one of the experts. He had experienced that the absence of some key role (by illness), made the whole process stagnate. Lastly, the whole panel agreed that the process was overly complex: "it will take ten hours of planning to divide five hours of work".

The panel addressed several aspects of the process that can be found and improved by process mining. These usually correspond with the aspects discussed in Section 5.3.2. For example, the high number of roles and relocating of the decision points. Still there are some difficulties. The fact that there are no KPIs defined makes it difficult to quantify improvements from a process mining project. Also, the panel supported the lack of filtering possibilities. The framework gives this model a score of 11. Based on the discussion with the panel, the score seems very plausible.

Generate preventive maintenance need (P.4.3.02)

It strikes the group that this process has very few roles in comparison with the previous process. While walking through the model, the panel could not come up with possible improvements. One of the experts explains that preventive maintenance is planned only when a request for it is given by the user. This is strange, since the long-term planning of preventive maintenance is fixed for years in advance and this should be backed by gauge readings (like millage of a vehicle) from SAP. The expert describes that the long-term planning is often neglected and that a significant proportion of the maintainers does not enter the gauge readings in SAP. It is noted that this is more a cultural problem than process-oriented. Equipping weapon systems and vehicles with sensors could automate the registration process and improve the planning as a whole.

When initiating a cultural change, process mining can help by providing information, showing the status quo, and monitoring the changes. However, for this process other business intelligence tools are more appropriate. Only activities in isolation have to be examined. For example, the number of times a meter reading is entered in SAP. Process mining could extract those data from SAP, but other tools or even a manual count with a spreadsheet can do this quicker. The decision framework gives this model a score of 12 points. This score is considered to be high, but based on personal experience, the researcher argued the potential to be medium. It is difficult to draw a conclusion from the evaluation, but it tends to medium as well.

Count stock and inventory (P.2.3.06)

The panel was confused on the role of technology in this process and it questioned whether the counting could be automated. After a discussion, the panel agreed that there is no alternative for its manual labour (the actual counting of the inventory). This makes the process more sensitive to (human) errors and reduces the logging possibilities by SAP. The panel also agreed that it is a complicated process. For example, the role of the blockades could not be explained. This also had to do with the fact that some activities were initiated in a prior process (P.2.3.05 Preparing stock and inventory count). Nevertheless, no concrete points of improvement were found.

The possibility of having multiple freestanding models that are bound to succeed each other, was not taken into account when developing the framework. The size of each model is a design choice. The organization could be modelled in one big process model or countless smaller process models (that are sometimes interdependent). It is inevitable that when several surrounding models are combined into one larger model, the outcome of the framework also changes. At the same time, the scope of the improvement project will grow, leveling out the effect. Still, more research is needed to confirm this. A score of 8 points was given to this model. This indicates little process mining potential. Since the panel was unable to come up with improvements, the score seems valid.

Create order (P.1.1.05)

This is again a complicated model, but this time the discussion quickly concentrated on the improvement of the process. The model shows several people acting as controllers, but the effect of these roles was questioned. More specifically, what do they actually do and how much work is it? According to personal experience of one of the experts, the controller's role in this process is more a formality than the described decision-making function. Another expert explains that even though the same role is described on several locations in the process, they all are unique in their tasks and responsibilities. The discussion continued on the possibility of rearranging some of the activities and later on the possibilities of combining some of them. Again, no consensus was reached on its ideal arrangement, but this process leaves room for improvement.

This model was given 15 points, which is high and indicates significant improvement potential. The discussion by the panel led to several questions that can be answered by process mining. For example, does the responsible supervision actually effect the process, or will process mining show that they forward all requests? This shows that the initial score was correct. The discussion also revealed a potential shortcoming in the way the processes are modelled. Namely that the different controllers performed different tasks and had different responsibilities, while being modelled as a single type of controller. Adding more details to the roles could improve this.

5.4.3. Results of evaluation

The method of evaluation was not ideal and its responds are subjective. Still, it can help with the evaluation of the decision framework. It was remarkable that the discussions during the first and last model were most dynamic. The panel was able to debate on potential improvements and these were in line with the possibilities of process mining. During the second discussion, improvement potential was also found. But this time it was also achievable by more simplistic tools (or even a spreadsheet). The third model was characterised by not being clarity regarding the role of technology. All in all, the evaluation shows significant resemblance with the scores of the framework.

Next to the evaluation of the framework, two other concerns came to light. According to one of the experts, the model of the first process does not correspond with how the process is executed in practise. This possibility was not taken into account during the development of the framework. However, it is questionable if it affects the framework and if so, how this can be taken into account. Second, the size of the model is design choice. When multiple models are combined, the outcome of the framework will change. It is expected that this effect is compensated because the scope of the improvement project will also grow. Still, more research is needed to confirm this.

6. Conclusion and Discussion

In this study, a methodology is developed to objectively classify process models from the Netherlands Ministry of Defence on their suitability of being analysed and improved by process mining as a business process improvement methodology. This study was carried out parallel to the Defence broad pilot Process Mining. The pilots goal was to learn more about process mining and to enthuse the organization for it. In this final chapter, a conclusion is given on both the outcome of the study and the course of the pilot. The chapter continuous with a reflection and several recommendations. Finally, the research limitations and opportunities for future research are discussed.

The goals of the pilot fall outside the scope of the thesis and are thus only briefly mentioned in this document. Still, experience and knowledge gained during the thesis, can help MoD with its pilot. That is why this chapter devotes attention to both the thesis and the pilot and perhaps reflects on material that is not explicitly described in the thesis. For MoD, a separate (internal) report will also comment on the pilot and give recommendations for using process mining.

6.1. Conclusion

As discussed in the introduction, the conclusion is divided into two parts. It starts with answering the research questions of the thesis and closes with commenting on the goals of the pilot.

6.1.1. Academic thesis

For the conclusion, the sub-questions of Section 2.3 are recalled. These questions need to be answered in order to answer the main research question.

Sub-question 1: What business process improvement methodologies are described in the literature?
In this study, Business Process Improvement methodology (BPI) was defined as: a collection of problem-solving methods, governed by a set of principles and a common philosophy, with the collective goal to improve another set of logically related tasks that use resources of the organization to provide defined results that support the organization's objectives. In the literature, many BPIs are described. Listing all of them is a challenging endeavour, since scholars and practitioners often combine the multiple BPIs to form new ones while interchanging the terms. The thesis discusses Lean, Six Sigma, Lean Six Sigma, TQM, Lean MRO, TPM, BPR, and BPI. They are characterised using various criteria, namely criticism, secondary effects, primary effects, methodologies, involvement, process view, theory, first mentioned, and origin.

Sub-question 2: How can process mining be used as a process analysis methodology?
Process mining is a promising business analysis methodology that uses event data to model processes on how they actually happened. Three purposes are distinguished: the discovery, conformance checking, and enhancement of process models. There are many ways to model and visualize these processes. This methodology of choice mainly depends on the data, goal of the process mining, and available software.

Sub-question 3: How can process mining be used as a business process improvement methodology?
On its own, process mining can assess processes and even suggest improvements, but a framework around it is necessary to guide the course and implementation of these improvements. Two methodologies are given extra attention: the L* life-cycle model (because of its popularity) and PM^2 (because of its structure). It was found that almost all discussed BPIs have a statistical or analytical background, but process mining excels in its ability to automatically convert data into

organized information. However, compared with the more mature BPIs, process mining lacks in its capabilities to implement improvements. The DMAIC cycle of Six Sigma, for example, shows a more structured approach for this step.

Sub-question 4: How does the Netherlands Ministry of Defence execute process mining projects in the Defence broad pilot Process Mining?
The thesis elaborates on PM², but while carrying out the Defence broad pilot Process Mining, room for improvement of this methodology was found. In combination with Six Sigma, PM² was revised to PM³, a methodology following the DMAIC-cycle, that embraces the philosophy of continuous improvement. This is shown in Figure 6.1.

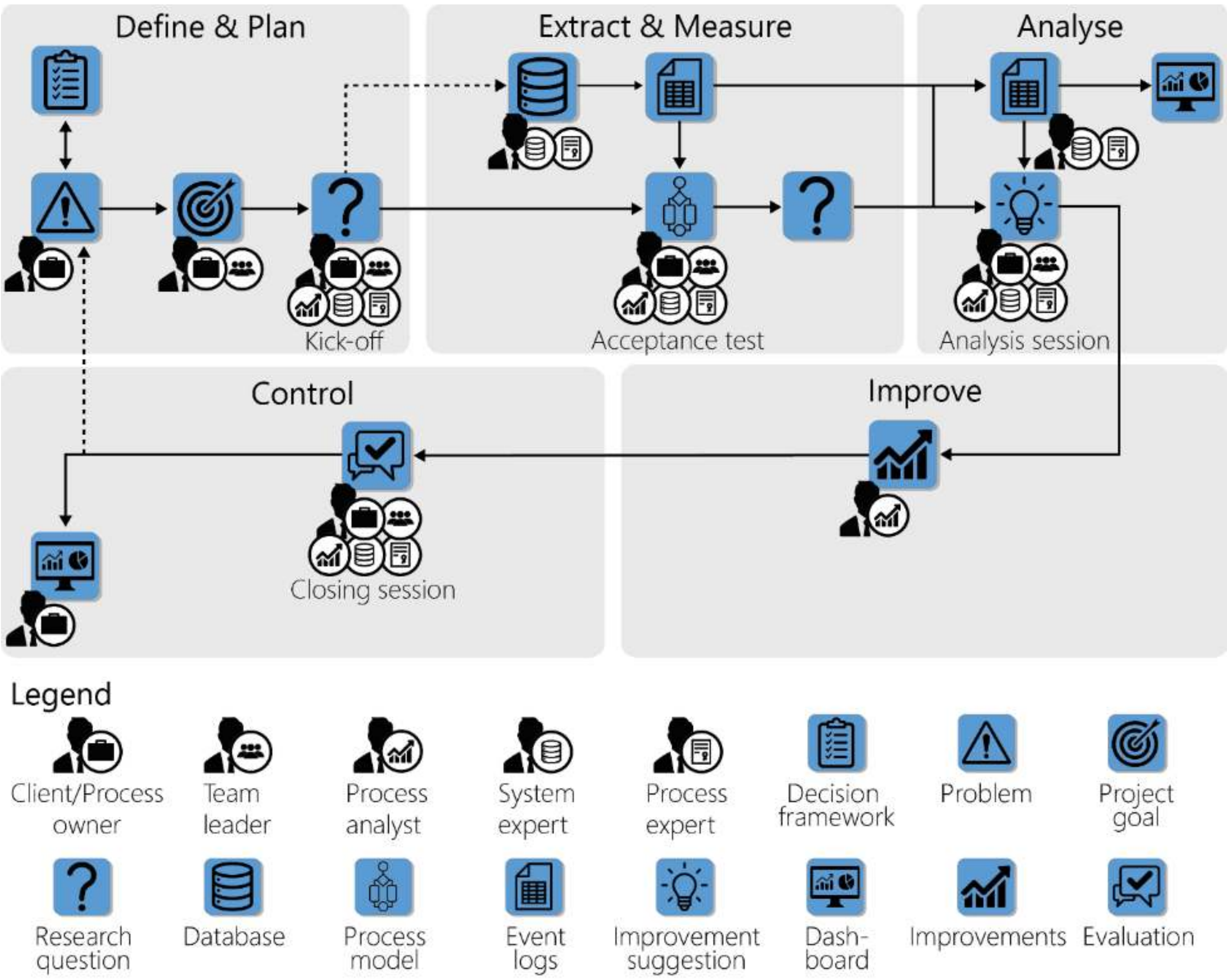


Figure 6.1: The PM³ model

Sub-question 5: What are the characteristics of processes at the Netherlands Ministry of Defence?
A conceptual business process model visualizes the generic business model and its attributes. MoD models its processes with Event-Driven Process Chains according to the ARIS-HOBE.

Sub-question 6: What processes at the Netherlands Ministry of Defence are suitable for process mining projects?
Twelve characteristics are described that can determine whether a process model of MoD is suitable for process mining as a BPI. These are: availability of the minimum required attributes, number of roles involved in the process, number of decision points, number of activities, lead time (average), quality of dataset, methods of importing data, possibility to filter on attributes, nature of process, process maturity (years of experience), availability of process model, and mandated is known. These characteristics are placed in a decision framework, where they all receives scores (as

shown in Table A.5.1). Lowest score possible is -12 and highest 17. A sample of fourteen actual process models of MoD was used to test the decision framework. On average, the processes scored 9,9 points. Based on their scores, the processes where labelled a to have a high, medium, or low process mining potential. These scores were in line with expectations.

Based on the answers to the sub-questions, attention can now be drawn to the main research question:
What kind of processes of the Netherlands Ministry of Defence can possibly benefit from using the process mining analysis method applied as a process improvement methodology?
Processes that benefit most from process mining, are the following. Processes that log their activities and these event logs contain at least a case ID, activity, and timestamp. They have a large number of roles involved, decision points, and activities. The lead time is short and the dataset is of high quality. It is beneficial if the dataset can be imported automatically into the process mining software (preferable with a standard adapter). During its analysis, having the possibility to filter on its attributes will improve the results. Preferably, the processes have a nature in manufacturing or auditing, are matured (has been applied for a number of years), the process models are available, and the persons how can implement changes, the mandated, are known.

6.1.1.1. Defence broad pilot Process Mining

MoD started the Defence broad pilot Process Mining for two reasons. First, to learn about the opportunities that process mining brings and on how MoD could utilize these. And second, to make the organization enthusiastic about process mining to accelerate the (possible) full-scale introduction after the pilot.

Even though the literature is reluctant in connecting case studies to quantitative results like cost reduction, it is unanimously positive on the potential of process mining. The literature discusses several aspects that influence the potential of process mining. Most important is the availability of high quality data. Several years ago, MoD integrated many of its processes and started using SAP to support these. Even though the quality of the resulting data was poor at start, but is getting significantly better every year, while even more processes are getting supported by SAP. This positive trend shows that the potential of process mining is still growing. And with the integrating of its processes, MoD started describing and centralizing its process models in the ARIS library. Already having process models available will speed up any improvement project. The literature also indicates that the success of process mining depends on the domain it is used in. For example, it could be specifically beneficial for manufacturing and auditing processes, while being useless for developing products. MoD is a large organisation and covers many domains, including the ones with a high potential. Lastly, MoD, with the DMO in specific, follows a matrix organisation structure. This proves useful when creating process mining teams. Employers are already accustomed to working in multidisciplinary teams from different departments. It thus can be concluded that process mining has significant potential.

Next to the gained knowledge, the thesis supports MoD in two ways. First, the decision framework helps to select the processes with the highest improvement potential. Second, the improved PM^3 methodology will improve all of MoD's process mining projects.

The second goal of the pilot is to make the organisation enthusiastic about process mining. During the course of the pilot, it received significant attention. Next to the treated cases (which involved all four armed services), several workshops and presentations were given. Many were reluctant at the start (asking about the added value in comparison to MoD's existing tools), but after learning about process mining they were eager to further pursue its potential.

6.2. Reflection and recommendations

Just like the conclusion, this section is divided into two parts. First a reflection and recommendations are given on the thesis, and thereafter on the pilot.

6.2.1. Academic thesis

When the thesis started, the research quickly focused on the improvement of process mining as a BPI, through executing and analysing a complete improvement project. It proved to be difficult to find a problem that was supported by data and also small enough to be examined by one person. An attempt was made to start with a dataset and then look for a problem. When there finally was data available, it first took days to get them in the format Disco required, to find out later that support from process experts was still needed to explain the mined models. This may be typical for the organisation, which can be labelled as quite cumbersome. After several fruitless attempts, the thesis gained a new direction: the development of a decision framework. The gained experience nevertheless proved useful.

During the research, various things stood out. For example, the ambiguity that exists among the many BPI's, while very little literature can be found that pursues to compare them. The presented overview contributes to this, but still more research is needed to truly gain insights on this topic. Second, it is remarkable that process mining is presented as a revolutionary instrument to improve processes, but very little attention is given on how to execute a complete project. One of the few methodologies available (PM^2) shows to be highly software depended. Several improvements have been made, resulting in PM^3. The developed decision framework helps with the consideration of using process mining. Based on the sample in the thesis, in half the cases the framework would give a negative advice. This illustrates the importance of such a framework.

Overall, process mining still seems to be a very valuable improvement methodology. The coming years, the software, data, and methodology will mature to become even more useful.

6.2.2. Defence broad pilot Process Mining

When the pilot started, MoD had very limited experience and knowledge on process mining. The topics of the five cases did not all reflect true process mining issues, but this did not withhold the team to begin with full dedication.

Already during the first case, several challenges arose. The goal was to determine whether maintenance personal was executing tasks only at or below the level of complexity it was certified for. After careful consideration, the conclusion was drawn that this case fell in the domain of data mining and not process mining (it was about the combination of two attributes in isolation and not a whole process). The topic was changed to the maintenance process. Soon thereafter, it became clear that a well scoped research question and support from management are essential. This led to a different project management approach and to PM^3. The second case aimed at finding out whether preventive maintenance tasks were adequately planned ahead. Experiences from the first case further embedded the procedure into the team, resulting in a smooth workflow. Indicating a steep learning curve. Even though the results of the analysis showed that very few tasks were planned, it still is unsure if the results will lead to actual changes. This illustrates another point of attention. The organisation should be willing to change. The incentive of the third case was that several logistic processes were being integrated. Project mining is deployed to visualize the integration of the old process models to the new, uniform, model. By now, the core process mining group is able to quickly guide the client and rest of the experts through the methodology. On the one hand, it is tiring to repeat the same story over and over, but on the other hand, the enthusiasm for process mining that it generates is motivating. At the time of writing, the last two cases are still ongoing.

For MoD, process mining can be a valuable instrument for improving processes. It can convert a gut feeling into clear numbers and models, which can form the start of an improvement project. It is thus recommended to make process mining available to the organisation. The PM^3 and decision framework can be of great value. The decision framework still requires several rounds of field testing before its operation can be determined with certainty. However, even in its current form, the framework will encourage people to think about the requirements for process mining, which by definition is good. When it is fully tested, it will help MoD order its improvement projects on their potential. This reduces costs and sequentially streamline MoD's logistics.

6.3. Research limitations

The literature collected gives a good insight in the possibilities of process mining. And the decision framework can assist MoD in the valuing of processes whether process mining is a suitable BPI. However, the results are mostly based on desk research. This is because the limitations in the possibilities of testing the models on real life improvement projects.

The decision framework developed in this study is tested by applying it on actual process models of MoD. Overall, the framework resulted in scores that were in line with the expectations of the author and a panel of system experts, but the processes were not actually, i.e. empirically, tested by using them in concrete process mining projects. This, of course, would be unrealistic, since a single process mining project can take weeks or even months, let alone fourteen projects. However, an empirical test would give valuable results that can test and improve and thus contribute to the validity of the (theoretical) framework. The selection of characteristics chosen here can be improved, just like the turning points of some characteristics (for example, to what degree are roles involved in the process?), the scores given to the characteristics, and the overall turning points when a process is suitable for mining and when not. Ideally, a dataset of process mining projects at MoD is created, but if this is not feasible another testing method is needed.

6.4. Future research

The thesis gives several opportunities for future research. First, the overview of nine BPIs can be further supplemented with more BPIs and properties. And the conceptual framework of a business process can be further developed to be used for other research as well. But most potential can be found in PM^3 and the decision framework.

The PM^3 methodology is never final. It is used during several cases, but these cases still have not yet led to measurable improvements (only important insights in the processes). Following these cases for a longer period can give new insights in the methodology. Also, the cases focus on MRO within MoD. It is unlikely that a change of domain or organisation makes a significant difference, but the possibility should be taken into account and further investigated.

For its designing and its testing, the decision framework only considers processes of MoD. The level of aggregation is of such a level that the framework should work on other models as well. However, this needs to be tested. An option would be to use the case studies described in the literature as a source. Also, during the implementation of the framework three new possible characteristics were discovered. Namely the number of possible endings, the number of interfaces to other software, and the presence of loops. Future researchers can look into its effects on a process mining project and whether they should be added to the framework.

During the evaluation of the framework, it came to light that the modeller had the possibility to combine or split its process models and therewith change its size. When several surrounding models are combined to one larger model, the outcome of the framework also changes. At the same time, the scope of the improvement project will change, evening out the effect. Still, more research is needed to confirm this.

A significant limitation of the framework is that it only scores on process mining. So, if it presents a very low score, it does not suggest an alternative BPI. This fell outside the scope of the thesis, but is nevertheless an interesting research opportunity.

To the author's opinion, the current greatest research opportunity lays in the combination of using process mining results with simulation models. Data provided by process mining covers all activities of a process: its throughput, lead times, its delays, etc. In simulation software, these data can be used for building an accurate model. The simulations are particular interested for supply chain management, where by reasoning, but also by trial-and-error, elements in the simulation can be changed till the model is improved till it is (almost) perfect.

References

All images with no source mentioned are produced by the author.

Adesola, S., & Baines, T. (2005). Developing and evaluating a methodology for business process improvement. *Business Process Management Journal, 11*(1), 37-46.

Agrawal, R., Gunopulos, D., & Leymann, F. (1998). Mining Process Models from Workflow Logs. *6th International Conference on Extending Database Technology* (pp. 469-483). Valencia, Spain: Springer-Verlag Berlin Heidelberg.

Ailenei, I.-M. (2011). *Process Mining Tools: A Comparative Analysis.* Eindhoven, the Netherlands: Eindhoven University of Technology.

Al-Aomar, R., Aljeneibi, S., & Almazroui, S. (2016). Reducing Operational Downtime in Service Processes: A Six Sigma Case Study. *2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)* (pp. 63-67). Korea: IEEE.

Algemene Rekenkamer. (2016). *Resultaten verantwoordingsonderzoek 2015 - Ministerie van Defensie (X).* Den Haag: Algemene Rekenkamer.

Al-kaabi, H., Potter, A., & Naim, M. (2007). An outsourcing decision model for airlines' MRO activities. *Journal of Quality in Maintenance Engineering, 13*(3), 217-227.

Andersson, R., Eriksson, H., & Torstensson, H. (2006). Similarities and differences between TQM, Six Sigma and Lean. *The TQM Magazine, 18*(3), 282-296.

Aversano, L., & Tortorella, M. (2015). A Review of Enterprise Modelling Studies. *17th International Conference on Enterprise Information Systems* (pp. 346-351). Barcelona, Spain: SCITEPRESS.

Ayeni, P., Ball, P., & Baines, T. (2016). Towards the strategic adoption of Lean in aviation Maintenance Repair and Overhaul (MRO) industry: An empirical study into the industry's Lean status. *Journal of Manufacturing Technology Management, 27*(1), 38-61.

Baan. (1996). *BAAN IV Orgware: Dynamic Enterprise Modeling.* Barneveld, the Netherlands: Baan Development.

Bakri, A. H., Rahim, A., Yusof, N., & Ahmad, R. (2012). Boosting Lean Producion via TPM. *Procedia - Social and Behavioral Sciences, 65*, 485-491.

Baluch, N., Abdullah, C. S., & Mohtar, S. (2012). TPM and Lean maintenance - A critical review. *Interdisciplinary Journal Of Contemporary Research In Business, 4*(2), 850-857.

Becker, R. M. (1998). *Lean Manufacturing and the Toyota Production System.* Retrieved October 13, 2016, from SAE International: http://www.sae.org/manufacturing/lean/column/leanjun01.htm

Bose, J. C., van der Aalst, W. M., Žliobaitė, I., & Pechenizkiy, M. (2011). Handling Concept Drift in Process Mining. *23rd International Conference on Advanced Information Systems Engineering* (pp. 391-405). London, UK: Springer-Verlag Berlin Heidelberg.

Bose, J. C., van der Aalst, W. M., Žliobaitė, I., & Pechenizkiy, M. (2014). Dealing With Concept Drifts in Process Mining. *Ieee Transactions On Neural Networks And Learning Systems, 25*(1).

Bose, J. R., Mans, R. S., & van der Aalst, W. M. (2013). Wanna improve process mining results? It's high time we consider data quality issues seriously. *4th IEEE Symposium on Computational Intelligence and Data Mining* (pp. 127-134). Singapore: IEEE Inc.

Bozkaya, M., Gabriels, J., & Werf, J. M. (2009). Process Diagnostics: a Method Based on Process Mining. *2009 International Conference on Information, Process, and Knowledge Management* (pp. 22-27). Cancun, Mexico: IEEE.

Burattin, A. (2015). *Process Mining Techniques in Business Environments.* Innsbruck, Austria: Springer-Verlag Berlin Heidelberg.

Carmona, J., & Gavaldà, R. (2012). Online Techniques for Dealing with Concept Drift in Process Mining. *11th International Symposium on Intelligent Data Analysis* (pp. 90-102). Helsinki, Finland: Springer-Verlag Berlin Heidelberg.

Celonis SE. (n.d.). *Product*. Retrieved March 27, 2017, from Celonis: http://www.celonis.com/en/product/

Childe, S., Maull, R., & Bennett, J. (1994). Frameworks for Understanding Business Process Re-engineering. *International Journal of Operations & Production Management, 14*(12), 22-34.

Cook, J. E., & Wolf, A. L. (1995). Automating Process Discovery through Event-Data Analysis. *17th International Conference on Software Engineering* (pp. 73-82). Seattle, USA: ACM.

Datta, A. (1998). Automating the Discovery of AS-IS Business Process Models: Probabilistic and Algorithmic Approaches. *Information Systems Research, 9*(3), 275-301.

Davenport, T. H. (1992). *Process Innovation: Reengineering Work through Information.* Boston, USA: Harvard Business Review Press .

de Jong, S. J., Wouter, W., & Van Blokland, B. (2016). Measuring lean implementation for maintenance service companies. *International Journal of Lean Six Sigma,, 7*(1), 35-61.

de Weerdt, J., Schupp, A., Vanderloock, A., & Baesens, B. (2013). Process Mining for the multi-faceted analysis of business processes—A case study in a financial services organization. *Computers in Industry, 64*, 57-67.

Dietz, J. L. (2001). DEMO: Towards a discipline of organisation engineering. *European Journal of Operational Research, 128*, 351-363.

Doumeingts, G., & Ducq, Y. (2001). Enterprise modelling techniques to improve efficiency of enterprises. *Production Planning & Control, 12*(2), 146-163.

Earl, M. J. (1994). The New and the Old of Business Process Redesign. *The Journal of Strategic Information Systems, 3*(1), 5-22.

Eriksson, H.-E., & Penker, M. (1998). *Business Modeling With UML: Business Patterns at Work.* New York, USA: John Wiley & Sons, Inc.

Es-Soufi, W., Yahia, E., & Roucoules, L. (2016). On the Use of Process Mining and Machine Learning to Support Decision Making in Systems Design. In R. Harik, L. Rivest, A. Bernard, B. Eynard, & A. Bouras, *Product Lifecycle Management for Digital Transformation of Industries* (pp. 56-66). Cham, Switzerland: Springer International Publishing AG.

Fei, D., Liqun, Z., GuangYun, N., & Xiaolei, X. (2013). A Algorithm for Detecting Concept Drift Based on Context in Process Mining. *4th International Conference on Digital Manufacturing & Automation* (pp. 5-8). Qingdao, China: Curran Associates, Inc., New York, USA.

Ferguson, D. (2007). Lean and six sigma: The same or different? *Management Services, 51*(3), 12-13.

Gartner, Inc. (2016). *Automated Business Process Discovery.* Stamford, USA: Gartner, Inc.

Gershon, M. (2010). Choosing Which Process Improvement Methodology to Implement. *Journal of Applied Business & Economics, 10*(5), 61-68.

Gonella, P. (2016). *Process mining: A database of applications.* Retrieved March 12, 2017, from HSPI Management Consulting: http://www.hspi.it/

Günther, C. W., & van der Aalst, W. M. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. *Business Process Management 2007* (pp. 328-343). Brisbane, Australia: Springer-Verlag Berlin Heidelberg.

Hammer, M. (1990). Reengineering Work: Don't Automate, Obliterate. *Harvard Business Review, 68*(4), 104-112.

Hammer, M., & Champy, J. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution.* London, UK: Nicholas Brealey Publishing.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining.* London, England: The MIT Press.

Harrington, H. (1991). *Business Process Improvement: The Breakthrough Strategy for Total Quality, Productivity, and Competitiveness.* New York, USA: McGraw-Hill Education.

Harry, M. J. (1998). Six Sigma: A Breakthrough Strategy for Profitability. *Quality Progress, 31*(5), 60-64.

Herbst, J. (2000). A Machine Learning Approach to Workflow Management, A Machine Learning Approach to Workflow. *11th European Conference on Machine Learning* (pp. 183–194). Barcelona, Spain: Springer-Verlag Berlin Heidelberg.

Hicks, B. J. (2007). Lean information management: Understanding and eliminating waste. *International Journal of Information Management, 27*, 233-249.

Holmes, K. (1992). *Total Quality Management.* Leatherhead, United Kingdom: Pira International.

Hompes, B., Buijs, J., van der Aalst, W. M., Dixit, P., & Buurman, J. (2017). Detecting changes in process behavior using comparative case clustering. *5th International Symposium on Data-Driven Process Discovery and Analysis* (pp. 54-75). Vienna, Austria: Springer-Verlag Berlin Heidelberg.

IEEE Computational Intelligence Society. (2016). *IEEE Std 1849-2016 - IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams.* New York, USA: IEEE Inc.

Jäger, D., Schleicher, A., & Westfechtel, B. (1999). Using UML for Software Process Modeling. *7th ACM SIGSOFT international symposium on Foundations of software engineering* (pp. 91-108). Toulouse, France: Springer-Verlag London.

Kanji, G. K. (1990). Total quality management: the second industrial revolution. *Total Quality Management, 1*(1), 3-12.

Kebede, M. (2015). *Comparative Evaluation of Process Mining Tools.* Tartu, Estonia: University of Tartu.

Kettinger, W. J., Teng, J. T., & Guha, S. (1997). Business Process Change: A Study of Methodologies, Techniques, and Tools. *MIS Quarterly, 21*(1), 55-98.

Kooij, P., & Rozinat, A. (2016, July 26). *Automation Platforms and Process Mining: A Powerful Combination.* Retrieved July 31, 2017, from Fluxicon: http://coda.fluxicon.com/assets/downloads/Articles/Automation-And-Process-Mining_A-Powerful-Combination.pdf

Krafcik, J. F. (1988). Triumph of the lean production system. *MIT Sloan Management Review, 30*(1), 41-52.

Lean Enterprise Institute. (2006). *Lean.org.* Retrieved October 11, 2016, from A Brief History of Lean: http://www.lean.org/WhatsLean/History.cfm

List, B., & Korherr, B. (2006). An Evaluation of Conceptual Business Process Modelling Languages. *21st ACM Symposium on Applied Computing* (pp. 1532 - 1539). Dijon, France: ACM.

Luengo, D., & Sepúlveda, M. (2012). Applying Clustering in Process Mining to Find Different Versions of a Business Process That Changes over Time. *Business Process Management Workshops 2011* (pp. 153-158). Clermont-Ferrand, France: Springer-Verlag Berlin Heidelberg.

Manuel, A. (2012, February 7). *Process Mining: Ana Aeroportos de Portugal.* Retrieved July 31, 2017, from BPTrends: http://www.bptrends.com/process-mining-ana-aeroportos-de-portugal/

Mertins, K., & Jochem, R. (2000). Integrated Enterprise Modeling. In A. Rolstadas, & B. Andersen, *Enterprise Modeling* (Vol. 560, pp. 309-318). Boston, USA: Springer-Verlag Berlin Heidelberg.

Nakajima, S. (1989). *TPM development program: implementing total productive maintenance.* Cambridge, Mass: Productivity Press.

Object Management Group. (2015, June). *Unified Modeling Language*. Retrieved July 17, 2017, from Object Management Group: http://www.omg.org/spec/UML/

Pawlak, A., Rosienkiewicz, M., & Chlebus, E. (2017). Design of experiments approach in AZ31 powder selective laser melting process optimization. *Archives of Civil and Mechanical Engineering, 17*(1), 9-18.

Rebuge, Â., & Ferreira, D. R. (2012). Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems, 37*(2), 99-116.

Rinke, A. (2017, March 5). *How Process Mining, AI And Machine Learning Can Transform Manufacturing Operations.* Retrieved July 28, 2017, from Manufacturing Business Technology: https://www.mbtmag.com/article/2017/05/how-process-mining-ai-and-machine-learning-can-transform-manufacturing-operations

Riss, J. O., James, L. T., & Thorsteinsson, U. (1997). A situational maintenance model. *International Journal of Quality & Reliability Management, 14*(4), 349-366.

Röglinger, M., Pöppelbuß, J., & Becker, J. (2012). Maturity models in business process management. *Business Process Management Journal, 18*(2), 328-346.
Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics, 61*, 224-236.
Royce, W. (1987). Managing the Development of Large Software Systems. *9th international conference on Software Engineering* (pp. 328-338). Monterey, USA : IEEE Computer Society Press.
Rozinat, A., Mans, R., Song, M., & van der Aalst, W. (2009). Discovering Simulation Models. *Information Systems, 4*(3), 305-327.
SAP SE. (n.d.). *Home*. Retrieved March 2, 2017, from SAP Software Solutions: http://www.sap.com/index.html
Sayer, N. J., & Williams, B. (2012). *Lean For Dummies.* Hoboken, USA: John Wiley & Sons, Inc.
Scheer, A.-W. (1994). ARIS Toolset: A Software Product is Born. *Information Systems, 19*(8), 607-624.
Scheer, A.-W., & Nüttgens, M. (2000). ARIS Architecture and Reference Models for Business Process Management. In W. v. Aalst, J. Desel, & A. Oberweis, *Business Process Management* (pp. 376-389). Heidelberg, Germany: Springer-Verlag Berlin Heidelberg.
Scheer, A.-W., Abolhassan, F., Jost, W., & Kirchmer, M. (2004). *Business Process Automation - ARIS in Practice.* Berlin, Germany: Springer-Verlag Berlin Heidelberg.
Sester, D. (2001, October). Motorola: A Tradition of Quality. *Quality, 40*(10), pp. 30-34.
Sheth, J. N. (1973). A Model of Industrial Buyer Behavior. *Journal of Marketing, 37*(4), 50-56.
Software AG. (n.d.). *Software AG*. Retrieved March 29, 2017, from ARIS Business Process Analysis (BPA): http://www2.softwareag.com/corporate/products/aris_alfabet/bpa/default.aspx
Sugimori, Y., Kusunoki, K., Cho, F., & Uchikawa, S. (1977). Toyota production system and Kanban system Materialization ofjust-in-time and respect-for-human system. *International Journal of Production Research, 15*(6), 553-564.
Suriadi, S., Andrews, R., ter Hofstede, A., & Wynn, M. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems, 64*, 132-150.
Thomas, A., Barton, R., & Chuke-Okafor, C. (2009). Applying lean six sigma in a small engineering company - a model for change. *Journal of Manufacturing Technology Management, 20*(1), 113-129.
van der Aalst, W. M. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes.* Berlin, Germany: Springer-Verlag Berlin Heidelberg.
van der Aalst, W. M. (2016). *Process Mining - Data Science in Action.* Berlin, Germany: Springer-Verlag Berlin Heidelberg.
van der Aalst, W. M., & Weijters, A. J. (2004). Process mining: a research agenda. *Computers in Industry, 53*(3), 231-244.
van der Aalst, W. M., Adriansyah, A., de Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., . . . Wynn, M. (2012). Process mining manifesto. *Business Process Management Workshops 2011* (pp. 169-194). Clermont-Ferrand, France: Springer-Verlag Berlin Heidelberg.
van der Aalst, W. M., Reijers, H., Weijters, A., van Dongen, B., Medeiros, A. A., Song, M., & Verbeek, H. (2007). Business process mining: An industrial application. *Information Systems*(32), 713-732.
van der Aalst, W. M., Weijters, A., & Maruster, L. (2004). Workflow Mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering, 16*(9), 1128–1142.
van Dongen, B. F., & van der Aalst, W. M. (2004). Multi-Phase Process Mining: Building Instance Graphs. *23rd International Conference on Conceptual Modeling* (pp. 362-376). Shanghai, China: Springer-Verlag Berlin Heidelberg.
van Dongen, B. F., & van der Aalst, W. M. (2005). A Meta Model for Process Mining Data. *17th International Conference on Advanced Information Systems Engineering* (pp. 309-320). Porto, Portugal: FEUP Edições.

van Eck, M. L., Lu, X., Leemans, S. J., & van der Aalst, W. M. (2015). PM^2: a Process Mining Project Methodology. *27th International Conference on Advanced Information Systems Engineering* (pp. 279-313). Stockholm, Sweden: Springer-Verlag Berlin Heidelberg.

Verstraete, D. (2014). *Process Mining in Practice: Comparative Study of Process Mining Software.* Ghent, Belgium: Universiteit Gent.

Weber, P., Bordbar, B., & Tiňo, P. (2011). Real-Time Detection of Process Change using Process Mining. *1st Imperial College Computing Student Workshop* (pp. 108-114). London, UK: Imperial College London.

Weijters, A. J., & van der Aalst, W. M. (2003). Rediscovering Workflow Models from Event- Based Data using Little Thumb. *Integrated Computer-Aided Engineering, 10*(2), 151–162.

Weske, M. (2012). *Business Process Management.* Potsdam, Germany: Springer-Verlag Berlin Heidelberg.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The Machine That Changed The World.* New York, N.Y.: Macmillin Publishing Company.

Zhang, Z., Liu, G., Jiang, Z., & Chen, Y. (2015). A Cloud-Based Framework for Lean Maintenance, Repair, and Overhaul of Complex Equipment. *Journal of Manufacturing Science and Engineering, 137*(4), 04090801-04090811.

Ministerie van Defensie

TU Delft