UNIVERSITY OF TURKU

ABSTRACT

- [ ] Bachelor's thesis
- [x] Master's thesis
- [ ] Licentiate's thesis
- [ ] Doctoral dissertation

Subject	Information Management	Date	31.05.2022
Author(s)	Laura Arias	Number of pages	78 + 4 appendices
Title	Initiating a continuous process mining implementation.		
Supervisor(s)	Ph.D. Poonacha Medappa		

Organizations are facing difficulties to initiate a continuous Process Mining implementation. It is well established that Process Mining helps organizations to improve their business processes through data insights rather than subjective opinions. As the Process Mining technology is becoming mature, the managerial side of the disciple requires more attention. Thus, this thesis investigates how to initiate a continuous Process Mining implementation. The aim of this research is to design a guideline that operationally supports organizations in initiating their implementation.

To create the artifact, the Design Science research method was utilized. A first version of the guideline was created by collecting and analyzing data from User interviews. The participants were selected because they actively participated in their company's Process Mining implementation. To confirm the guideline, Validation interviews were conducted with Experts in Process Mining. These Process Mining managers, consultants, and doctors inputted the guideline with metrics and operational example and validated it.

The results showed 3 important pillars for initiating a continuous Process Mining implementation: the Process-, the Data-, and the Organizational- Pillars. It was found that for a successful implementation, the pillars must synergize. To support this journey, the IPMI guideline (Initiating Process Mining Implementation), provides activities and operational tools.

Another interesting finding of this research is the importance of the Organizational Pillar. This broad term encompasses various activities such as vision and strategy definition, change management, project team creation, and project governance. The implication of this finding is that the Organizational pillar leverages companies' ability to transform a small-scale implementation pilot project into a company-wide Process Mining capability.

This guideline technically and socially embeds the Process Mining capability in organizations.

Keywords	Process Mining, implementation, Data Pillar, Process Pillar, Organizational Pillar, DIDOV.

INITIATING A CONTINUOUS PROCESS MINING IMPLEMENTATION

An operational guideline for a successful implementation

Master's Thesis in
Information Management

Author(s):
Laura Arias

Supervisors:
Ph.D. Poonacha Medappa
Prof. Dr. Hannu Salmela

Company supervisors:
Dimphy Beaujean
Ana Lagoa

31.05.2022
Amsterdam

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE OF ABBREVIATIONS

ADKAR	Awareness-Desire-Knowledge-Ability-Reinforcement
BPM	Business Process Management
DFSS	Design for Six Sigma
DIDOV	Define-Identify-Design-Optimize-Verify
DIDVO	Define-Identify-Design-Verify- Optimize
DMAIC	Define-Measure-Analyse-Improve-Control
IoC	Internet of Content
IoE	Internet of Events
IoL	Internet of Locations
IoP	Internet of People
IoT	Internet of Things
IPMI methodology	Initiating Process Mining Implementation
KPI	Key Performance Indicators

NLP	Natural Language Processing
OPR	Optical Character Recognition
Philips DA	Philips Domestic Appliances
PPMS	Process Mining for Six Sigma
SIPOC	Supplier, Input, Process, Output, Customer
XES	eXtensible Event Stream

1 INTRODUCTION

This chapter introduces the Process Mining state of the art and highlights the motivations for this thesis. The motivations for this research originated in Philips Domestic Appliances' intention to initiate a Process Mining implementation. Therefore, the background of the company be presented. This chapter will be concluded by the announcement of the research questions and the structure of the research.

1.1 Background

Process Mining is a young and growing research discipline in the Big Data field (IEEE TASK FORCE ON PROCESS MINING, 2012). According to Professor Wil van der Aalst, the godfather of the discipline, *"Process Mining is becoming a more mainstream activity. In countries like Germany and The Netherlands, most of the larger organizations have applied Process Mining"* (Gartner, 2020). Process Mining offers a complete set of tools to generate insights founded on facts and, supports **process improvement** (van der Aalst, 2016). According to Gartner's Market Guide for Process Mining (2020), Process Mining focuses on the **as-is processes** as opposed to the intended process.

As an illustrative example, Figure 1 displays a Process Mining on a Purchase Order process.

Figure 1: Example of Process Mining Purchase Order

1.1.1 Problem Statement

Wil van der Aalst explains the reasons for the paradoxical slow adoption of Process Mining in an interview for Gartner: *"the technology is accessible for any organization. What should improve is the scale of adoption in organizations. […] Data quality and people remain the two main hurdles for widespread adoption"* (Gartner, 2020). To demonstrate a further understanding of this statement, one needs to compare the theory to the reality. It is worth noting that in research assumptions are made so the reality is easier to understand and model. On the one hand, in the reality, a Process Mining analysis is performed *"by extracting knowledge from event logs readily available in today's systems"* (van der Aalst, 2011). On the other hand, in the literature, the assumption is made that at each step of the process it is possible to collect enough of those event logs from the information systems. In other words, it is assumed that the minimal event logs requirements are systematically met (Weijters, 2004). However, in the reality, for most organizations, **data is not always integrated into the Information System, nor is easily accessible**. Even for organizations where the data is present, it can be **incomplete** or **hidden** among a massive amount of other information. Hence the choice of a suitable process to get started with Process Mining becomes harder. Weijters (2004) explains the challenge that most organizations are facing when trying to implement Process Mining: *"Process Mining generates a number of scientific and practical challenges (e.g., which processes can be discovered and how much data is needed to provide useful information)"*.

The existing literature offers incomplete guidance in this direction. This challenge represents an opportunity for further investigation in an appropriate manner.

1.1.2 Company Introduction

This research is conducted as part of the researcher's master's degree internship at Philips Domestic Appliances. The company agreed to take part in this thesis. In 2021, Philips DA disentangled from Royal Philips who already implemented Process Mining (van der Aalst, 2016). The ex-subsidiary benefits from clear Business Processes as the spine of its activity. As part of the Domestic Appliances Exuviate Program (i.e., guideline, strategy, and planning for the disentanglement), Philips DA aims to become digital-first: *"The first important step of this transformation is to design, build and implement the digital core including new technologies and process simplifications, enabling to deliver an exceptional service to our consumers daily"* (Philips Domestic Appliances,

2021). As per this statement, the interest in Process Mining became evident when Dimphy Beaujean, the supervisor of this thesis at Philips DA, described her vision during the hiring interview of the author: *"We would like to turn raw data into information and knowledge, to improve our business processes"*. During the second hiring interview, the author of this thesis introduced Dimphy Beaujean's department (Global Operational Excellence) to Process Mining. The peers agreed that Process Mining would be a suitable and beneficial tool, **but they did not feel ready nor knew how to start the technical implementation**. The need for a guideline on how to initiate the implementation of Process Mining emerged.

In Philips DA's Integrated Supply Chain, the sub-department Operational Excellence is leading a Coffee Excellence program dedicated to the improvement of the manufacturing processes. As part of this initiative, Philips DA is thinking about implementing Process Mining. This study aims at providing a guideline for Philips DA and for other companies who want to initiate a continuous Process Mining implementation.

1.2 Research Motivations and relevance

To evaluate the aim of the research, an emphasis on the relevance is required. According to the Cambridge Advanced Learner's Dictionary and Thesaurus (2022), relevance is defined as *"the degree to which something is related or useful to what is happening or being talked about"*. In research from Priniski (2017), relevance is described as *"the value a task has for an individual because it is useful for achieving current or future goals"* (p. 11–29). This definition can be extended to the scientific relevance (whether this research will improve the knowledge base and fill a gap in the existing literature) and to the business relevance (whether this research can be transposed to real-life settings to better understand the society, solve a real-world problem, and add value to businesses). This study contains 3 main motivations that will be presented in this chapter.

1.2.1 Worldwide growth of the Process Mining market

The Process Mining market is estimated to **grow with a Compound Annual Growth Rate of 49.33% from 2021 to 2028** (Fortune Business Insights, 2021). Process Mining changes the way companies are facing business process improvements. As an important emphasis is made on the benefits of Process Mining, this discipline is growing fast.

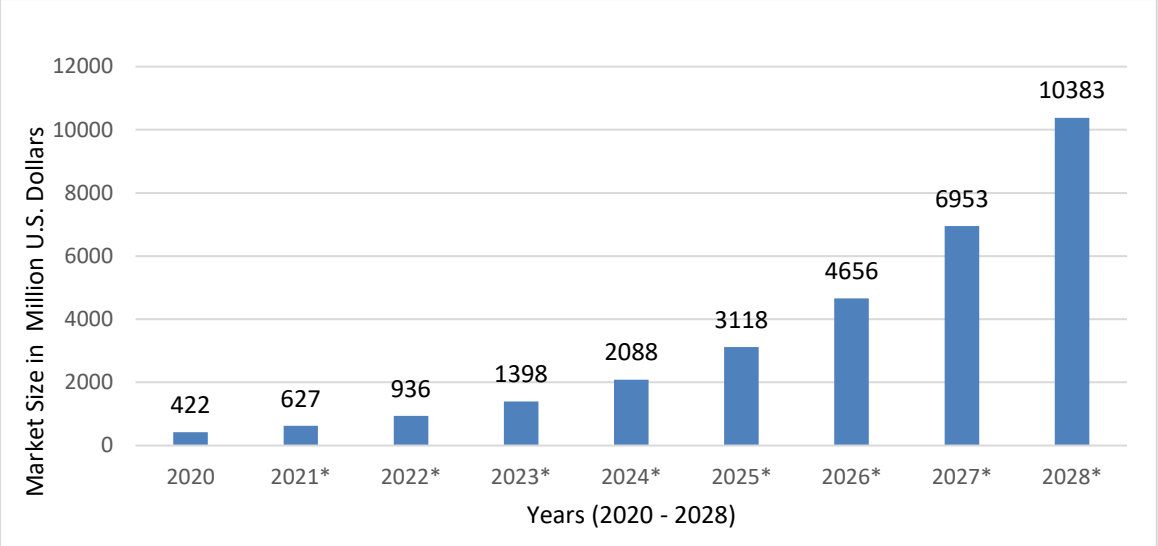


Figure 2: Global Process Mining software market size from 2020 to 2028, (Fortune Business insights, 2021)

Figure 2 illustrates the size of the global Process Mining market in 2020 and its forecasted growth for the years 2021 to 2028. The market is expected to increase from $422 million in 2020 to more than $10 billion in 2028 (Fortune Business Insights, 2021).

Consequently, gaining knowledge on how to initiate a continuous Process Mining implementation in companies is valuable. Organizations can implement Process Mining to discover, enhance and check the conformance of their business processes, basing their approach on data-based insights rather than subjective opinions or incomplete views.

1.2.2 Philips Domestic Appliances' challenge toward Process Mining

Philips DA is facing a challenge when trying to implement Process Mining. The below-described challenge illustrates the relevance of creating a guideline for the implementation of Process Mining for all companies.

Dimphy Beaujean, the Global Operational Excellence leader at Philips DA explains:

> *"We are very interested in Process Mining and are convinced it could support our approach to Business Processes improvement as part of our Operational Excellence program. The problem is that we don't know where to start: we don't know whether Process Mining would really be useful for us, if it is the case, we probably have the needed data among an extended volume of data in our IT systems, but we don't know exactly what type of data we need, how to use it nor what process we should begin with. Therefore, we would like to have a handbook explaining to us how we should implement Process Mining at Philips Domestic Appliances."*

As this testimony demonstrates, there is a gap between the potential benefits of Process Mining and the **hurdles to implement the technology**.

1.2.3 Transposition of Philips DA's challenge to other companies, acknowledged by the literature

The challenge that Philips DA faces is transposable to other companies. Organizations have difficulties dealing with and extracting value from their large amount of data, making the implementation of Process Mining more fastidious.

The disparities listed in Chapter 1.2.2 have been confirmed by the literature: *"For data and analytics leaders, Process Mining delivers insights that enable faster, smarter decisions and stronger performance on an organization's most critical priorities"* (Kerremans, 2021). However, companies are **unable to understand and manage their Big Data** as they are not suitably structured. One of the organizations' main hurdles is to understand and use their raw data to turn them into knowledge and create value (van der Aalst, 2011). During an interview for Gartner's Market Guide for Process Mining (2020), Wil van der Aalst tries to understand these disparities:

> *"Data quality and people remain the two main hurdles for widespread adoption. Typically, 80% of the efforts and time are spent on locating, selecting, extracting, and transforming the process. The time needed to apply Process Mining is short (say 20%) once the data are available in the right format. Process Mining often reveals data quality problems that need to be dealt with urgently".*

This challenge increases the relevance of this study. Providing Philips DA with a guideline on how to initiate a continuous Process Mining implementation will also help other companies to address this problem.

The motivations emphasized in this chapter shape the aim of this thesis. This research will contribute to the academic and practical worlds by providing a guideline for continuously initiating the implementation of Process for Philips DA and other companies.

1.3 Research question

This thesis will propose an answer to the following research questions:

How can an organization initiate a continuous Process Mining implementation?

a. Which Data measures should an organization take to initiate a continuous Process Mining implementation?

b. Which Organizational measures should an organization take to initiate a continuous Process Mining implementation?

c. Which Processes measures should an organization take to initiate a continuous Process Mining implementation?

This guideline will enable organizations to **improve their business processes through data insights** rather than subjective opinions. It will satisfy the data, people and governance, and process requirements.

Since the goal of this research is to create a guideline for initiating a continuous Process Mining implementation, the Design Science methodology will be employed to create the artifact. An adaptation of the DIDOV framework will be used as the spine of this work. The guideline will be articulated around the 5 steps of this framework. A reliable piece of research would improve the quality of this thesis. Therefore, the guideline will be constructed around three sources of data (triangulation): (1) literature review, (2) 10 user interviews with employees from Philips DA, Royal Philips, Heineken, and Whirlpool who are currently or already did successfully implement process mining in their companies. The guideline will firstly be designed based on these 2 sources of data. (3) Afterward, the guideline will be presented to 4 experts in Process Mining to be validated. The expert interviewed are process mining consultants, doctors, or head of the process mining capability in large companies.

1.4 Structure of the research

After presenting the Background, the Research Motivations and relevance in Chapter 1, Chapter 2 will conduct a thorough Literature review on Process Mining, on the Three pillars for a continuous implementation, and on the DIDOV framework. In Chapter 3, the Research Method and Study design will be imparted. Afterward, the collected data (Semi-structured user interviews and Unstructured validation interviews) will be introduced, the Research Quality will be evaluated, and the Research Data management Plan will be communicated. In Chapter 4, the results of this thesis will be analyzed, leading to the design of the IPMI (Initiating Process Mining Implementation) guideline. In Chapters 5 the discussions and limitations will be provided. Finally, this research will be concluded in Chapter 6.

2 LITERATURE REVIEW

This chapter provides the state-of-the-art of the Process Mining discipline. The definition, benefits, and context are presented. This chapter will then introduce the notion of event logs, the main Process Mining techniques and three relevant implementation methodologies.

Van der Aalst (2011, p.8) introduces Process Mining as:

> *"Process Mining is a relatively young research discipline that sits between [...] data mining on the one hand and process modeling and analysis on the other hand".*

Therefore, existing literature will be studied on **data** and **processes**. Moreover, as Wil van der Aalst mentioned in an interview for Gartner's Market Guide for Process Mining (2020), one of the main challenges when implementing process mining is *People.* Therefore, existing literature on People and Governance will be analyzed as part of the **Organizational** Pillar.

2.1 Process Mining

Members of the IEEE Task Force on Process Mining published a Manifesto declaring various principles and intentions to promote the field in both the academical and business world. The Manifesto states that Process Mining should exploit event data in a meaningful way. This would enable the ability for organizations *"to provide insights, identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures, and streamline processes"* (IEEE TASK FORCE ON PROCESS MINING, 201, p.174)

Process Mining has created number of success stories: *"Within Order-to-Cash, Siemens has increased its automation rate by 24% and reduced rework by 11% globally, resulting in 10 million fewer manual touches per year"* (Celonis, 2022), Ingram Micro implemented Process Mining on it Order-to-Cash process. "*Ingram Micro saved 50,000 labor hours by improving processes*" (Celonis, 2022). Uber is also a great success story example with *"$20M efficiency gains through handling time improvements*" (Introduction - Process Mining, 2022).

The research on Process Mining started in 1999 at Eindhoven University of Technology. The discipline has mainly been created by Wil van der Aalst. As the author contributed to most of the research in the field, the main definitions and concepts introduced in this thesis were researched by van der Aalst (2011, 2012, 2013, 2016). After the fundamentals of the field were defined, Process Mining spread to both the business and the academical world, enabling new research and usages.

2.1.1 From Big Data to Big Event Data

The digital universe is exponentially growing, both in term of the quantity and the role of data. This phenomenon refers to the Big Data era (van der Aalst, 2016). In a research article from Hilbert and Lopez (2011), it is found that society switched from being analog to digital. Many other studies are supporting this finding, some examples are provided bellow:

- In April 2014 the IDC Digital Universe Study evaluated to 4 Zettabytes the amount of digital information stored in devices. The study forecasts that the volume of digital information will raise to 44 Zettabytes by 2020. The study exemplifies 44 Zettabytes as *"6.6 stacks of iPads from Earth to the Moon".*
- According to Zomaya and Sakr (2017), the amount of data generated, consumed, and integrated to computational systems is expanding. This is enabled by the constant availably and use of Internet technologies and applications.

As described above, the **growth of data** is proof that information systems are increasingly important. Moreover, according to van der Aalst (2016), beside the quantity, the role of data also participates in the expansion of the digital universe. Data are therefore increasingly important in nowadays business processes. The author adds that for the above reasons, the development of the digital universe (also referred to as *Big Data*) is more and more aligned to the physical universe. In research from van der Aalst (2016) it is explained that the goal of Big Data is to turn data into **value**.

Van der Aalst (2016) associates Process Mining to the Big Data technologies. Big Event Data are the aggregation of centillions of generated event data. Most companies struggle to extract valuable insights from their Big Event Data.

2.1.2 Event data and Internet of Events

According to van der Aalst (2013), the end goal of data is not to gather more and more of them (p.1). On the contrary, data ought to be used to **enhance business activities** and empower new ones. Hence, event data are the starting point to enable information creation. Therefore, the Manifesto on Process Mining dedicated their very first Guiding Principle to Event Data. It is stated that event data must be considered as "*First-class citizens*" (IEEE TASK FORCE ON PROCESS MINING, 2012, p.179-180). This research also informs that the outcome of Process Mining is systematically influenced by the **quality** of the input: the **event data**. Great attention should be paid to the quality of event data. They should systematically and automatically be recorded in a complete, safe, reliable, and structured manner (IEEE TASK FORCE ON PROCESS MINING, 2012, p.179-180).

Event data can occur in any place, for example: in a company's information system (e.g., purchase order to a supplier), in a device or a machine (e.g., ATM), in a social network (e.g., e-amils)

(van der Aalst, 2023, p. 1-2). Therefore, they are classified in three categories: life-events, machine-events, and organization-events.

The **Internet of Events (IoE)** is defined by van der Aalst (2013, p.2) as:

"We use the term the Internet of Events (IoE) to refer to all event data available".

In other words, the Internet of Events can be defined as the constantly increasing amount of event data and their new role in organization. Hence, they are enablers for Process Mining.

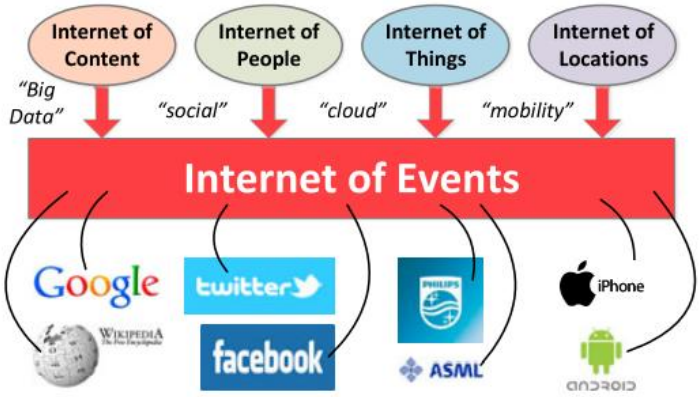


Figure 3: Internet of Events, van der Aalst (2013)

From Figure 3 two information can be concluded. First, Events data find their roots in multiples sources from the Internet. Second, the IoE is composed out of four parts: (1) the Internet of Content (IoC), (2) the Internet of People, (3) the Internet of Things, (4) the Internet of Locations.

Each part composing the Internet of Events in Figure 3 is defined by van der Aalst (2013) and illustrated by mean of examples in Table 1.

IoE composition	**Explanation**	**Example**
The Internet of Content (IoC)	Information generated by humans with the goal of increasing knowledge on any subject.	Web-pages, online articles, SharePoint articles, social media feed …
The Internet of People (IoP)	Data generated by social interactions and networking.	E-mails, Microsoft Teams interactions, LinkedIn messages.
The Internet of Things (IoT)	Physical devices with a unique ID, generating data and connected to a network and to internet.	Smart Watches, home security camera, connected cars and speakers.
The Internet of Locations (IoL)	Any data containing geographical, geospatial or movement dimensions. Enabled by the common usage of mobile devices (e.g., smartphones)	GPS, location from which a tweet is sent, step/ activity measurement.

Table 1: Composition of Internet of Events

2.1.3 Event Logs

Event logs are the foundation of Process Mining. Van der Aalst (2013) defines event logs as the **blueprint** (in the digital world) corresponding to an activity in the **process** (in the physical world). Events belong to a specific process instance (a case) and are ordered. Events can be seen as "*one "run" of the process*" (van der Aalst, 2013, p. 11). Events have properties that will be explained and exemplified in this sub-chapter.

Event logs have a structure and properties that are essential for Process Mining. According to van der Aalst (2016), data from event logs should refer to only one *process*. One *single process instance* is defined as a *case*. Therefore, each event in the log can only belong to one *case*. A process is defined as "*a collection of activities such that the life-cycle of a single instance is described*" (van der Aalst, 2016, p.128). In each *case*, some *activities* are happening (for example, order submission, invoice creation, payment confirmation). Assuming that the **events** in the case are **organized**, the **case id** and the **activity** are the minimal required information needed to run Process Mining. Only considering these pieces of information would enable to discover at an aggregated level what and how often activities are performed and establish root cause analysis. However, the Process Mining analysis can be much more useful if the event logs include more characteristics. **Attributes** like cost, resource (role, authorization data, …) or timestamp (beginning and/or end of each event, time the activity was offered to the resource, …) can enable performance analysis.

Case id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	…
1	35654423	30-12-2010:11.02	register request	Pete	50	…
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	…
	35654425	05-01-2011:15.12	check ticket	Mike	100	…
	35654426	06-01-2011:11.18	decide	Sara	200	…
	35654427	07-01-2011:14.24	reject request	Pete	200	…
2	35654483	30-12-2010:11.32	register request	Mike	50	…
	35654485	30-12-2010:12.12	check ticket	Mike	100	…
	35654487	30-12-2010:14.16	examine casually	Pete	400	…
	35654488	05-01-2011:11.22	decide	Sara	200	…
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	…
3	35654521	30-12-2010:14.32	register request	Pete	50	…
	35654522	30-12-2010:15.06	examine casually	Mike	400	…
	35654524	30-12-2010:16.34	check ticket	Ellen	100	…
	35654525	06-01-2011:09.18	decide	Sara	200	…
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	…
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	…
	35654530	08-01-2011:11.43	check ticket	Pete	100	…
	35654531	09-01-2011:09.55	decide	Sara	200	…
	35654533	15-01-2011:10.45	pay compensation	Ellen	200	…
4	35654641	06-01-2011:15.02	register request	Pete	50	…
	35654643	07-01-2011:12.06	check ticket	Mike	100	…
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400	…
	35654645	09-01-2011:12.02	decide	Sara	200	…
	35654647	12-01-2011:15.44	reject request	Ellen	200	…
…	…	…	…	…	…	…

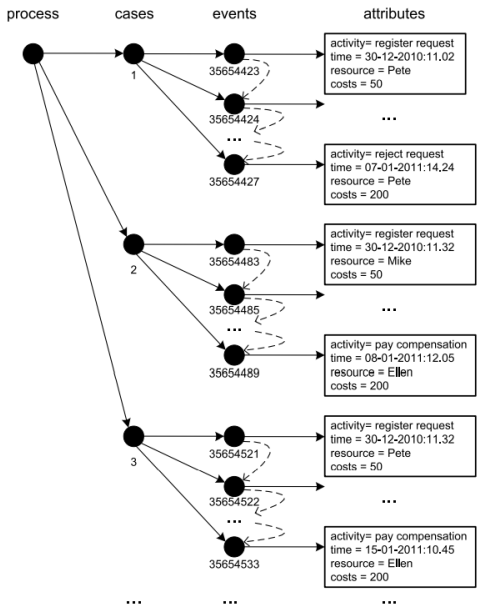


Figure 4: Example fragment of event logs, related to their tree structure, duplicated from van der Aalst, 2016

Figure 4 is composed of an example table of an event log (on the left) and a schema of its tree structure (on the right). These event logs belong to the process of handling request for compensation. The schema on the right shows that the process is the highest instance, under it the events are grouped by case (i.e., single instance of a process). The case is displayed in the first column of the table. Each line of the table corresponds to a single event id (2nd column), and each event is linked to its activity and its attributes (columns: timestamp, resource). The tree schema shows that the cases are belonging to the same process, and that each case is composed of a set of events ids. There again, each event id corresponds to an activity that has attributes.

2.1.4 Discovery, enhancement, conformance

Previous research in the field highlighted three goals of Process Mining: (1) **discovery**, (2) **conformance checking**, (3) **enhancing** of processes (van der Aalst, 2016). Figure 5 summarizes the input and outputs of each technique. Moreover, in research from van der Aalst (2011, 2012, 2016) and IEEE TASK FORCE ON PROCESS MINING (2012) the three goals are further defined. The explanations provided in the previously mentioned literature are combined bellow.

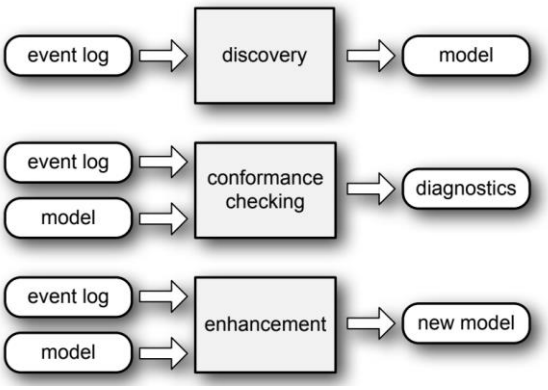


Figure 5: Input and output of the three types of Process Mining (IEEE TASK FORCE ON PROCESS MINING, 2012)

Process discovery (1) is seen as the primary focus of Process Mining (van der Aalst, 2016). The aim of Process discovery is to uncover a model. Figure 5 shows that the discovery technique uses the event log as an input without making use of any other kind of a-priory information. The models are built, by design, on recorded events.

The second type of Process Mining technique is conformance checking (2). Figure 5 indicates that this technique uses an existing process model and its generated event logs as inputs, to obtain a diagnosis as an output. Eventually, the model used as an input can be created by hand or via process discovery (1). The alignment between the input and the output models is examined to verify if the as-is process conforms to the intended process and the other way around. In practice,

an example could be: a process model (intended process) requires that sales over one thousand euros are sent with a discount code for a future order. Examining the event logs will help determine if the as-is process is conforming to the process model. Therefore, this technique can be used to expose, explain, and measure the deviations.

The last type of Process Mining, enhancement (3), aims at improving or extending a current process. Figure 5 illustrates that this technique is using the generated event logs and their model to create a new enriched model. It is possible to make characteristics of the process transparent (i.e., bottlenecks, service levels, cycle times and occurrences), to analyze them so an enhanced model can be shaped. While the second technique (i.e., conformance checking) would quantify the alignment between the real process and the process model, the goal of this third technique is to modify the process model. As mentioned above, the enhancement type of Process Mining can be used to improve or extend a process model, these terms are clarified bellow:

- Repair: improve the process model to better align with the reality. An example is provided in ***Process Mining: Data Science in Action*** from van der Aalst (2016): in the process model, two activities are modeled sequentially (one after the other) but in reality, they are occurring simultaneously. Therefore, the model can be repaired. (p. 33)
- Extension: using event logs to add a new perspective to the model. For example, one could use time stamps to add a time perspective and therefore measure throughput times. (p. 33)

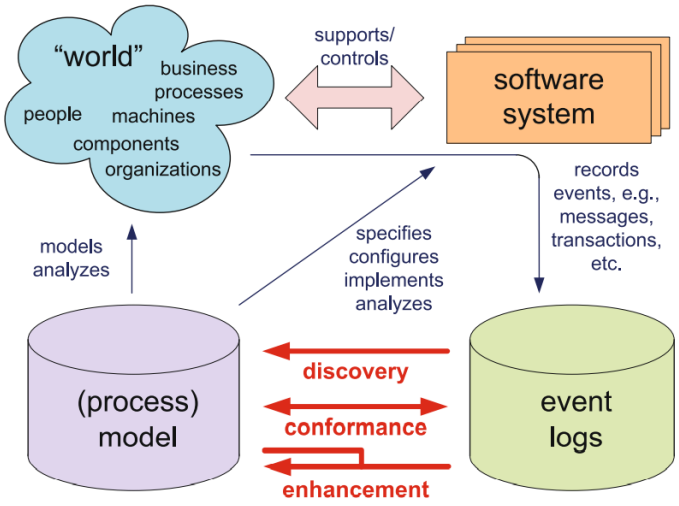


Figure 6: Positioning of the three main types of Process Mining: discovery, conformance, and enhancement (van der Aalst, 2016)

The above figure (Figure 6) indicates that Process Mining stands as a bridge between (process) models on the one hand, and real processes supported by their data on the other hand. The digital world laying in the software systems and the physical world are more and more aligned, they support and control each other. Information Systems are recording colossal number of blueprints and their attributes (messages, transactions, information about performed actions, …). In

the Process Mining field, these data are called Event Logs. It is assumed that companies Information Systems can systematically record the event logs referring to an activity in the process. Very often, those existing data are dispersed over multiples Information Systems and stored in an unstructured way. Therefore, companies may struggle to identify and collect those event logs. This preparation phase is a fundamental stage of Process Mining performance (van der Aalst, 2016).

2.1.5 Process Mining implementation frameworks

To implement a continuous Process Mining project, multiple methodologies have been introduced by the literature. Since there no existing framework scoping the initiating phase of the implementation, the presented frameworks will consider the whole life cycle of the implementation. However, only the initiating phase will be deepened.

2.1.5.1 L Life-cycle model describing a Process Mining project, applied to lasagna and spaghetti processes (van der Aalst, 2016)*

In the book ***Process Mining: Data Science in Action,*** van der Aalst (2016, p. 387-420) created a framework (L* Life-cycle model) for implementing Process Mining. This framework applies in totality to the Lasagna processes (structured process), but only partially to the Spaghetti processes (unstructured process). In this section, both types of processes will be differentiated, and the framework will be presented. Conclusion will be provided on the most and least relevant points of the framework for this thesis.

Processes are represented on a continuum ranging from highly structured processes to unstructured processes.

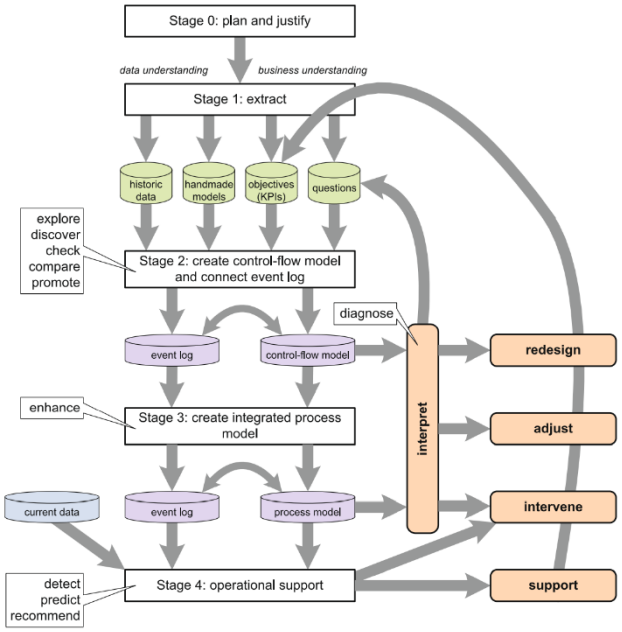
- Structured process (Lasagna processes): the pattern of activities is frequent, repeatable, stakeholders have an accurate estimation of inputs and outputs. Structured processes are often suitable for automation. Van der Aalst (2016, p.387) defines a structured process: "*a process is a Lasagna process if with limited efforts it is possible to create an agreed-upon process model that has a fitness of at least 0.8, i.e., more than 80% of the events happen as planned and stakeholders confirm the validity of the model".* Therefore, it is possible to apply all steps (from 0 until 4) of the L* life cycle model for implementing Process Mining.
- Unstructured process (Spaghetti processes): for each activity; information, data, inputs, and outputs are challenging to discern. This type of process is relying on subjective intuitions, experience, lack of information and trial-and-errors. Therefore, only the first steps of the

L* Life-cycle model are applicable (from step 0 until 2 included). Steps 3 and 4 cannot be applied because of variability issues.

Figure 7 (van der Aalst, 2016) displays **the 5 high-level stages** of the life-cycle model describing a Process Mining project: (0): **plan and justify**, (1) **extract**, (2) create control flow-model and connect event logs, (3) create integrated process model, (4) operational support.

Even if the L* life cycle model is not applicable for the end stages of the implementation, this constraint does not apply to this thesis because of the chosen scope. The scope of this study relates to stages 0 and 1 of the L* life cycle models, further steps are out of the scope of this research. Consequently, stages 0 and 1 are the only stages that will be detailed. Therefore, as both types of processes are suitable, this thesis can plentifully benefit from information of this model.

In the initiating phase (step 0), van der Aalst suggests justifying the type of Process Mining project. The three existing types of Process Mining defined by van der Aalst are:



- **Data-driven / curiosity-driven**: the trigger for the launch of the project is the availability of event data. This exploratory project is not led by specific questions or goals but by the enthusiasm of stakeholders

- **Question-driven**: this type of project aims to answer a specific question related t the chosen process. For example, why is the delivery date often not respected during the weekend?

- **Goal driven**: the aim is to enhance a process with regard to its KPIs. For example, improve lead time.

Figure 7: L* life cycle model, possible high-level stages of a Process Mining Project (van der Aalst, 2016, p. 394)

Van der Aalst advocates that it is easier for companies without Process Mining experience to start with a question-driven project. The step zero of the L* life cycle model also urges companies to create an agenda including milestones, allocate resources and monitor progress.

In the extraction phase (1), van der Aalst recommends extracting event data from systems, models from domain experts and objectives and questions from management. Regarding the data extraction, earlier in the book, van der Aalst explains how to turn raw data into event logs. While this information is important, more basic questions remain unanswered by this model. For example**:** How to define the required data? How to assess their availability? How to locate them? How

to collect them? Additionally, in a goal-driven project, the questions and objective should be formulated in stage 1 by interviewing stakeholders (e.g., domain experts, process executers and customers, management).

Taking a step back, one can conclude that these steps zero and one are **succinct and high level**. This part of the L* life cycle model should be deepened by the literature. One pillar is lacking in this model: the *organizational* part (e.g., gathering a team, leveraging desire, enhance/create Process Mining knowledge, creation of a strategy, of a continuous capability…). Moreover, even if some pieces of advice are given, **only limited explanation is provided to organization on how to apply and articulate the recommendations**. Therefore, other implementation frameworks need to be studied. Nevertheless, the directions provided are relevant and useful. In particular, the suggestion of starting by a question-driven project, the planning activities and the objective and questions definition will be included and deepened in the guideline.

2.1.5.2 PM²

In this sub-chapter, the overview of the PM² methodology and its characteristics will be presented (van Eck et al., 2015). Moreover, the stages that are part of the scope of this research will be described in detail.

PM² is a methodology helping organizations to implement Process Mining projects. Especially, this methodology applies to project aimed at improving process performance or compliance. Figure 8 displays the high-level stages of the methodology. The particularity of this methodology is that it analyses each step of the methodology considering its inputs and outputs. The methodology begins with an initialization phase that groups Stage 1 (**Planning**) and Stage 2 (**Extraction**), In this phase goals are defined (for example, cost reduction of X% on the given process, increase speed of execution of the process) and translated into research questions. They enable to scope the data extraction. In the following phase (Analysis iterations), the research questions are iteratively refined and resolved: in Stages 3, 4, 5, the data is processed, mined and the outcome is then evaluated. If the result is satisfactory, it is used to improve the process (Stage 6).

This methodology operates by analyzing inputs and outputs objects at each Stage. The initialization phase (Stages 1 and 2) is part of the scope of this research; however, the further Stages (3,4,5,6) are not. Therefore, **only stages 1 and 2 will be thoroughly presented**.

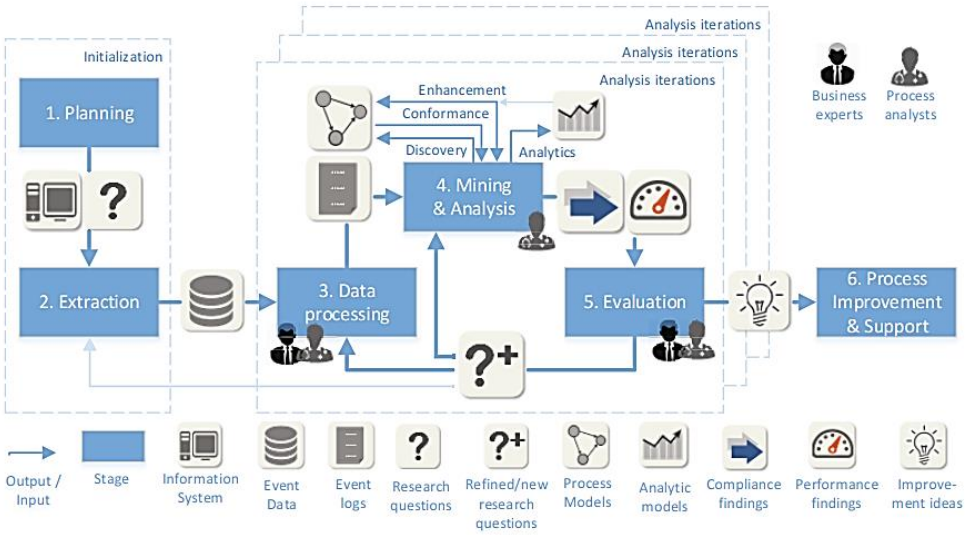


Figure 8: Overview of the PM² methodology

Figure 9 presents the overview of Stage 1 in term of inputs, activities, and output.

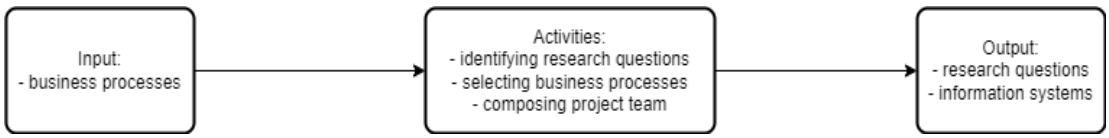


Figure 9: PM², Stage 1 Planning

The goal is this Stage is to launch the project and to enable the outputs: defining the research questions and scoping the information systems which are supporting the execution of the business process to be mined. The inputs of the first Stage are the company's business processes.

Three activities are mentioned in Stage one. They may happen in any order.

- **Selecting business processes**: To choose the right process, some criteria are to be taken into consideration. For example, the characteristics of the process, the quality of the event data, the changeability (ability to take actions to improve the process execution accordingly to the findings). This activity enables the definition of the information systems in which the system execution lays.
- **Identifying research questions**: the first step of this activity is to define goals. The goals must then be turn into research questions. PM² gives a definition of the term research question: "*questions related to the selected process that can be answered using event data*". The research questions should refer to characteristics of the process (for example, cost, time…). The research question can be specific or vaguer. In this last scenario, multiple iterative analysis will be performed to refine and answer the research questions.
- **Composing project team**: the team should be composed of multiple experts in each field. Some roles are identified in Table 2.

Function	Explanation	Importance
Business owners	Are in charge of the business processes	Secondary role
Business experts	Know the business aspect and executions of the processes	Primary role
System Experts	Are familiar with the IT aspect of the processes and the systems supporting the processes	Secondary role
Process Analysts	Skilled in analyzing processes and applying Process Mining techniques	Primary role

Table 2: PM²: team roles, (duplicated from PM², 2015)

By the end of Stage 1, the definition of the research questions and of the information system are forming the outputs.

Figure 10 displays the overview of the inputs, activities, and outputs of Stage 2 (Extraction).

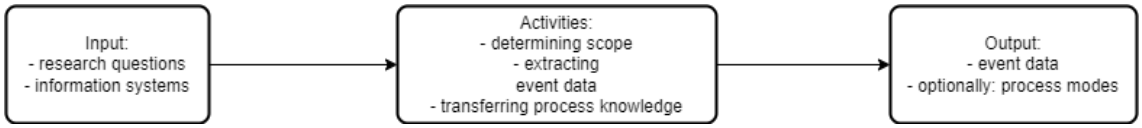


Figure 10: PM², Stage 2 Extraction

The aim of the second stage is to extract event data (defined as: "*a collection of events without predefined case notion or event classes*" and optionally process models. To perform this Stage, some inputs are needed: the research questions and the information systems. In Stage 2, three activities are identified by the methodology. The first activity should be performed first, the second and third activity can be performed simultaneously.

- **Determining scope**: the scope of the data extraction must be defined prior to extraction. To execute this activity, some criteria are to be taken into consideration: the granularity level, the period, the data attributes, the correlation between data.
- **Extracting event data**: after the execution of the first activity, the event data can be created. Therefore, the data from the selected process model are collected from the information systems. They are combined into a single collection of events.
- **Transferring process knowledge**: business experts and process analysts meet in interviews and thought showering sessions knowledge related to the process (for example, process documentation) and the data attributes are discussed. This activity aids the business analysts to be more effective when processing and mining the data in the later Stages.

As a conclusion, this methodology is adopting a less-technical perspective than the L* Lifecycle methodology described in Chapter 2.1.5.1. For the first time, the team creation, and the idea of different roles in the implementation has been researched. However, this methodology remains high-level and has a broader scope than the one adopted in this thesis. Moreover, this methodology

offers a guideline for a **one-time Process Mining project**. The idea of implementing a continuous capability is not mentioned. The guideline that will be design in this research will operationally detail how to best achieve the activities mentioned in the PM² methodology.

2.1.5.3 PPMS

Process Mining for Six Sigma (PPMS) is a methodology for **continuous process improvement** projects using Process Mining as a tool. It was created by Turetken (2020). On top of the guideline creation, the authors extended an existing Process Mining tool (Process Gold) to support the Six Sigma part of the PPMS. This artifact is an adaptation of the PM² methodology presented in section 2.1.5.2. The artifact is built with the DMAIC framework as its spine. DMAIC is a Six Sigma continuous improvement framework. It is composed of five steps: Define, Measure, Analyse, Improve, Control. The steps are inspired by the PM² methodology but, because of the adaptation to DMAIC, the order may differ. Figure 11 displays the eight steps of the artifact integrated in the DMAIC framework:

- Define: Planning, Data Preparation (preliminary), Mining and Analysis (Exploratory)
- Measure: Data Preparation
- Analyse: Mining & Analysis (Explanatory)
- Improve: Process Improvement
- Control: Monitoring, Evaluation

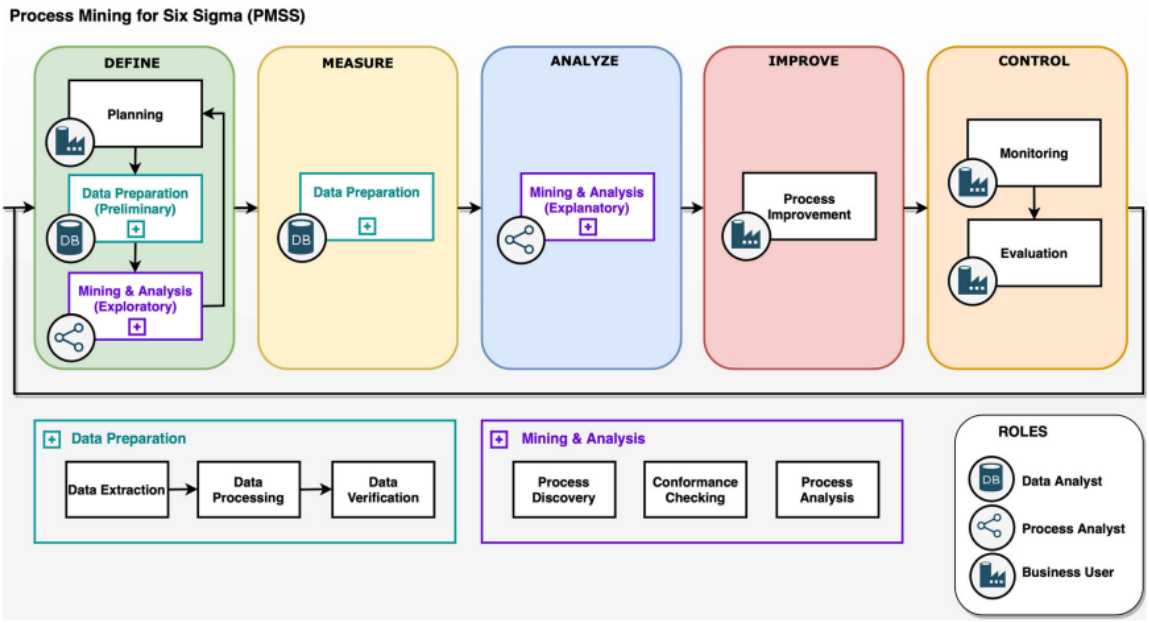


Figure 11: PPMS graphical overview

The DMAIC is envisioned as a cycle. This means that the last phase is used as feedback for beginning a new Define phase. The lowest part of Figure 11 provides the legend of the Figure, it

also indicates the three key roles in the artifact: the data analyst, the process analyst and the business owner.

Since the scope of this research focuses on the initiation of the implementation, only the specificities of the Define phase will be explained below (Planning, Data Preparation (preliminary) and Mining and Analysis (exploratory).

Table 3 describes the responsible role, inputs, activities, and output for each step. In the Define phase, the three steps (Planning, Data Preparation (preliminary) and Mining and Analysis (exploratory) are performed iteratively until the Business Goal is specific. The clarity of the overall Business Goal is an important success factor pointed by the methodology. Each activity within the three steps can also be conducted iteratively.

	Responsible	Input	Activities	Output
Planning	Business user	Information about business processes Potential problems First insights New business problems	Define business goals Identify and select business processes and supporting systems Identify business questions Compose a project team Create preliminary business case	Business goals Business questions Selected processes and existing inf. Systems Preliminary business case
Data preparation (preliminary)	Data analyst	Business goals Business questions Selected processes and existing inf. Systems Potential problems	Preliminary data extraction Preliminary data processing Preliminary data verification	Event logs Audit trail Data description
Mining and analysis (exploratory)	Process analyst	Event logs Audit trail Data description	Exploratory process discovery Exploratory conformance checking Exploratory process analysis	First insights Additional data (e.g., extra data attributes) Business problems
Data preparation	Data analyst	First insights Business problems Additional data (e.g., extra data attributes)	Data extraction Data processing Data verification	Event logs Audit trail Data description
Mining and analysis (explanatory)	Process analyst	Event logs Business problems Audit trail Data description	Process discovery Conformance checking Process analysis	Improvement opportunities
Process improvement	Business user	Improvement opportunities	Assess the impact of improvement opportunities/alternatives Implement improvements/process changes	Process changes towards business goals Process performance indicators
Monitoring	Business user	Process changes towards business goals Process performance indicators	Diagnose Identify new business problems	Impact of process changes New business problems
Evaluation	Business user	Impact of process changes Process performance indicators	Verify and validate Supporting operations	Verified and validated improvement

Table 3: PPMS: activities, input, output, (duplicated from Turetken,2020)

The goal of the Planning step is to collect a large and broad amount of information. This information is used as inputs, they can refer to the business processes, problems, business related insights… After performing the activities depicted in the Planning stage, the business goals and questions are defined, the process is chosen and linked to the related information systems. This

enables the shaping of the preliminary business case. This step is mainly led by the business users, they are providing information about the domain knowledge and other relevant information.

The objective of the preliminary data preparation is to quickly extract, process and validate the data required for the chosen process(es). This activity is performed to enable both the exploratory mining & analysis step, and to enrich the definition of the business case. In this step, the leading role is the data analyst.

The aim of the exploratory mining & analysis step is to enrich the definition of business problems for the chosen process. This step uses data prepared in the previous step as input and analyses it using one or more of the three Process Mining techniques (discovery, conformance, enhancement).

These three steps can be conducted iteratively until the business goals are well-defined.

Reflecting on the PPMS methodology, the continuous approach is interesting as the idea of continuity and improvement is born. However, the scope of this methodology remains **broad and high-level** as this methodology was built on PM², presented in 2.1.5.2. This thesis will present a more operational guideline for companies, focusing on the initiating stage of the implementation. Moreover, the framework used is different between the PPMS methodology (using an adaptation of DMAIC) and this thesis (using an adaptation of DIDOV). This is because the goal of both pieces of research is different. On the one hand, the PPMS methodology aims at continuous process improvement projects using Process Mining as a tool. Therefore, the DMAIC is used in the PPMS research because of the *improvement* goal. On the other hand, in this thesis, the goal is to **initiate** a continuous Process Mining implementation. Therefore, the framework will differ to suit the *implementation* needs.

2.2 Three pillars for a continuous implementation

2.2.1 Data Pillar

2.2.1.1 How Process Mining relates to Data Mining

According to (Hand et al., 2001), data mining is a set of tools used with the aim of uncovering knowledge from large data bases. The data are analyzed through an algorithm, then models are built to summarize and highlight the key insights.

Van der Aalst (2016) highlights the main similarity between Process Mining and data mining: they are both data driven. However, most of the **data mining methods are not process-centric**.

In other words, most of the data mining techniques cannot comprehend, uncover, or analyze process models in terms of Business Process Management Notation. Only a few data mining techniques (for example, episode mining or sequence mining) can understand the process dimension. Yet, they do not recognize the end-to-end processes.

2.2.1.2 Data quality

The **quality** of the event logs is of primary importance and should be considered from the initiating phase of the implementation. The quality of the data belonging to the process to be analyzed will heavily determine the quality of the results. As Process Mining is a different discipline than data mining, both definitions of *data quality* differ. In other words, the definition of data quality in the context of data mining cannot apply to the context of Process Mining. Thus, this thesis will study data quality in the context of Process Mining.

A general definition can be provided for data quality in the context of Process Mining: a high quality of data in the context of Process Mining is **trustworthy, complete, have well-defined semantics and respect privacy and security** (IEEE TASK FORCE ON PROCESS MINING, 2012). In this sub-chapter, three sources will be presented to understand and assess the main event data quality issues.

The Manifesto on Process Mining (IEEE TASK FORCE ON PROCESS MINING, 2012) proposed an assessment table to judge the quality of the data for Process Mining (Table 4). In this table, the quality of data is ranged from the highest level (five stars) to the lowest level (one star). In the highest levels of quality, events are **recorded automatically**, using a **systematic semantic** approach (signifying the existence of ontologies), are **complete**, **trustworthy** and **respect privacy** standards. At a mid-quality level, the data matches the normally matches the reality. Data can be automatically recorded but not in a systematic way (the existence of ontology is not guaranteed), and data are not necessarily complete. At the lowest level of quality, the recorded data do not match the reality. Data is manually entered and therefore is subject to human-error. Events are often incomplete and inaccurate. Even if technically speaking, it is possible to conduct a Process Mining analysis on any data that respect the requirements previously mentioned, it is advised to run the Process Mining exercise only on the highest level of data quality (4 and 5). When lowering the quality of the data, the main risk is that the data does not correspond to the reality and therefore the outcome of the Process Mining analysis is unreliable.

Level	Characterization
★★★★★	Highest level: the event log is of excellent quality (i.e., trustworthy and complete) and events are well-defined. Events are recorded in an automatic, systematic, reliable, and safe manner. Privacy and security considerations are addressed adequately. Moreover, the events recorded (and all of their attributes) have clear semantics. This implies the existence of one or more ontologies. Events and their attributes point to this ontology. *Example:* semantically annotated logs of BPM systems.
★★★★	Events are recorded automatically and in a systematic and reliable manner, i.e., logs are trustworthy and complete. Unlike the systems operating at level ★★★, notions such as process instance (case) and activity are supported in an explicit manner. *Example:* the events logs of traditional BPM/workflow systems.
★★★	Events are recorded automatically, but no systematic approach is followed to record events. However, unlike logs at level ★★, there is some level of guarantee that the events recorded match reality (i.e., the event log is trustworthy but not necessarily complete). Consider, for example, the events recorded by an ERP system. Although events need to be extracted from a variety of tables, the information can be assumed to be correct (e.g., it is safe to assume that a payment recorded by the ERP actually exists and vice versa). *Examples:* tables in ERP systems, events logs of CRM systems, transaction logs of messaging systems, event logs of high-tech systems, etc.
★★	Events are recorded automatically, i.e., as a by-product of some information system. Coverage varies, i.e., no systematic approach is followed to decide which events are recorded. Moreover, it is possible to bypass the information system. Hence, events may be missing or not recorded properly. *Examples:* event logs of document and product management systems, error logs of embedded systems, worksheets of service engineers, etc.
★	Lowest level: event logs are of poor quality. Recorded events may not correspond to reality and events may be missing. Event logs for which events are recorded by hand typically have such characteristics. *Examples:* trails left in paper documents routed through the organization ("yellow notes"), paper-based medical records, etc.

Table 4: Maturity levels for event logs (Source: IEEE TASK FORCE ON PROCESS MINING, 2012)

In research on the improvements of Process Mining results, Bose et al. (2013, pp. 127–134) found four types of recurring problem in data quality for event logs: **incorrectness** of data (for example, human-error in the entry creates a mismatch between data in the information system and the reality), **incompleteness** of data (for example, missing unique identifier), **impreciseness** of data (for example, imprecise timing) and **irrelevant data** (for example, very large set of irrelevant attributes slowing down the analysis).

Mannhardt et al., (2019, p. 612) conducted research on **privacy-preserving Process Mining.** They found that more and more data are collected and stored by organizations. This causes privacy and security challenges. On the one hand, these challenges have been considered by the Data Mining field, leading to progress in privacy-preserving technology. However, on the other hand, these issues are not yet researched for the Process Mining context: "*privacy-preserving process mining is still in its infancy*" Mannhardt et al., (2019, p. 612).

Therefore, the data quality assessment method provided by the IEEE TASK FORCE ON PROCESS MINING (2012) will be utilized in this thesis.

2.2.2 Process Pillar

Process Mining is tightly linked with **Business Process Management (BPM).** Before considering the similarities and disparities between the two disciplines, both terms *processes* and *BPM* would benefit from being defined.

A **process** is traditionally defined as "*A process is a coherent set of actions carried out by a collaborating set of roles to achieve a goal*." (Ould, 2005, p. 32). In other words, these activities are executed in teams to the objective of a defined outcome. Processes can refer to different areas of the company such as operational processes, management processes or supporting processes. Processes can be considered through various level of granularity. The processes can be measured by performance indicators to measure their success based on the objective of each process.

To define the field of **Business Process Management** in a complete way, two views should be considered. On the one hand, Business Process Management is defined as: "*The discipline that combines knowledge from information technology and knowledge from management sciences and applies both to operational business processes*" (IEEE TASK FORCE ON PROCESS MINING, 2012, p. 193). On the other hand, according to van der Aalst (2016, p.16) BPM refers to: "*The discipline that combines approaches for the design, execution, control, measurement and optimization of business processes*.". Van der Aalst adds that originally the aim of the discipline was to (re)design and implement processes, while now the discipline tends to the monitoring and the diagnostic of processes. In other words, the BPM field is pivoting from being model-driven to being data-driven. Process Mining techniques are nowadays often employed in the redesign activity of BPM.

2.2.3 Organizational Pillar

Projects are made by people for people. When implementing new IT projects, the *organizational* pillar is key for success and should not be underestimated. The *Organizational* factor represents 80% of projects failure, this is partly explained by the fact that managers often have more *technical* competencies than *people* competencies (Armstrong, 2022). Another widely recognized reason for IT projects failure is the resistance to change. Therefore, the social context is important to consider since the beginning of the implementation to ensure a good deployment at each step of the project.

2.2.3.1 Change Management: the ADKAR model

In the ADKAR model, 5 objectives are recognized as key factors for a **successful change** (Figure 12). The 5 objectives should be met in the right order (Awareness, Desire, Knowledge, Ability, Reinforcement). In the book ***ADKAR: A Model for Change in Business, Government, and Our Community*** (Hiatt, 2022) the objectives are decomposed in set of activities including relevant stakeholders.

The model can be initiated when a need for change is identified. Each of the objectives should then be achieved. The **Awareness** refers to the perception of the need for change and the perception of their reasons; to the personal and general benefits; and to the dangers for not changing. The **Desire** is the enthusiasm for taking part in change, this pillar is personal and is influenced by the individual's situation. The **Knowledge** is a more technical part that refers to the information on how to implement the change. This pillar can be trained. The **Ability** represents the fulfilment of the actions leading to the change. The **Reinforcement** is the factors and actions aiming to maintain the change. This last objective (i.e., Reinforcement) will not be utilized in this research as it is not part of the scope. Figure 12 summarizes the main factors influencing each objective, linked with their enabling elements.

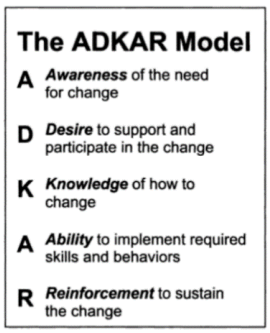


Figure 12: Overview of the ADKAR model

In this thesis, an adaptation of the ADKAR model for change management will be used. The first four activities will be included in the guideline, but they order may differ as it will adapt best to the context of the implementation of Process Mining.

2.2.3.2 Knowledge development

In ***The Career Architect Development Planner*** from Michael Lombardo and Robert Eichinger (1996), a theory for developing knowledge is established. According to the authors, **3 types of learnings** are required for a full comprehension. The different types of learning have a different weight in the total knowledge (100%).

The first type of learning is theoretical: **trainings** and coursework. This is equal to 10% of the total learnings. The second type of learning is **hands-on-job and** completing challenging assignments: it is learned through the practical application of the first 10% of knowledge. This type of learning corresponds to 70% of the total learnings. Finally, the last type of learnings is the **reflection** and discussion on previously acquired knowledge. This can be done by exchanging

with peers: feedback sessions, coaching sessions and collective discussions on what went well or less well. The total learning is therefore equal to 10% + 70% + 20% = 100% of the knowledge.

This combination of learning sources is beneficial in the context of Process Mining and will therefore be utilized in the guideline.

2.2.3.3 IT Governance

IT governance should be considered from the beginning of the implementation. IT governance aims to enhance value in the business and mitigate the risks. IT governance activities are also referred to as "*meta-management*" (Ebert et al., 2020, pp. 13–20). Weill et al. (2004, p.10) follow a similar perspective: they address three questions that must be answered in any IT related project.

> *"(1) What decisions must be made to ensure effective management and use of IT? (2) Who should make these decisions? (3) How will these decisions be made and monitored?"*

To answer those questions, **five types of decisions** should iteratively be addressed when initiating a new IT-related project Weill et al. (2004, p.10):

> *"(1) IT principles: Clarifying the business role of IT. (2) IT architecture: Defining integration and standardization requirements. (3) IT infrastructure: Determining shared and enabling services. (4) Business application needs: Specifying the business need for purchased or internally developed IT applications. (5) IT investment and prioritization: Choosing which initiatives to fund and how much to spend".*

Finally, the decision rights must be clarified. **6 archetypes for attributing decision rights** are elaborated by Weill et al (2004, p.12).:

> *"(1) Business Monarchy: Top Managers. (2) IT Monarchy: IT specialists. (3) Feudal: Each business unit making independent decisions. (4) Federal: Combination of the corporate center and the business units with or without IT people involved. (5) IT duopoly: IT group and one other groups (either management or business unit leaders). (6) Anarchy: Isolated individual or small group decision making"*

These questions and decisions will be considered when establishing the guideline for initiating a continuous Process Mining implementation.

Lars Reinkemeyer, the Vice-President of Customer Transformation at Celonis adopts a different view in the article **how to leverage CoE to accelerate execution excellence at scale,** published in the research report The Global State of Operational Excellence (2021, pp. 75–76). The author starts by explaining the importance of Centers of Excellence. This entity supports, commands and control data-driven process transformation and improvements in companies. This department also enables change management, executive sponsorship while making sure that improvement initiatives are measurable and aligned with the company's strategy. Moreover, the author indicates that most Process Mining projects are led by Center of Excellence.

Secondly, the author explains where the Centers of Excellence should stand in the company, and therefore, who is owns the decision rights. Reinkemeyer explains: *"Some CoEs [Centers of Excellence] have their organizational home in the IT department, while others reside on the business side. […] More and more, we see CoEs hosted in shared services units. […] There are a number of other possible organizational setups. These include a holding structure or private equity firm. […] Or a CoE led by a chief digital officer"*. However, even considering the great number of possibilities, the author affirms that the challenge does not reside in *who* owns the decision rights. The challenge resides is "*bridging the gap*" between all stakeholders. *"How will you align a technical CoE with the requirements of the business, or vice versa?"*

2.3 DIDOV framework

2.3.1 Six Sigma

Six Sigma is a methodology aiming to continuously improve the **quality** of activities and initiatives in the business (George, 2002). Six Sigma was created in 1986 by Motorola and is now globally recognized and adopted (Tjahjono et al., 2010).

Turetken (2020) explains that the σ from Six Sigma correspond to the standard deviation of a normal distribution. If a process is driven at one σ: 690 000 cases maximum will be defective on one million in total (minimum 31% are not deviating from the standard). If a process is driven at 2 σ, the percentage of non-deviating cases will rise to 69,2. At Six σ, there are 99.99966% of cases that are executed correctly (3.4 defects on 1 million cases). There is a positive correlation between the increase of the σ and the quality.

To achieve this quality level, the Six Sigma methodology disposes on multiple tools.

2.3.2 Design for Six Sigma

Design for Six Sigma (DFSS) is a subset of Six Sigma. It applies Six Sigma's principles to new products, processes, technologies or services. The goal of this approach -as for Six Sigma- is to enhance the quality while reducing its variation. The added value in DFSS the consideration for the early stages of the project: the **earliest decisions** made when designing a new product or service can have primary importance effect on the **success** of the activities (Vipin et al., 2013, pp. 369–374). Therefore, DFSS is also referred to as a problem-prevention methodology (Creveling et al., 2003).

DFSS consist of multiple **data-driven frameworks** used to (re-)design a new technology, product or service (Francisco et al., 2020). Some examples are: DMADV (define, measure, analyze, design, and verify), DDOV (Define, Design, Optimize and Validate), ICOV (Identify, Characterize, Optimize and Validate). In this thesis, an adaptation of the DIDOV (Define, Identify, Design, Optimize, Verify) will be used.

2.3.3 DIDOV

DIDOV stands for **Define, Identify, Design, Optimize, Verify** (Vipin et al., 2013, pp. 369–374). In this section, each step of the framework will be explained. Each of the DIDOV steps can iteratively be repeated until the desired outcome is reached. As the DIDOV cycle is used for the implementation of new services or products, it is usually not an iterative cycle in its whole. It *can* be repeated if the Verify stage is not successfully verified. In this case, the cycle can begin again from the commencement or from any in-between phase. Usually, the DIDOV cycle can be followed by a DMAIC cycle (Define, Measure, Analyze, Improve, Control) to continuously improve the quality of the implemented service.

According to Philips's training document on the DIDOV methodology (2007)[1]:

Define: This first activity sets the foundation of the project. At the beginning of the project, a set of activities should be performed to define: business needs for launching the project, a vision of the current/ future state, the stakeholders, a set of potential benefits / goals. Hence, the process that will be impacted must be chosen.

Identify: The project should be broken down into manageable and prioritized pieces, and the needed resources should be identified. The project tracking method and a change management

[1] Internal Philips's training document that is not publicly available.

method should also be considered. Afterward, the process knowledge must be identified and exploited, this can be through performing a SIPOC analysis (Supplier, Input, Process, Output, Customer). To maintain a high quality, the *usage environment* of the new product or service, and the quality targets should be identified.

Design: The goal of the Design phase is to select the best concept for the implementation. Eventually, a pilot can be released at the end of this stage. The Design phase consist of a set potential solution that meet the customer's needs and predict the ability of successfully and continuously implementing the service/ product. The best concept is then selected and detailed.

Optimize: The goal of this phase to critically examine the detailed design and preparing it for launch. The potential quality gaps and listed and solutioned.

Verify: In this last step, the optimized design is verified by the team member and the sponsor. If the quality is not sufficient, the project can reiterate from any of the previous steps until the quality goals are met.

The DIDOV cycle is well suited for this thesis. It has been chosen over other frameworks mentioned in chapter 2.3.2 as it is designed for the **implementation of new services or products**, and it is a **data-driven** and **proactive methodology** (Vipin et al., 2013, pp. 369–374).

An **adaptation of the DIDOV** framework will be utilized in this thesis to support the continuous initiation of Process Mining implementation. Moreover, it enables to design a guideline, verify, and optimize it. This is important for the **reliability** of the implementation. Finally, the choice of a Six-Sigma based framework was natural as all the company interviewed are using Six Sigma in their daily operations.

3 RESEARCH METHODOLOGY

3.1 Research Method

The goal of this thesis is to design a guideline for initiating a continuous Process Mining implementation. Because an artifact will be created, the Design Science methodology will be used as the research method.

In ***Design Science Methodology for Information Systems and Software Engineering***, Wieringa (2014) defines Design Science as:

> "The design and investigation of artifacts in context" (p.10).

Artifacts can be software or hardware components, a business process, a method, a conceptual structure… In the case of this thesis, the artifact is the guideline using an adaptation the DIDOV framework. The context can be an organization, people, values, norms, budgets, goals, processes or methods… The problem context is given to the research and cannot be changed. In this case, the problem context is the initiation of Process Mining implementation. Wieringa adds that the artifact in itself does not solve any problem. It is the interaction between the artifact and the problem context that contribute to solve the problem.

According to Wieringa, there are two kinds of research problems in Design Science: the Design problems and the Knowledge questions. On the one hand, the Design problems is triggered by a problem and a need for a change in the real world. The solution can be multiple designs and is evaluated by the utility. The utility depends on the stakeholders' goals. On the other hand, the knowledge questions are a request to understand the real world. In this case, the answer is a proposition envisioned as one unique solution. It is evaluated by the truth and therefore does not depend on stakeholders' goals.

This guideline uses the **Design problem** approach.

A template (Table 5) for defining the design problem is provided. Fulfilling the template is valuable even if not all information is known at first glance, and even if a hypothesis must be made. The reason is that it could facilitate the discovery of lacking information that are required to scope the research (Wieringa, 2014).

Template for design problems (aka technical research questions). Not all parts to be filled in may be clear at the start of the project.
• Improve <a problem context> • by <(re)designing an artifact>

- that satisfies <some requirements>
- in order to <help stakeholders achieve some goals>.

Duplicated from Wieringa (2014, p. 16).

Table 5: Template for Design problems (Wieringa, 2014)

With the intention of shaping the design problem, the following subchapters will define and fulfill each of the four parts of the template provided by Wieringa.

3.1.1 <a problem context>

Organizations used to make decisions founded on partially biased human judgment. This is because data was only partially available reliable. Thanks to today's information systems, data (and big data) are generally available, reliable, and complete. The challenge lays in understanding them. Process Mining is a tool that enables organizations to make clever use of their data to enhance the company's activities and processes. A continuous Process Mining implementation enables many improvements in organizations. To only name a few, the capabilities include data-based insights, root cause analysis, process transparency and enhancement, anticipation of issues, corrective measures, bottleneck-, and compliance violation- detection (IEEE TASK FORCE ON PROCESS MINING, 2012, pp. 164–194).

The potential benefits for using Process Mining in a company are obvious. However, the implementation of Process Mining in companies is not generalized yet. One of the reasons is that companies are not sure how to initiate their Process Mining implementation.

Therefore, the initiation of Process Mining implementation in companies should be improved. The problem context of the design problem is **the initiation of Process Mining implementation**.

3.1.2 <(re)designing an artifact>

When implementing a technology for the first time, Design for Six Sigma can be used to achieve a high level of quality.

Design for Six Sigma comes from the Six Sigma methodology. The goal of Six Sigma is to improve the quality while reducing the variance. The variance measures the deviation from a standard. There are two main fields within Six Sigma:

- Process improvements, using DMAIC framework for example
- Design for Six Sigma which is made for process / service / technology implementation

Within Design for Six Sigma multiple frameworks were created, including the DIDOV framework. DIDOV stands for Define – Identify – Design – Optimize – Verify. A deeper explanation of the DIDOV cycle can be found in chapter 2.3.
The artifact created in this thesis will be a guideline using an **adaptation of the DIDOV framework**.

3.1.3 <some requirements>

To create a guideline for initiating a continuous Process Mining implementation, the utilized DIDOV framework should be adapted to satisfy some requirements. 3 pillars are defined as the requirements for this guideline:

- **The Data Pillar**
- **The Process Pillar**
- **The Organizational pillar**

Each of the pillar is considered as satisfying the requirements if they successfully pass the last phase (Verify phase) of the DIDOV framework used in the guideline. The pillars are built on the literature review and the interviews, and they are strengthened by the literature. A more specific definition of each pillar can be found in chapter 2.2.

3.1.4 <help stakeholders achieve some goals>

Organizations have always tried to improve their business processes. This could translate by reducing waiting times and delays, overstock, errors or risks, while improving customer satisfaction, productivity… The end goal is to offer to customers the best quality, at the lowest cost and the quickest time.

This objective has stayed the same since hundreds of years. What changed is the way to approach the challenge. Before, subjective decisions were taken by humans based on instinct, experience, or traditions. Today, companies are aiming to take decisions based on data insights.

Process Mining is a tool that is part of this bigger initiative. Ultimately, this guideline is designed to help stakeholder achieve some goals. This goal is: **to continuously improve business processes through data insights rather than subjective opinions**.

3.1.5 Filling the template for design problems

Based on the elements mentioned previously, the template for design problems is filled as follow:

Template for design problems (aka technical research questions). Not all parts to be filled in may be clear at the start of the project.
• Improve: the initiation of Process Mining implementation. • By: designing a guideline that follows an adaptation of the DIDOV framework. • That satisfies: the Data-, the Organizational-, and the Process- Pillars. • In order to: continuously improve business processes through data insights rather than subjective opinions.

Duplicated from Wieringa (2014, p. 16).

Table 6: Filled template for design problems

3.2 Study design

To solve the research question "*how to initiate a continuous Process Mining implementation?*", a guideline will be designed, it will help companies in this journey.

This thesis is built on the Design Science approach, as an **artifact** will be created. The methodology from Wieringa (2014) is well suited for this thesis. It was conceived for the Information Systems fields, and it enables the researcher to start their study from a **real-world problem**, define some requirements in order to (re)create an artifact that helps stakeholders to achieve their goals.

This thesis is embedded in Wieringa's methodology and has been constructed around it (Figure 13). Both Philips DA and the researcher of this thesis envisioned to promote the continuous improvement of business processes through data insights rather than subjective opinions. This is the reason for the common agreement between stakeholders to dedicate the graduation internship to Process Mining implementation. At the same time, the researcher of this thesis enhanced their existing knowledge and developed new ones on the Process Mining field. This has been done through university classes, literature readings, market study, trainings, and certification on vendors platforms.

These led to 2 outcomes. On the one hand, a problem in the real world has been identified: the **implementation of Process Mining is not as easy as companies would primarily think**, and some companies might feel lost and do not know how to initiate their Process Mining implementation. The initiation of Process Mining implementation can benefit from being improved, which would help Process Mining to be **more widely adopted by companies**. On the other hand, while studying various Process Mining implementation frameworks (i.e., L* life cycle, PM², PPMS), it also appeared that a detailed and operational guideline for initiating Process Mining implementation in companies was lacking in the literature.

These 2 discoveries guided the problem context formulation *the initiation of Process Mining implementation*. To define some requirements for the artifact creation, the literature has been studied. Some quotations can efficiently **justify the choice of the 3 pillars**: Process-, Organizational-, and Data Pillars:

- Choice of the Process- and Data- Pillars: Van der Aalst (2011) defines Process Mining as: "*a relatively young research discipline that sits between […] data mining on the one hand and process modeling and analysis on the other hand*" (p.8).
- Choice of the Data- and Organizational- Pillars: "*the technology is accessible for any organization. What should improve is the scale of adoption in organizations. […] Data quality and people remain the two main hurdles for widespread adoption*" (Wil van der Aalst in an interview for Gartner, 2020, p. 1-2).
- Choice of the Organizational Pillar, and decision to use an adaptation of the DIDOV framework: "*In my view, there are too many small pilot projects. To reap the true benefits of process mining, one should not look for quick wins only. Process mining should be a continuous activity […]. Organizations that were able to lift process mining to the enterprise level have benefited most*". (Wil van der Aalst in an interview for Gartner, 2020, p. 2).

Then, for each pillar the literature has thoroughly been studied (IEEE TASK FORCE ON PROCESS MINING (2012), Hiatt (2022), van der Aalst (2016), …

To complete the definition of the requirements, 10 **user interviews** have been realized. The interviewees were selected because they participated in their company's respective successful Process Mining implementation. The significant number of interviewees, their different backgrounds (Data Engineers, Operational Excellence, Project Managers, Business Process Experts…), and their various provenance (Philips DA, Whirlpool, Heineken, Royal Philips) enabled a **broad exploration of the required activities for each Pillar**.

The newly collected data enabled the design of the IPMI guideline (Initiating Process Mining Implementation). The guideline was modeled using an adaptation of the DIDOV framework as its spine. The adaptation of the DIDOV framework was used because is to support the continuous implementation of new technologies, services or goods. Moreover, this framework was particularly appreciated by the experts (validation interviews) because it enabled to verify all the data, process, and organizational requirements, by design, before the moving forward with analysis. This is important as it eases trust from the organization in the Process Mining technology.

The research emphasized importance on designing a valid guideline that will effectively be used by real-world companies. Therefore, the guideline needed to be **validated** by experts. The 4 selected experts are **Process Mining professionals** who dedicated their careers to helping other companies implement Process Mining. With the help of their inputs, the guideline has been enriched by metrics and real operational examples. All experts validated the guideline. In turn, the literature review has been updated to always triangulate information (Lombardo & Eichinger (1996), Reinkemeyer (2021), Mannhardt et al., (2019))

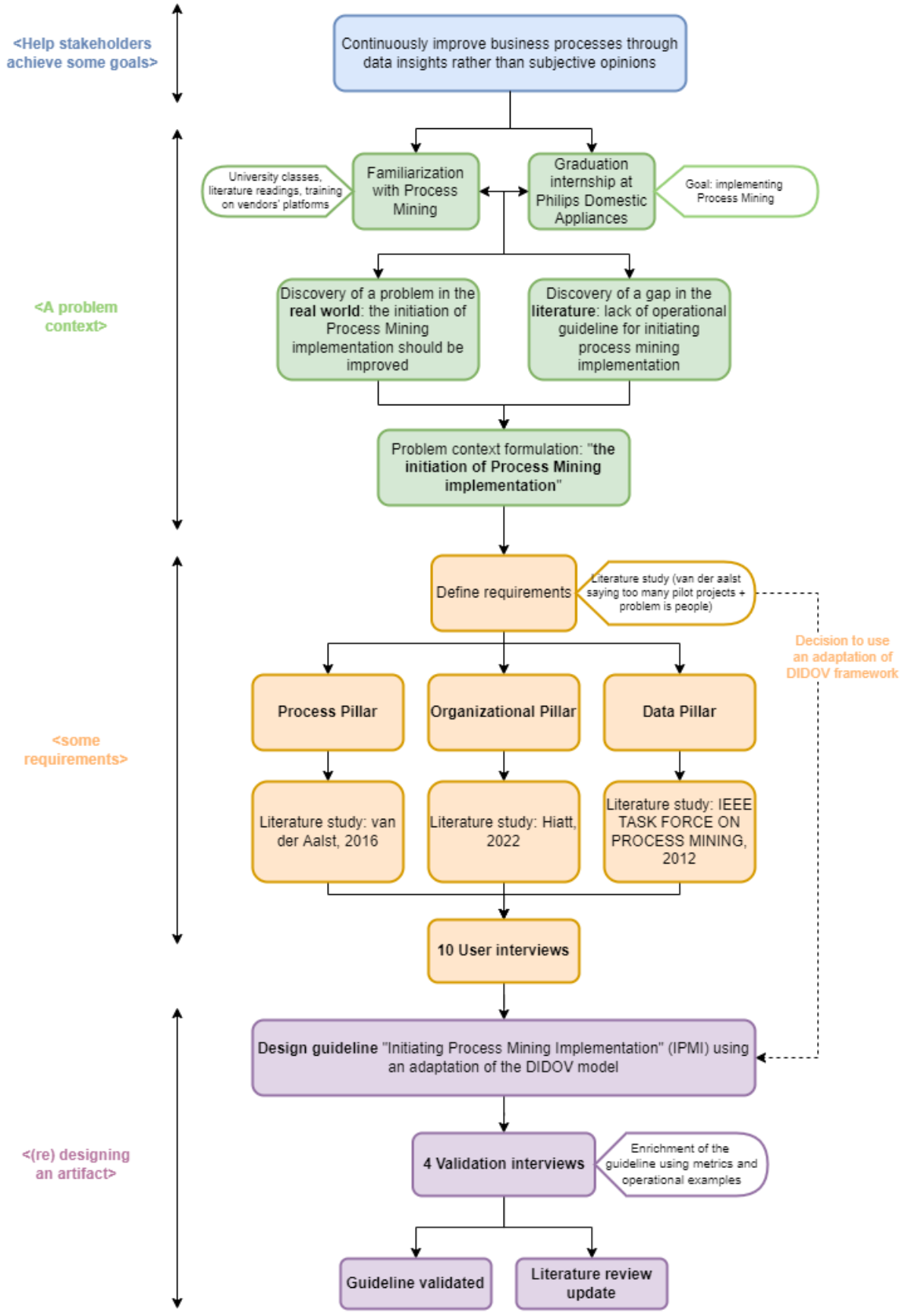


Figure 13: Study design

3.3 Semi-structured user interviews

3.3.1 Data collection

Interviews are conducted with **users of Process Mining**. The goal was to learn from the interviewee's **challenges** and main **activities** performed during the implementation. The insights gathered from the user interviews directly contributed to the creation of the guideline. The interviews were conducted in a **semi-structured** way. This format was the most suitable compared to fully structured or unstructured interviews. The semi-structured interviews enabled to focus and gain insights on the three pillars (Data, Organizational, and Process) while letting freedom to the informant for sharing their own experience and motivating their answers in their own words (Robert Wood Johnson Foundation, 2008). Afterward, interviewees were invited to reflect on the best practices and pitfalls of a Process Mining implementation.

Semi-structured interviews was a useful data collection method as the interviewer can improvise follow-up questions based on the informant's answers (Rubin & Rubin, 2005, Kallio et al., 2016). This benefit was enabled by ad-hoc preparation. As a matter of fact, semi-structured interviews require to develop a strong knowledge base on the research topic previously to the interviews (Wengraf 2001, Kelly 2010). Therefore, the author of this thesis prepared an interview guide that covers the main topics of the study (Table 9). The structure of the interview was the same for all informants, but the time dedicated for exploring each part variated accordingly to the participant's inputs.

3.3.1.1 Sampling

This research aims at designing a guideline that helps companies to initiate a continuous Process Mining implementation. As this methodology aims to be used by companies in the real world, it was important to collect data from **companies who have implemented** or are implementing Process Mining. To create a guideline as free of bias as possible, data from **four companies** were collected. This will improve the transferability of the results.

To select companies who have experience with Process Mining, the professional network of Philips Domestic Appliances was used. A list of potential companies was raised from Philips Domestic Appliances database. The list was analyzed, and four companies were finally selected on the following criteria: three companies were selected for their successful implementation of Process Mining (Royal Philips, Heineken, Whirlpool) and one company was selected because it is currently performing the implementation (Philips Domestic Appliances).

Next, the contact person in each of the companies was identified and emailed. The goal of this interaction was to verify that Process Mining was indeed successfully implemented in the company and ask for the contact details of some relevant colleagues who were part of the Process Mining implementation themselves. An emphasis was put on requesting a right balance in the profiles of interviewees. The variety of profile is determined by the domain in which a person is knowledgeable; for example, a data/ ERP expert differentiates from a project manager or from a business process expert. One to three contact profiles were shared by company.

3.3.1.2 Protocol

The first stage of the protocol was to enter in contact with the identified persons for the interviews. An introductory email was sent to the potential interviewees. In this email, the motivations for this research were expressed by explaining the need for a guideline that would help companies initiate the Process Mining implementation. Moreover, an emphasis was put on the importance to take to consider the Organizational Pillar on top of the more traditional Data pillar and Process pillar. Finally, the email intended to confirm that the potential interviewees were indeed closely involved in the implementation. When the targeted interviewees answered positively, they were asked to be interviewed.

Originally 12 experts were identified as suitable potential interviewees for this research and an invitation was sent to them. However, two of them indicated that they were only supporting the implementation project but were not closely involved. Therefore, no invitation was sent to them. Moreover, one person did not have sufficient time to dedicate to an interview within a short time frame. In order to keep an equilibrated balance between the different profiles, the researcher asked to the person whom did not have enough time to provide the contact details from one of their direct colleague who was also part of the implementation. An introductory email was sent to this new potential interviewee to be certain that they would have the same knowledge domain than their direct colleague who did not have sufficient time for the interview. Finally, 10 interviewees were selected for participating in the user interviews. The selected interviewees did not receive the interview guide in advance to prevent biased answers. All interviewees self-assessed as being good to very good in the following fields: Process Mining knowledge, Data skills, Knowledge of the Process, and Organizational skills.

Table 7 shows the interviewees' function and company. In Table 8, an explanation of the roles of the interviewees can be found, accompanied by its usefulness for the research.

ID	Company	Function

A	Philips Domestic Appliances	Automation & Industry 4.0 engineer
B		Regional Operational Excellence Lead
D		Global Operational Excellence Leader
F	Royal Philips	Business Process Owner - Source to Pay process
C		Director Business Process Owner – Order to Cash – Order management
G		Center of Excellence Lead for Insights and Analytics
E	Whirlpool	Head of Supply Chain Operational Excellence
H		Product Management Officer Supply Chain
J		Operational Excellence Supply Chain Manager
I	Heineken	IT Project Manager

Table 7: Interviewees' function and company

Function	Description	Usefulness for the research
Automation & Industry 4.0 engineer / Lead for Insights and Analytics	Experienced Business Intelligence and analytics professional. Skilled in data visualisation, extraction, modelling, and warehousing. Typically works on ERP implementations, this enables to quick scan data structures and translate these into solid solutions.	This technical role can enable the data scoping, localization and extraction during the project. It can help understand the large quantity of data collected. Main skill: data.
Business Process Owner / Business Process Expert	Understands the big picture of the process and empower end users, create and maintain documentation to provide visibility to other stakeholders, data driven person who understands both operational and information system languages.	This role can provide in depth insights on the chosen process, especially on both its execution and the way it is monitored in the information system. Main skills: process and data
Operational Excellence	Often acts as a project manager, specialised in business processes improvement. Uses continuous improvement to meet performance objectives and enable stakeholders and resources.	This role is involved in multiple business processes improvement projects and has a generalizable view on man-aging Process led projects. Main skills: project management and processes.

IT Project manager	Set goals and implement projects and strategies, specialised in IT, data, technologies. Can understand both business and IT domains.	This role is involved in multiple IT or Business Intelligence projects and has a generalizable view on managing IT led projects. Main skills: project management and IT/ Data.

Table 8: function description of interviewees and suitability for the thesis

3.3.1.3 Interview structure

A literature review was conducted prior to the interviews. This helped to construct the questions. The literature targeted various domains: change management domain, data quality, characteristics of processes. However, the literature did not target other domains that came as an **interesting outcome** of the interviews, for example, the **governance or the performance indicators**.

It is important to note that at this point of the research, before the interviews, the Organizational pillar was simply called People pillar. This is because in previous literature the *team creation activity* was the only one that did not fall under the Data or Governance categories. During the interviews, the whole governance field came as a highly important thematic, even though it was not considered before. To make the interviews guide easier to read and understand, the pillar will be called Organizational Pillar in the Table 9. However, interviewees were introduced to the pillar as the People pillar.

Table 9 depicts the user interview structure. The semi-structured interviews are useful for this research because they enabled to **systematically define relevant activities** for each of the pillars, allowing **interviewees' own experience sharing**. These activities will then be included in the designed guideline. As the goal of this interview is to discover relevant activities for the three pillars, the semi-structured interviews are suitable and effective: the potential for interviewee to divagate and not properly answer the question is reduced thanks to the follow-up questions. Some freedom is provided to interviewee's for sharing their examples and point of views.

The structure of the interview presented in Table 9 was replicated similarly for every interviewee. This includes the **introduction to the interview** (Part 0), the **experience sharing part** (Part 1) and the **reflection part** (Part 2).

In Part 1: Choice was made to let the questions related to each pillar vague. The goal is to let the informant provide the pieces of information that seem important to them. If they do not

understand the question, divagate, need clarifications or specifications; examples are provided, and informants are encouraged to provide their own ones.

In Part 2: Once the 3 pillars were filled with enough content (relevant activities), informants were provided with the flexibility to share their pieces of advice. This could be related to the three pillars, or also involve an unexpected activity or domain, which does not belong to any pillars.

The length of the interview was 1 hours for the 10 interviews. The time was divided as following: 5 minutes of Part 0 (introduction), 35 minutes of Part 1, 5 minutes of summary, 15 minutes (Part 2). The interviews were audio-recorded to let the interviewer focus on follow-up questions rather than on immediately writing down the answers. The participants freely agreed to the data collection and retention condition. Those conditions are specified in Section 3.6: Research Data management Plan

Part 0: Introduction	
Theme	**Content**
Data protection	Express gratitude cooperating Request permission to record the audio. According to GDPR, data will be stored for 5 years after the publication of the thesis and then deleted. (This is because the data collective is non-sensitive). The interview will be anonymized.
State goal of interview	Explain the aim of the thesis ("designing a guideline for helping companies to initiate a continuous Process Mining implementation. The guideline will be developed around 3 pillars: Process; People and Governance; and data").
Explain structure of interview	Structured interview: 75%. The researcher will ask about the main activities performed, and challenges encountered for each of the pillars. Specify that the interviewees should keep in mind that the questions are related to the implementation of Process Mining and not its current usage. Semi-structured interview: 25%. The researcher asks to the interviewees to take a step back. A discussion will open with the following theme: main lessons learned, best practices, pitfalls, advice to any company who would like to engage in Process Mining.

Part 1: Experience sharing

Theme	Questions	Purpose
Personal background	For how long have you been working at X? What are your main responsibilities and tasks? (X = the company where the interviewee working / worked during the Process Mining implementation)	Understand the interviewee's role in the organization. Question used as an icebreaker.
Organization situation with Process Mining	For how long is X using Process Mining? To what extent has Process Mining been implemented in X? What were the main motives for implementing Process Mining?	Determine the implementation maturity and successfulness and understand its motives. Increases transferability.
Individual experience with Process Mining	What was your role in the implementation? How did you start the implementation? What were the very few first steps?	Understand the informant's experience with Process Mining implementation, and their role. Increases transferability.
Process pillar	What were the main activities performed related to the process? What were the main challenges related to the process and how did you overcome them? Examples: choice of the process, scope of the process, granularity level of the steps of the process, understanding the process…	Build the process pillar of the guideline: make the participant explain and elaborate a critical view on the key steps, challenges, and main choices regarding the process.
Data pillar	What were the main activities performed related to the data? What were the main challenges related to the data and how did you overcome them? Examples: data collection / availability / quality within the process chosen, definition of event logs, data scoping and extraction, quantity, or complexity of data…	Build the process pillar of the guideline: make the participant explain and elaborate a critical view on the key steps, challenges, and main choices regarding the data.

Organizational[2] pillar	What were the main activities performed related to the people? What were the main challenges related to the people and how did you overcome them? Examples: team creation, change management, managing the project, getting full commitment from team members, resources, planning, support from management…	Build the process pillar of the guideline: make the participant explain and elaborate a critical view on the key steps, challenges, and main choices regarding the people.
Summary of activities and challenges mentioned by interviewee. Thank interviewee for the first 75% of interview.		
Part 2: Reflection on implementation		
Ask to take a step back and reflect on knowledge and experience. Opening of a conversation on the following theme: key success factors/best practices and pitfalls when implementing Process Mining (regarding data, people and governance, and process but also maybe any other area). What advice would you give to a company considering Process Mining implementation?		Let the possibility to explore any activity or challenge that was not part of the three pillars. For the activities/ pitfalls/ best practices mentioned related to the pillars; interpret their relative importance.

Table 9: Interview structure

3.3.2 Data analysis

Once the data were collected, they needed to be analyzed in a structured way to produce a reliable result and design the guideline. Braun and Clarke (2006) proposed a six-step method for analyzing qualitative data collected from interviews. The six summarized steps are: (1) Familiarizing yourself with your data, (2) Systematically coding interesting features across the whole data set to define information, (3) Colleting codes into potential themes, searching for themes, (4) Review and map the themes, (5) Refine, and define and name themes, (6) producing the report.

This methodology was followed for the analysis of the semi-structured user interviews.

- **Step 1**: The 10 interviews were audio-recorded, which enabled the interviewer to fully focus on understanding the interviewee's information. After the interviews, the audio records were listened and a 2/3 pages of each interview was made. The author of this thesis listened

[2] Initially called People pillar, instead of Organizational. No question was asked on the governance. Governance was spontaneously evoked by some informants.

each interview 2 times to be certain all relevant information would be included in the summary. Even if writing a summary of the interviews requires more time and efforts, this method was chosen overusing an automatic transcribed software. This is because it enabled the interviewer to become fully familiarized with the collected data.

- **Step 2:** Using a knowledge base developed during the literature review phase, the key topics and information of each interview were coded. These codes are visible in Appendix 1: User interviews summary. For example, when the implementation team was trained, the code *knowledge* is visible on the interview notes. Or for example, when the sponsor of the project communicated about it to the larger organization, the word *awareness* is visible in the interview notes.
- **Step 3:** The recurring codes were analyzed, and some trends were identifying, resulting in the creation of potential themes. For example, the codes containing *knowledge* or *awareness* data were grouped into a *change management* theme.
- **Step 4:** After creating the themes, they were mapped and review to ensure that they contain enough data and that the data is similar enough to belong to the same theme. Some themes were merged or on the opposite some themes were divided into two themes.
- **Step 5:** The newly formed themes were detailed, defined, and named
- **Step 6:** Finally, the analyzed data collected through interview, combined with the literature, was used to design the guideline.

3.4 Unstructured validation interviews

To address the research question "*how to initiate a continuous Process Mining implementation?*" and design a credible guideline, the **findings of this study must be validated.** Therefore, this thesis uses data **triangulation** to increase the credibility. By testing the convergence of empirically collected data, the goal is to overcome biases.

In this thesis, existing literature has been combined with user interviews to create the guideline. The result was then improved and **validated by Process Mining professionals**. To enhance the quality of the study and make the guideline more operational for companies, experts provided some recommendations. For example, to insert metrics and operational pieces of advice in the activities. Based on the enhancement of the findings with the new inputs, the guideline was validated by the experts.

The expert interview approach was adopted because, as Process Mining consultants, those professionals have **experience with several implementations**, in various companies and industries, working together with a wide range of stakeholder. Through multiple applications, and the experts had the opportunity to oversee different implementation strategies. Therefore, their view is generalizable and credible.

In this section, a short introduction of the 4 Process Mining experts involved in the validation interview is provided. Interviews lasted between 1 hour 10 minutes, and 1 hour 40 minutes

- Expert 1: is a researcher and consultant with 10 years' experience in Process Mining. He is a former Process Mining Doctoral candidate. He started his academical career at the Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, under the supervision of Professor Wil van der Aalst, where he published numerous pieces of research. Presently, Expert 1 is the CEO and founder of his Process Mining consultancy company where he supports organizations in implementing Process Mining in a continuous way.
- Expert 2: is a Process Mining specialist and entrepreneur. He created his consulting company 5 years ago. Expert 2' s visions is to democratize the adoption of Process Mining. In his experience, the data localization and mapping are complex stages in the implementation and might discourage some companies. Therefore, he created a Process Mining application that supports companies in localizing and understanding their information system data fields for various processes. The software also helps companies to build the events logs. The vision of the company is to create a community that promotes information sharing while being free.
- Expert 3: is a Process Mining manager working at Heineken. He has more than 14 years of experience in managing and auditing processes and technology. Expert 3: built and is the head of the Process Mining Center of Excellence. This department governs and grow the word-wide Process Mining capability for the company that employs more than 80 000 people, operates in more than 190 countries with 300 different brands.
- Expert 4: is Consultant and PhD Candidate in the PhD for Professionals Program (PPP) at CentER, research institute at Tilburg School of Economics and Management, under the supervision of Hans Weigand. He gained experience with Process Mining through various positions at ASML and Ponthus (IT and data consultancy firm).

The experts were contacted one month before the interview. During the first exchange, the motivations for writing this thesis were presented and the credibility of the experts was assessed through their previous experiences. The researcher asked to the Process Mining professionals if

they would agree to be interviewed as experts to help improve and validate the guideline. The guideline was sent by email before the interview. This enabled the expert to reflect on the artifact prior to the meeting and focus on communicating on the improvements leads and validation during the interview.

3.5 Research Quality

The quality of the designed guideline depends on the quality of this research. Therefore, assessing the quality of this thesis is important. Four terms are commonly employed to assess quality in non-quantitative research. Those terms are: credibility, confirmability, dependability and transferability (Korstjens & Moser, 2017, Houghton et al., 2013, Steinke, 2004). These terms will iteratively be defined, and the quality of this thesis will be assessed accordingly to each criterion.

(1) **Credibility**: is defined by accuracy of research's findings, especially considering the interpretation of participant's original view (Korstjens & Moser, 2017). In this thesis, multiple sources of data were used to enhance the credibility. The literature was supported by a combination of 10 user interviews and 4 validation interviews. Moreover, at the end of each interview, the researcher summed-up all mentioned points to the interviewee, to be certain to not over-interpret the information.

(2) **Transferability**: is defined by the extent to which the findings can be transferred to other situations or participants (Korstjens & Moser, 2017). In this thesis, the transferability is an important criterion as the intend of the guideline is to be used by any company. To improve the transferability, each group of interviewees (users and experts) came from 4 different companies each. Furthermore, each interviewee was asked to provide a thick description. On top of their behavior and experiences, they communicated about the context of Process Mining in their company, and in their personal background. This enables to consider interviewees actions critically and meaningfully during their respective Process Mining implementation.

(3) **Dependability**: is defined by the stability of findings over time and over conditions (Korstjens & Moser, 2017). To increase the dependability of the thesis, the data collected was thoroughly explained (sampling, protocol, interview structure). Moreover, the method for analyzing it was explained and the annotations of the data analysis can be found in Appendix 1: User interviews summary. The research steps were thoroughly explained in the Study design.

(4) **Confirmability**: is defined by the extent to which other researchers may validate this thesis' findings. A confirmable study finds its results from data and not from the researcher's interpretations (Korstjens & Moser, 2017). To enable other researchers to corroborate this research, the guideline has been validated by multiple Process Mining professionals. Moreover, to verify the confirmability of results, all steps of the analysis of data were explicated.

3.6 Research Data management Plan

This sub-chapter is based on a template provided by the University of Turku. It helps researchers managing the data used in this thesis. The researcher of this thesis paid great attention to planning the data collection, requesting the right to collect data (interviewee's consent, safety and structure of storage both during the research and after…). The completed template is available in Appendix 2: Research data management plan.

4 ANALYSIS AND RESULTS

In this Chapter, the **IPMI guideline (Initiating Process Mining Implementation)** and its **activities** will be presented. All the **tools, metrics, assessments methods and operational examples** will be detailed.

Based on the literature and the 10 user's interviews, the IPMI methodology has been created. Afterward, the guideline has been improved and validated during the validation interviews with 4 Process Mining professional consultants. The goal of the IPMI methodology is to help companies initiate a continuous Process Mining implementation. By using this guideline, companies have the opportunity to technically and socially embed Process Mining as an organization-capability.

For the purpose of this guideline, existing theories and models have been adapted. The adaptations and the reasons for the change are hereby explained.

- DIDOV framework: Experts 1 and 4 suggested to use an adaptation of the DIDOV framework instead of the original one. The modification consists in interchanging the Optimize and the Verify phases. The rationale is that the adaptation corresponds better to the context of Process Mining. The Experts emphasized the importance of continuously verifying all steps taken in the implementation and then optimize them. Hence, it makes more sense in the Process Mining context to first Verify the activities performed, and then to optimize them.
- ADKAR model for change management: User Interviewee [B, C, E, F and I] specifically mentioned the importance of change management when implementing a new technology. Therefore, an adaptation of ADKAR model will be used. There are two reasons for using an adaptation of this theory instead of the original one. The first reason is the scope of this thesis. As this paper focuses on the initiation phase of the implementation, the last step of the ADKAR model (Reinforcement) is by design excluded from this thesis. The way the ADKAR model will be adapted is by changing the order of the activities to fit into the adaptation of the DIDOV model.

4.1 Overview of the IPMI guideline (Initiating Process Mining Implementation)

The **scope** of the guideline begins with the definition of the vision for using Process Mining in the company. It ends with the sponsor's agreement to perform the analysis, if and only if, all activities of the 3 pillars are successfully performed.

The IPMI guideline is represented in Figure 14. The guideline is composed of 5 main phases which are Define – Identify – Design –Verify – Optimize, corresponding to the adaptation of the DIDOV

model. The 5 phases are happening **sequentially, from left to right**. Therefore, the first phase is *Define* and the last phase is *Optimize.* At the end of the last phase is depicted a feedback loop connecting back to the first phase (Define). This arrow indicates that if the activities taking place in the last phase are not verified, it is possible to start the DIDVO cycle again. There is also an arrow pointing to the right: if the activities are successfully verified, the extraction can happen, and the analysis can be performed (outside of the bellow-presented guideline).

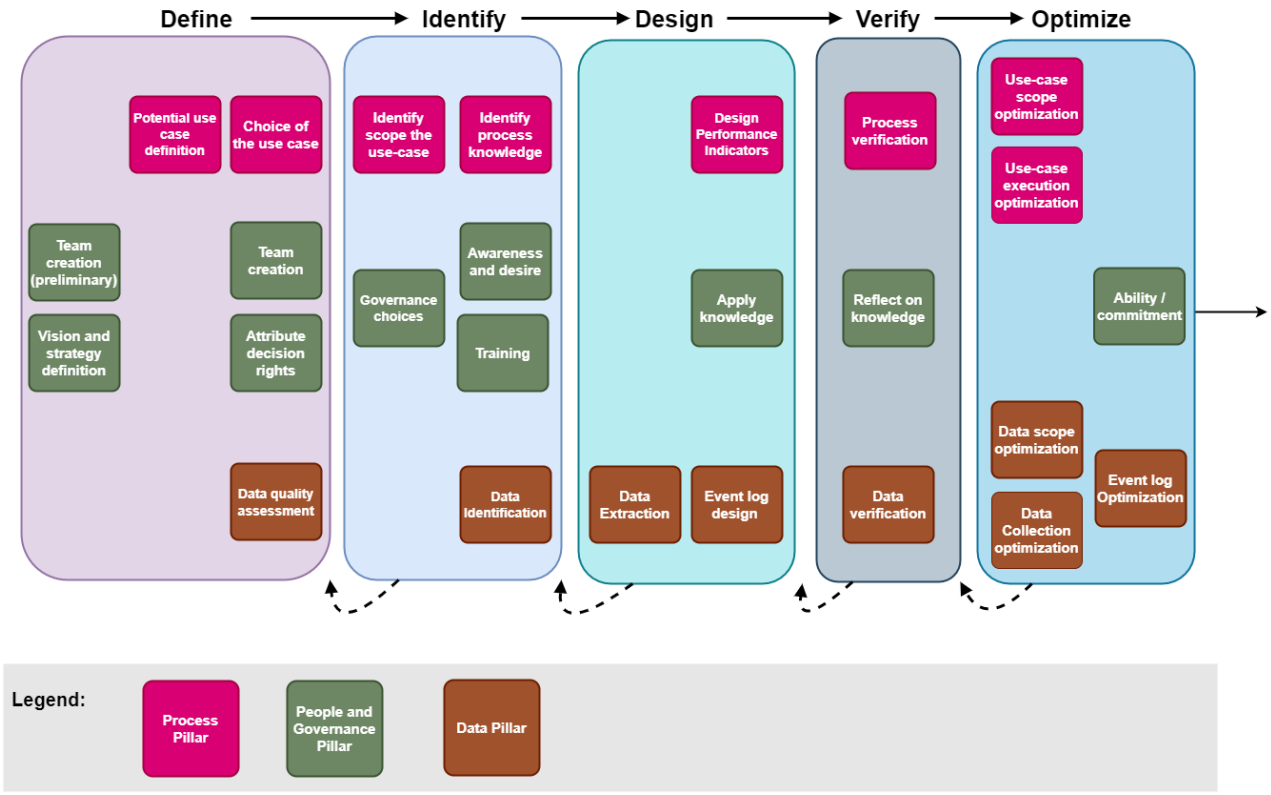


Figure 14: IPMI guideline (Initiating Process Mining Implementation)

The activities are grouped in 3 pillars, represented by the pink, kaki, and brown small rectangles. The pillars are: *process, people and governance* and *data*. The activities linked with each pillar are interacting with the other pillars. The pillars considered are the foundations of this guideline. They are the main pain points of a Process Mining implementation identified during the interviews and through the literature review. For a successful and continuous Process Mining implementation, the pillars must synergize with each other. An assumption made in this guideline is that the technology (the Process Mining software) is easily accessible. As a matter of fact, there are many vendors worldwide and there also exists open-source Process Mining software such as ProM.

Similarly, to the phases, the activities are happening sequentially from left to right. Therefore, the first activity is the sanitary check, and the last activities are *process verification, ability,* and

data verification. Some activities are depicted on top and under each other: they happen simultaneously without any specific order required.

In Table 10, for each activity, the **input, output, and the most impacting roles** are identified. Each activity is also connected to its corresponding DIDVO phase. However, users of this guideline should keep in mind that all roles should closely be involved.

DIDVO phase	Activity	Input	Output	Key roles
Define	Team creation (preliminary)	Employee's functions and characteristics	Fixed team members	Sponsor
	Vision and strategy definition	Company's general vision	Clear vision and purpose for using Process Mining	Sponsor
	Potential use-case definition	Company's processes and Business Process Management software	Set of potential use-cases.	Business Process Expert
	Choice of the use-case	Set of potential use-cases	Chosen use-case	Project manager
	Data quality assessment	Information systems	"*Yes or no advice*" to launch a Process Mining project	Performed by: Information System expert Confirmed by: Business Process Expert
	Team creation	Employee's functions, characteristics, and process chosen	Variable team members, complete team.	Project manager
	Attribute decision rights	Complete team	Definition of decision rights	Sponsor
Identify	Governance choices	Decision rights	Project governance, Resources enablement	Persons defined as owning decision rights
	Identify scope of the use-case	Chosen use-case	Narrowed use-case	Project manager
	Awareness and desire	Governance choices and the team members	Awareness and Desire building	Sponsor, project manager
	Training	Resources enabled by the project governance	10% of knowledge on Process Mining	Process Mining expert

	Identify Process Knowledge	Chosen use-case	Process mapped	Project manager, business process expert, manager of the execution of the process
	Data localization and mapping	Use-case chosen, data quality assessment and information system.	Relevant data fields identification	Project manager, business process expert and information system expert
Design	Data extraction	relevant data fields identified.	Data extracted	Information System expert
	Event log (preliminary)	Data mapped and the 10% of knowledge	Preliminary event log	Information System expert
	Performance indicator	10% of knowledge, use-case and data mapped	Statical and dynamical performance indicators	Project manager, Process Mining expert and Business Process Expert.
	Apply knowledge	10% of knowledge	70% knowledge	Process Mining expert
Verify	Process verification	Performance indicator and broken-down process and a prioritized plan for the process	Agreement to move to the phases of Process Mining	Project manager, Business Process Expert and Process Mining expert
	Data verification	Preliminary event logs and the performance indicator	Leads for improving the data pillar	Project manager
	Group reflection	10% and 70% of knowledge	20% of knowledge	Project manager
Optimize	Use-case scope optimization	Process mapped and the performance indicator	Reduced use-case scope and prioritized plan	Project manager and the Process Mining expert
	Use-case execution optimization	Preliminary event log	Optimized process execution	Project manager, business process expert
	Data scope optimization	Event log and the data mapping	Optimized data scope	Information System expert, Process Mining expert, project manager
	Data collection optimization	Preliminary event log	Optimized data collection	Project manager, business process expert and information system expert

	Event log optimization	Preliminary event logs and performance indicator	Complete event log	Project manager, Process Mining expert, Information System expert
	Ability	Change management and project governance activities	Sign-off moment to pursue implementation. Final milestone.	Project manager, sponsor

Table 10: IPMI, inputs, activities, outputs and key roles

In the following chapters, each activity is detailed and attributed to their key team members.

4.2 Define phase

The Define phase combines five steps, those steps are the foundation of any Process Mining project. They enable to define if the company is ready for Process Mining, to define the project with the right process and the right team members and initiates the reflection on the project governance.

4.2.1 Team creation (preliminary)

The goal of this activity is to **define the fixed team members**. *Fixed team members* refer to members who will be part of the project regardless of the chosen process. The inputs for selecting the right team are the employees' functions and characteristics. The fixed team members are selected via a 2 rounds assessment.

They will primarily be chosen accordingly to their function in the company. The four described roles were mentioned by all User interviewees, and then confirmed during all Validation interviews.

If there are multiple potential employees eligible for each role in the project, the second selection criteria are based on the employees' characteristics. This argument was provided by User interviewee [D]. [D]'s bottom line is "*we should always make the project evolve in the direction of people's enthusiasm*". Table 11 depicts the functions and roles of the fixed team members that should be involved in the project.

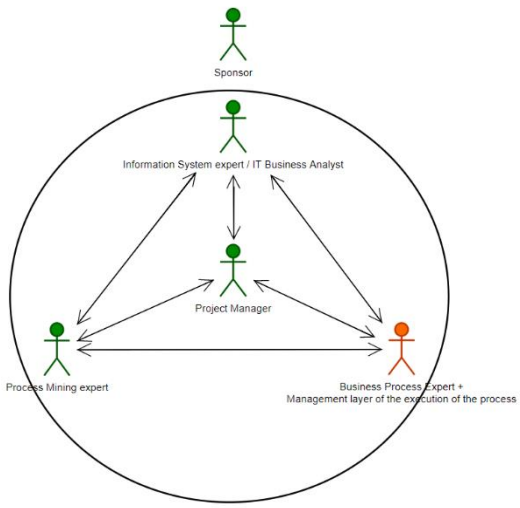
1st selection round: employee's functions / roles	
Function of the fixed team members	**Role of the Fixed team members**
The project manager.	Has a central role in the implementation, they are accountable for the making the project successful. They will coordinate the dialogue between the other team members, facilitate the access to resources and report to the sponsor.

	For example, the project manager can come from the operational excellence department or an internal consulting department.
The sponsor.	Is responsible for the project. They are the customer of the project as they will show the success to the outside world. The sponsors are a credible source for communicating about the project outside of the team. The sponsor is responsible for the main choices and authorizes the access to resources. For the Process Mining to be useful, the sponsor must have the authority to modify the execution of the process.
The Information System expert / data engineer	The information system expert / data engineer technically knows what fields of data are collected, where they are stored, how to access and extract them. This role can also be referred to IT Business Expert.
The Process Mining expert.	To initiate a successful Process Mining implementation, this role is key. It facilitates the right decisions from the beginning with consideration of the Process Mining requirements. The Process Mining expert can share best practices, provide pieces of advice, and help in overcoming challenges or doubts. If the Process Mining Expert is an employee and has the required skills, it is possible to merge this role with the Project Manager's role (this is impossible if the Process Mining expert is a consultant for example).
2nd selection round: employee's characteristics	
Enthusiastic about the project	
Previous successful experience in Business Intelligence / digital skills	
Ability to communicate with various stakeholders (from different expertise domain or hierarchy level)	
Ability to think in term of processes	
Is willing to free-up time/ go the extra mile for the project	

Table 11: defining the right fixed team

There is the possibility that two roles can be occupied by one person. For example, it would not bring any additional challenge if the project manager were also the Process Mining expert.

Figure 15 shows in green the fixed team members, the arrows represent the members collaborating with each other during the implementation. The project manager is placed in the center as



they are the one facilitating the exchanges and being responsible for moving the project forward. The sponsor is placed outside of the circle as they are not involved in the daily project and only takes high level decisions

. The outcome of this activity is the definition of fixed team members. Ultimately, the sponsor is the one responsible for gathering the team and enabling them to spend time on the implementation.

Figure 15: team members' relationship structure

4.2.2 Vision and strategy definition

The goal of this activity is to **define a vision and a strategy for initiating a process mining implementation**. This activity should be performed by the sponsor of the project. The input for this activity is the company's general vision. To continuously implement Process Mining and scale it up from a pilot project to a company's capability, the project should be initiated by a high-level strategy and vision for the use of Process Mining. The best way to implement Process Mining is to implement it as a tool that will support the general company's vision and strategy. This activity enables to define a plan to embed Process Mining both technically and socially in the company. To help companies defining their Process Mining strategy, this guideline provides a template that should be filled (Table 12). Process Mining implementations should be led by questions (van der Aalst, 2016), therefore, this template was designed accordingly to the guiding principle.

Currently, where is the company?	What is the mission of the company? What are the values of the company? What is the company's relative position in their industry? (This can be done via a SWOT analysis)
In the future, where is the company going?	What is the company's vision? What improvements should be achieved to realize the vision?
Is Process Mining the right tool?	Can Process Mining support the company's vision achievement? How? What is the purpose of using Process Mining in the company's context?

Table 12: Vision and strategy definition template for using Process Mining

The outcome of this activity is a clear vision and purpose for using Process Mining in the company.

4.2.3 Potential use case definition

The goal of this activity is to **define potential use cases to initiate the Process Mining implementation**. The input for this activity is the company's processes and Business Process Management software.

Step 1: In this activity all existing value chains and their processes in the company will be mapped to define potential use-cases. Figure 16 depicts how the 3 elements relate to each other in their different granularity levels. Figure 17 provides an example for more clarity. In Figure 16 the value chain is *Order to Cash* (presented in white at the top-left corner of the Figure). When zooming in, processes like *Strategy to Supply Chain* or *Forecast to Plan* are depicted (at the top-center of the Figure). At the highest level of granularity (the most specific), the use cases are presented: *Schedule production, Order materials* … In Figure 17, the use-cases presented are part of the Plan to Finished Good process and are displayed in the center and bottom of the Figure).

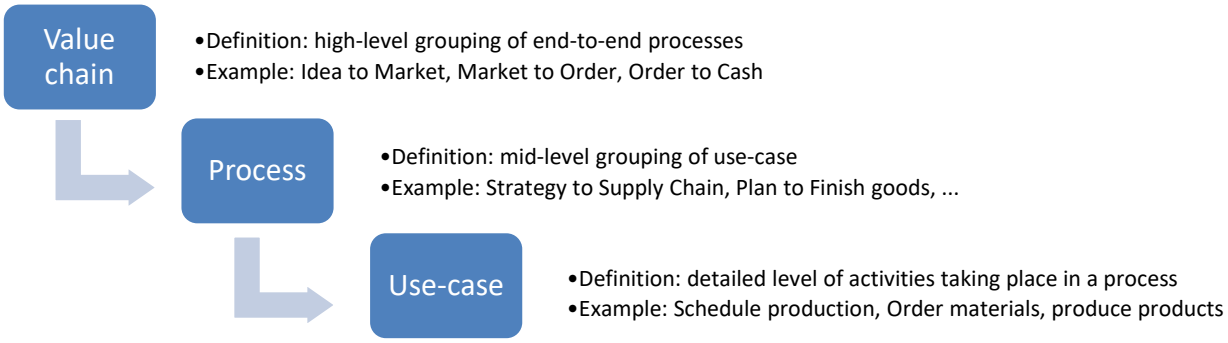


Figure 16: Relation between value chain, process and use case

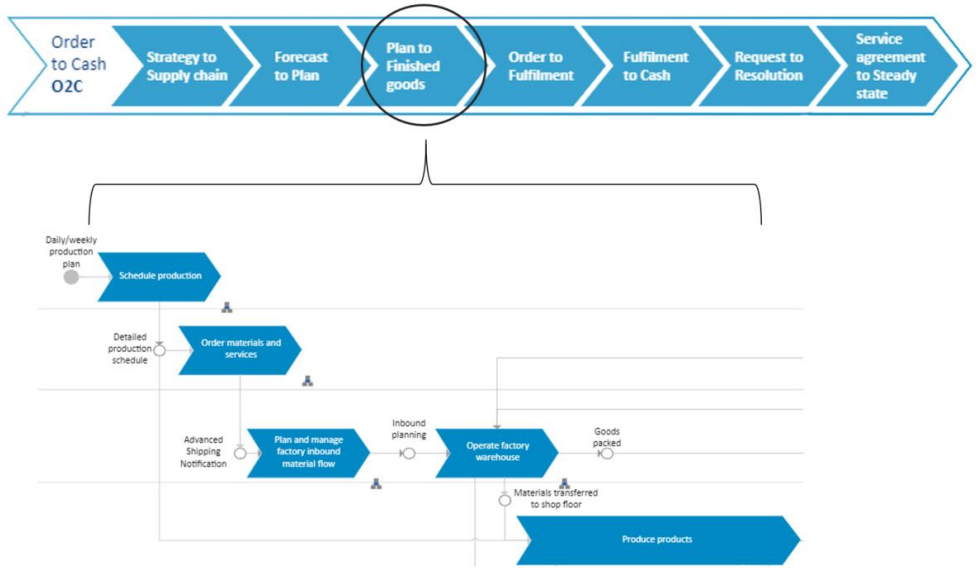


Figure 17: Example of Value Chain, Process and Use-Case

As per this example (Figure 17), the Business Process Expert should map the different processes and use-cases in their value chain. To facilitate this step, the Business Process Expert should exploit knowledge from the company's Business Process Management software.

Step 2: the Business Process Expert should interview with the manager of the execution of each process. The goal of this interview is to define some high-level pain points / business problems. As a Process Mining implementation should always be led by questions, those problems should be turned into research questions. The research questions will be useful during the implementation to identify the steps of the process, the data flowing and therefore to design the performance indicator and the event logs. It is important to define research questions to understand the root causes of the problems. Some examples of research questions can be "*why are the deliveries often late on week-ends?*" or "*why are the invoices often sent without approval?*".

Afterward, to keep a structured overview of how Process Mining can help this use-case, the use-cases should be categorized into types of benefits. This guideline found 5 main types of benefits:

- Process improvement
- Risk and Control
- Performance Measuring
- Automation Opportunity Identification
- Digital Transformation (Example, support in new ERP implementation)

Defining a benefit is important, because when a process will be chosen: the benefit will act as a north star. The list of benefits provided by this guideline can be enhanced accordingly to the companies specifies.

Step 3: the Business Process Expert and the Project manager should combine all information in the *prioritization of potential use-cases* table (Table 13). This table will be enriched during the next activities of this guideline. In this table, all use-cases are firstly grouped by value-chain, and secondly by process. Then, at least one benefit per use-case should be defined.

A template for the Prioritization of Potential Use-Cases is provided in Table 13.

Value Chain	**Process**	**Use Case**	**Benefit**
Order to Cash	Process 1	Use-Case P1 A	Risk and Control
		Use-Case P1 B	Process improvement
	Process 2	Use-Case P2 A	Automation Opportunity
		Use-Case P2 B	Risk and Control
		Use-Case P2 C	Digital Transformation
…	…	…	…

Table 13: Prioritization of Potential Use-Cases

The outcome of this activity is a set of potential use-cases.

4.2.4 Choice of the use-case

The goal of this activity is to **choose one or two use-cases to launch the Process Mining project**. The input is the set of use-cases defined in chapter 4.2.3. To define the chosen use-case, different approaches have been recommended by the interviewees. All 3 suggested methods are useful and have been validated by their practical and successful use in companies. As the choice of the process is a crucial step in the process mining implementation, all 3 methods suggested by the interviewees will be presented. However, the first method is recommended as it sits between the simplest but least detailed method (2nd method) and the most complete but most complex method (3rd method). Any of the 2 other methods (presented in 4.2.4.2 and 4.2.4.3) can be used in complement to the first method (4.2.4.1) to confirm the choice of the use-case. As all methods are valid, the user of this guideline might prefer choosing the 2nd or 3rd method instead of the first if they match the company's situation better. Therefore, the up and down sides of each method will be discussed to help users of this guideline to make a conscious choice.

4.2.4.1 *Use-case choice: Method 1, balance between strategic alignment, scalability, impact, data, Process Mining necessity (Expert 3)*

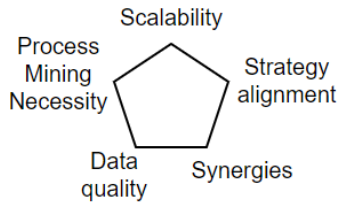


Figure 18: Use-case choice, method 1

This first method to choose a use-case balances 5 criteria: strategic alignment, scalability, synergies, data quality and Process Mining necessity (**Error! Reference source not found.**). This method has b een introduced by Expert 3 during a validation interview.

For this method, the Table 13: Prioritization of Potential Use-Cases created in 4.2.3, should be used as an input and will be enhanced in this activity. Each potential use-case should be assessed on a scale from 0/5 to 5/5 with the following criteria: strategic alignment, scalability, synergies, data quality and Process Mining necessity.

The scalability can be defined in two ways, both are important. Firstly, it is easily possible to focus on a small part of the process and then look at the bigger picture. For example, it is possible to start with a small use-case, and then add more steps to capture the end-to-end process. The second way to define scalability is: the use-case happens in a similar manner in another location. For example, in another factory, in another market… It is therefore possible to re-use most of the knowledge and amortize efforts and investments put into the implementation to have a bigger return on investment.

The strategic alignment refers to the alignment between the type of benefit for each use case and the company's general vision and strategy for the coming years. It is important that Process Mining is used as a tool to support the company's visions.

The synergies criterion refers to the synergies between the benefits of process mining on a specific use-case and the general initiatives taken to improve this same use-case in the company. It is recommended (especially the first times) to perform process mining as part of an existing use-case improvement initiative: this helps process mining to receive more sponsor's priority, resources and convince stakeholders of Process Mining potential benefits. It also favorizes trust and engagement from stakeholders and team members.

The data quality criterion refers to the quality of data evaluated in the Data quality assessment activity presented in Section 4.2.5. For each use-case the quality of data should be defined and used as a criterion to choose the right process. For the first implementation, it is recommended to choose a use case where data comes from one single information system, rather than multiple ones. This is because it will simplify the event log creation. It can be argued that the potential for improvements is bigger for a process that lays in multiple information systems/emails/ Excel files

than a process that lays into one single information system. For example, the chances of manual entry errors are bigger. This argument is valid but taking into consideration that the audience for this guideline targets a novice audience: multiple sources of data increase the complexity and therefore the chances of failure. If the first Process Mining project of the company fails, the trust in the Process Mining tool will be broken. The suggestion is to start with a use-case that generates data in one single information system. Then, in the future when the process mining capability will be mastered, choose processes that generates data in multiple information systems.

The Process Mining necessity: Process Mining can be expensive and / or require a lot of efforts to invest. Therefore, it is important to first reflect on: is process mining the only tool / technology that can enable the achievement of the desired benefits for this use-case?

To complete this activity, and choose one use-case, the following template should be completed by the companies using this guideline (Table 14). The use-case having the highest grade should be chosen. This table was originated in section 4.2.3. The use-case with the highest grade should be chosen.

Value Chain	**Process**	**Use Case**	**Benefit**	**Scalability**	**Impact**	**Data quality**	**Process Mining necessity**	**Synergies**
Order to Cash	Process 1	Use-Case A	Risk and Control	/5	/5	/5	/5	/5
		Use-Case B	Process improvement	/5	/5	/5	/5	/5
	Process 2	Use-Case C	Automation Opportunity	…	…	…	…	…
		Use-Case D	Risk and Control	…	…	…	…	…
		Use-Case E	Digital Transformation	…	…	…	…	…
…	…	…	…	…	…	…	…	…

Table 14: template for choosing the use-case, method 1

This method is chosen as it takes the most relevant criteria into considerations and enables the users of this guideline to choose the right use-case while being able to compare them at the same

time. This methodology has the best ratio completeness/ complexity compared to the methodologies presented in 4.2.4.2 and 4.2.4.3.

4.2.4.2 Use-case choice: Method 2, balance between frequency, impact, and complexity (expert: 2)

This alternative method for choosing a use-case is discussed in Appendix 3: Use-case choice: Method 2, balance between frequency, impact, and complexity (expert: 2).

4.2.4.3 Use-case choice: Method 3, three rounds assessment

This alternative method for choosing a use-case is discussed in Appendix 4: Use-case choice: Method 3, three rounds assessment.

4.2.5 Data quality assessment

The goal of the data quality assessment is to answer the question: ***is the company ready for Process Mining?*** Or more specifically, ***is the use-case ready for Process Mining?*** The inputs for this activity are the information systems**.** This activity should be carried out by a technical team such as Information System experts or Data Engineers. It would be useful to present the results of this assessment to the Business Process Expert to confirm that this assessment is indeed reflecting the reality. This activity should be performed at the same time and as part of the choice the use-case (section 4.2.4). Table 15 presents a tool for data quality assessment to be used on each use-case. The tool is an adaptation from the Table 4: Maturity levels for event logs (Source: IEEE TASK FORCE ON PROCESS MINING, 2012) presented in section 2.2.1.

Level	**Characterization**
5/5 Highest level	In the company, data is of excellent quality (trustworthy, complete, well defined). The data is generated and stored in an automatic, systematic, reliable, and safe manner. The collected data addresses privacy and security considerations. The collected data has clear semantics, which implies the existence of ontologies.
4/5	In the company, data are recorded automatically systematically, and reliably. They are trustworthy and complete. Process instances are supported, which means that information does not need to be extracted from a variety of tables but is combined into one.

3/5	In the company, data is recorded automatically but not systematically. The present data is likely to be trustworthy. However, the collected data is not complete. There is no structured way to collect data, which lays in various tables.
2/5	In the company, data is recorded automatically but is not treated as primary importance: there is no systematic approach to record data which is incomplete and only poorly trustworthy. It is possible to bypass information in the system: many data are collected inadequately or not collected at all.
1/5 Lowest level	In the company, most data are not recorded. The few collected data are likely to be collected by hand. They do not match reality, data is missing.

Table 15: adaptation of the maturity levels for event logs (Source: IEEE TASK FORCE ON PROCESS MINING, 2012)

If the company assess the quality of their data on the specific use-case as level 4/5 or 5/5: the use-case is considered are ready to initiate a Process Mining project.

If the company assess the quality of their data collection procedures on the specific use-case as 3/5, the use-case readiness to initiate Process Mining is considered as on the edge. Further discussions should be opened to also consider criteria like "*is top management supporting the project?*", "*is there a group of employees ready go to the extra-mile to carry out the project?*" … At this point, only the company can make the decision whether they are willing to put in consequent efforts to make the project successful.

If the company assess their data quality on the specific use-case as 1/5 or 2/5: this guideline recommends that the company first improve the general quality of their data before thinking about a Process Mining project. Process Mining is only recommended to companies considering data as first-class citizen.

The output of this activity is a *yes or no advice* for launching a Process Mining project. If the questions are answered negatively,

4.2.6 Team creation

The goal of this activity is to define the variable team members. The term *variable team members,* refers to the members who are involved because of their knowledge on the chosen process. This activity should be performed by the project manager. The inputs for selecting the right team are the employee's functions and their characteristics but also the process chosen. The variable team members are selected via a 2 rounds assessment: The they should primarily be chosen accordingly to their function in the company. If there are multiple potential employees eligible for each role in the project, the second selection criterion is based

on the employees' characteristics. Table 16 depicts the functions and roles of the variable team members who should be involved in the project.

Figure 15 shows the variable team members in orange, collaborating with the fixed team members during the implementation.

The outcome of this activity is the variable team members of the project. The fixed team members and the variable team members together are forming the complete implementation team.

1st selection round: employee's functions / roles	
Function of the variable team members	**Role of the variable team members**
The Business Process Expert.	This role is the bridge between the business side and the technical side. Business process experts know how the process is supposed to happen, and how the data is collected in the process.
The management layer of the process execution.	Through interviews, this role can contribute to clarify the execution to the process. The managers of the execution of the process are the ones feeling the pain. They can explain the daily execution of the process in a way that is closer to the reality than what is statically documented.
2nd selection round: employee's characteristics	
Enthusiastic about the project	
Previous successful experience in Business Intelligence / digital skills	
Ability to communicate with various stakeholders (from different expertise domain or hierarchy level)	
Ability to think in term of processes	
Is willing to free-up time/ go the extra mile for the project	

Table 16: Function and role of the variable team members

4.2.7 Attribute decision rights

The goal of this activity is to **define and attribute the decision rights for the project**. The input is the complete project team. The sponsor is responsible for this activity.

In this activity, the owner of the decision right is defined, but the decisions themselves will be taken in 4.3.1: Governance choices. The governance choices will mainly concern the strategic role of the technology in the business, technical choices, resources provided for the implementation and prioritization.

Who **can** own the decision rights? The owner of the decision rights can be business executives, or IT executives, or operating business unit. It can also be a combination of some (or all) of the roles previously mentioned.

Who **should** own the decision rights? One single answer cannot be provided to this question. As a matter of fact, Expert 3 recommends that the business executive owns the decision rights, and, especially, that the Business Units should never own them. Expert 3's rationale is that Process Mining implementation can cost a lot, both in terms of money and time. Therefore, the business executives must make sure to control the investments, to favorize economies of scale, to build a standard capability in the headquarters and to create a common way of working among the different business units. Expert 1 recommends that the business units own the decision rights. Expert 1's rationale is that Process Mining is likely to be rejected if the users of the technology do not have the possibility to customize its usage to their own needs. Finally, Weill and Ross (2004) claim that the most successful combination of decision rights ownership is IT executive, with either business executives or business unit's leaders.

Consequently, this guideline recommends a different perspective: most of the efforts should not be put into *who* owns the decision rights, but into maintaining the information and requirements aligned between all stakeholders (The Global State of Operational Excellence, 2021, pp. 75–76)

Finally, this guideline will still provide some characteristics that can help companies decide how to organize their decision rights regarding the implementation. The ownership of decision rights might vary accordingly to: strategic and performance goals, organizational structure, governance experience, size and diversity, industry and regional differences (Weill and Ross, 2004).

The outcome of this activity is the definition of the decision rights for the project.

4.3 Identify

The Identify stage breaks down the project into smaller and manageable pieces. To do so, the process and the data are mapped. The goal of the project and its needed resources are clearly identified. In order to prepare for the next phase (Design), the knowledge should also be trained.

4.3.1 Governance choices

The goal of this activity is to **identify the governance choices for the implementation**. The input for this activity is the decision rights defined in 4.2.6. This activity is considered as being of pri-

mary importance by this guideline. This is because establishing a strategy for the governance enables projects to grow from being experimental to being widely operationalized. For each subject, the team members owning the decision rights should identify a strategy for governing the project.

This guideline lists some key decisions that should be answered to help companies. Users of this guideline are encouraged to complement and enhance this list accordingly to their specific situation.

Key governance decisions
What will be the strategic role of Process Mining in the company?
What is the envisioned planning for the implementation?
Who will be the user of Process Mining? (Will each business unit continuously monitor their activity, will a central team of analyst generate reports for the business units, …) This decision impacts the cost of the project with the number of Process Mining software licenses offered and the training provided, but it also impacts the transparency of information in the company and the trustiness toward the users.
Is the company ready to change and enhance their processes in the future, accordingly to the analysis' findings? How will this happen?
Will the project be built internally, externally (with / without the help of consultants)? If the project is built externally, when, and how will the competency be moved in-house?
What is the budget secured for the project?
What is the time allowed for the team members to spend on the project? Here it is important to note that all user interviewees pointed out that having time is the most crucial resource for making the implementation successful. This is because Process Mining is broad, it requires knowledges and involvement from a multiplicity of stakeholders.
Will Process Mining be utilized instead, or on top of traditional dashboards? Will this take place in two different tools, or all in the Process Mining tool, or all in the traditional dashboard?

Table 17: Key governance decisions

The outcome of this activity is the identified project governance and the enablement of resources.

4.3.2 Identify scope the use case

The goal of this activity is to **identify the scope of the use-case**. The input for this activity is the chosen use-case. The activity should be performed by the project manager.

A finding from the user interviews is that a use-case with a narrower scope has more chances to make the implementation successful. It is best practice to narrow down the scope of

the use-case. If a use-case happens in multiple locations, factories, business units, operating companies, only one location should be used for initiating the Process Mining implementation.

One concrete example of this is: a company owns multiple factories. All factories must purchase and pay goods to their suppliers (this is the use-case). Therefore, the company should choose to initiate the process mining implementation in only one of the locations instead of working on all of them at the same time.

The rationale is that the use-case will be easier to map, and the data will require less efforts to understand and localize.

The output of this activity is the narrowed use-case.

4.3.3 Awareness and desire

The goal of this activity is to **build awareness and create desire** around Process Mining. The input for this activity is the governance choices and the team members. The project manager and the sponsor should work together to communicate with stakeholders from outside of the implementation team.

Based on Expert 3's inputs, this guideline proposes to build a communication plan based on the characteristics of the stakeholders.

Stakeholder	**Subject of communication**	**Tool**
High level executives	High level possible savings, return on investment…	Quarterly business reviews
1 champion per use-case (representative operational use-case executor)	Working on identifying incontestable reasons for change and potential benefits of Process Mining. Sponsorship, Coaching	Bi-weekly working group Microsoft Teams
Community of Process Mining (future) user	Sponsor communicates about reasons for change and benefits. The credibility of the sender is essential. Preventing circulation of misinformation. Understand factors of resistance to change (impact for own situation, fear of losing comfort, do not understand their personal gains…) and deal with it (for example by showcasing the personal gains for each specific situation (less manual work, gain time…)	Internal Yammer posts: success stories, small trainings, newsletter, advancements in implementation

Rest of the company	Communicating on implementation's milestones achievement and small successes.	SharePoint publications

Table 18: Awareness and Desire communication plan

The outcome of this activity is the building of awareness and the creation of desire.

4.3.4 Training

The goal of this activity is to begin with **developing the knowledge**. Training the team members from an early stage enables to make insightful choices and to build strong competencies on a longer period. This also offers the possibility to manage resistances and empower members who are the least comfortable with technologies. The Process Mining expert should enable trainings for the team members involved in the implementation and for the final users of the Process Mining tool. The input for this activity is the resources enabled by the government choices.

This guideline recommends some resources to train knowledge.

Type of knowledge	**Who**	**Direct access**
Course	Coursera (courses platform), Eindhoven University of Technology and Professor Wil van der Aalst	*Process Mining: Data science in Action*
	Celonis (vendor) and Wil van der Aalst	*Process Mining, From theory to execution*. *Celonis also offers free online trainings on both their software and the Process Mining discipline in general: Celonis Academy*
	Future Learn and Eindhoven University of Technology	*Introduction to Process Mining and ProM. ProM is the open-source Process Mining framework created by Wil van der Aalst.*
	More courses can be found on the official Process Mining website: processmining.org	
Book / Literature	Wil van der Aalst	***Process Mining: Data Science in Action***

		The first chapters are accessible to non-specialist readers.
Conference	International Conference on Process Mining	The Process Mining Conference Series
Reliable Internet Websites	Official website	processmining.org
	Task Force on Process Mining	IEEE Task Force on Process Mining
	Gartner	Gartner Market Guide for Process Mining
General advice:		
If the implementation is built with the help of a vendor, they will offer trainings If the project team is composed of a Process Mining expert, their knowledge should be exploited, and an in-house training is encouraged to be provided		

Table 19: Resources to train knowledge

Only 10% of the knowledge is gained from trainings. The rest of the knowledge is gained from applying the knowledge and reflecting on it (Lombardo et al., 1996). Therefore, the output of this activity is 10% of knowledge on Process Mining. The remaining knowledge will be provided in later activities to ensure 100% of knowledge at the end of the implementation.

4.3.5 Identify Process Knowledge

The goal of this activity is to **exploit** as much as possible the **knowledge of the process**. This enables to gain time when locating and mapping the data and when the designing the performance indicators. The input for this stage is the chosen process.

To reach the goal of this activity, this guideline recommends filling the SIPOC (Table 20) of the use case (Supplier, Input, Process, Output, Customer). The SIPOC is a Six Sigma tool. This guideline provides an explicative template to help companies. For more explanations, an example of a filled SIPOC is available in Appendix 2

S	I	P	O	C
Supplier	**Inputs**	**Process**	**Output**	**Customer**
Who supplies / enables the process?	What are the resources needed	What are the steps of the process?	What is created / what results	Who benefits from the process?

	for executing the process?		from the process?	

Table 20: SIPOC of the chosen use-case

To fill the SIPOC, the project manager should conduct focus groups and/ or interviews with:

- The business process expert: to identify all steps of the process at a high level and map them. This enables a clear identification of all steps in the process mapping.
- The manager of the execution of the process: to identify all the pain points, the needs, the constraints, and the steps on a lower level. This enables a thorough understanding of the process.

Each step and activities of the chosen process should be identified and mapped. If the process is continuously monitored through a Business Intelligence dashboard, the Key Performance Indicators (KPI) should also be identified and mapped.

The Process Mining expert could also support this activity. The outcome of this activity is the process mapped and optionally the performance indicator identified.

4.3.6 Data localization and mapping

The goal of this activity is to **translate the execution of the process from the physical world, into the execution of the process in the digital world**. In this activity, the team identifies and create an overview of the data flowing in the Information System based on the chosen use-case. The input for this activity is the process chosen, the data quality assessment and the information system. This guideline recommends performing this activity in parallel with the process mapping. This activity should be performed by the project manager animating a workshop with the business process expert and the information system expert.

Step 1: the team members should identify in which information system the use-case is laying.

Step 2: for each step of the process, the transactions operated in the physical world should be translated to the digital world. In other words, the Business Process Expert should help translate business language into technical data fields language.

Step 3: those digital transactions should be localized and mapped in the information system. This enables to have an overview of what data field exist. Identifying which data will be extracted beforehand is important. An Information System such as SAP has more than 10 000 tables, therefore methodically identifying which data are relevant is key to not discourage the team.

Important: a very important piece of advice at this stage is to write down each table accessed in the Information System. Documenting the accessed data enables auditability: when creating the

event logs (in the next phase), the Information System expert should be able to explain where the data comes from at any moment. Starting to document the data prior to the event log creation enables to adopt a structured approach and help non-IT stakeholders to understand data as well. On top of the auditability criteria, documenting data supports reproducibility and scalability.

The outcome of this activity is the identification of relevant data fields used in each transaction during the execution of the use-case.

4.4 Design

The Design transforms ideas into tangible and monitorable indicators. In this phase, the knowledge built previously is applied by extracting the identified data, designing the event logs and the performance indicator. Those topics will be developed in this chapter.

4.4.1 Data extraction

The goal of this activity is to extract the data, so that they event logs can be created in the next activity. The input for this activity is the relevant data fields identified. This activity should be performed by the Information System expert.

When extracting the data from various Information Systems sources such as SAP or Aris (a Business Process Management software), the data must be stored in XES format (eXtensible Event Stream). XES is the official interchange format for event logs adopted by the IEEE Task Force on Process Mining. To extract data directly in XES format, this guideline references XESame as a tool. XESame is relatively easy to use as it does not require any programming skills (Verbeek et al., 2011, p.60-75).

If the use-case chosen is monitored through multiple information systems, this guideline recommends using Task Mining. It will help create the event logs and include steps of the use-case that were not integrated in the Information System and therefore would not usually generate event logs. Task Mining uses Optical Character Recognition (OCR), Natural Language Processing (NLP) and Machine Learning algorithms. Task Mining tools can be installed on employee's computers to collect data like clicks, emails, consulting a spreadsheet… On this base Task Mining software can generate useful data for creating event logs such as timestamp, activity, case ID…

The output for this activity is the data extracted.

4.4.2 Event log (preliminary)

The goal of this activity is to design a first version of the event logs. The inputs are the data extracted and the 10% of knowledge on Process Mining. This activity should be carried out by the Process Mining expert, the project manager and the information system expert.

Step 1, **transforming the data**: if the data comes from multiple sources, it must be transformed to fit operational needs such as syntax and semantics and quality level (van der Aalst, 2016, p.142). This can be done by loading the data into the company's data warehouse to unify the data.

Step 2, **designing the event logs**: based of the unified data, the event log must contain for each activity of the use-case a timestamp, and the unique case identifier that is followed through all steps of the use-case. If a case identifier is not available at this stage, the information system expert should at least make sure that the event is sorted by chronological order.

The outcome of this activity is a preliminary event log.

4.4.3 Performance indicator

The goal of this activity is to **design a set of performance measurement** to assess the success of the Process Mining implementation. The input for this activity is the process and the data mapped, and the 10% of knowledge on Process Mining. This step should be executed by the project manager, the Process Mining expert, and the Business Process Expert.

This guideline helps to design the performance indicators. Two kind of performance indicator are identified: the static ones, and the dynamic one.

1/ Traditionally, static key performance indicators (KPI) are already created in companies. They refer to the quality, the speed or the cost. Those KPI were exist without Process Mining.

2/ Process Mining will unlock access to dynamic performance indicator that should be designed before the implementation.

Concretely, what are the dynamic performance indicators and why are they important?

There is not one answer to this question, as it depends on the chosen use-case and company's objectives. This guideline will help companies to define their own dynamic performance measurements.

They will enable a new way of analyzing the use-case and will reply to new questions. They enable the company to better prepare for the analysis phase of the use-case, even if this happens outside of the scope of this thesis. Nevertheless, it is best practice to design those performance indicators beforehand. This preparation increases the chances of success for the implementation:

by defining the dynamic performance indicators, the company will be more efficient and know what to look for and how to analyze the graphs provided by thee Process Mining software. According to Expert 4, if designing the performance indicators is not performed beforehand, chances are that the team will be disoriented during the analysis phase and loose trust in the Process Mining technology. Lastly, the dynamic performance indicators enable to make the implementation project measurable, allowing to *sell* it to the rest of the company and in turn building trust around the technology.

How can a company define the dynamic performance measurements? Each company can define their own performance indicators by replying to the following questions:

- What is the value / the benefit expected, and how to measure it? For example, the value expected can be to optimize the working capital. This can be measured by quantifying how often are the invoices paid in advance to suppliers. Other examples are: labor productivity, growing revenue… Here, the research questions established in 4.2.3 can help defining what should be measured.
- What is the worst-case scenario for this use-case? What is the absolute nightmare, that the executors of the process do not want to see happen? Looking at the most important violations, define a set of rules that must be respected. Then, reflect on where/how the use-case can deviate from the defined set of rules and how can it be measured. For example, 4 eyes principle, or a certain order for executing steps…

Piece of advice: during this activity, there should be a clear decision on how to combine the dynamic and the static performance indicators. For example, if a Power BI dashboard displaying static KPIs already exists, is it going to be used next to Process Mining? Will the 2 tools be embedded? Will one replace the other?

The outcome of this activity is the design of both statical and dynamical performance indicators.

4.4.4 Apply knowledge

The goal of this activity to **create the biggest part of the team's knowledge**. The input for this activity is the 10% of knowledge on Process Mining. While executing the activities *Performance indicator* and *Event Log (preliminary)* the knowledge is trained by on-the-job training. The goal is to work on a challenging assignment and familiarize with tasks that will become part of working environment. During this activity, the Process Mining expert takes the role of a mentor and help the peers to gain knowledge via hands-on experience. This type of learning represents 70% of the

total learnings (Lombardo et al., 1996), this also serves as the outcome of this activity. The last 20% of knowledge will be provided in Section 4.5.3: Group reflection.

4.5 Verify

In the verify phase, the outcome of all three pillars is reviewed. Team members should make sure that all the information is complete and accurate. This stage is of primary importance because it enables trust with stakeholders. If, when the project moves forward to the analysis part, after the scope of this thesis, stakeholders discover problems in the process or in the data: trust in the Process Mining technology might be broken. User Interviewees [A, G] emphasis that the more the Information System is customized, the more the Verification stage will require time and efforts.

4.5.1 Process verification

The goal of this activity is to **verify that all elements in the process pillar are complete and accurate**. The input for this activity is the performance indicator.

The project manager must lead their team to verify whether the process mapped in 4.3.5 corresponds to the reality of the execution of the process. It should also be verified whether the performance indicators designed in 4.4.3 can be measured considering the process modeled, and whether the performance indicators are relevant considering the envisioned benefit for the use-case.

The output of this stage is some leads for improving the process pillar.

4.5.2 Data verification

The goal of this activity is to **verify that all elements of the data pillar are complete and accurate**. The input for this activity is the preliminary event log and performance indicators.

The project manager must lead the Information System expert and the Process Mining expert review the preliminary event log to verify whether it is complete and accurate. Especially it must be verified whether the event log designed makes sense compared to both the real execution of the process and the process modeled in 4.3.5. It should also be verified if the attributes of the event logs enable to measure the designed performance indicators from section 4.4.3.

The output of this stage is some leads for improving the data pillar.

4.5.3 Group reflection

The goal of this activity is to **verify the knowledge of the team on Process Mining**. The input for this activity is the 10% and the 70% of knowledge. This guideline recommends that the Process Mining expert organizes feedback sessions, coaching sessions and collective discussions on what went well or less well. Knowledge can also be developed by meeting with companies who already have experience with Process Mining (listening to their best practices and pitfalls)

This type of learning represents 20% of the total learning (Lombardo et al., 1996), this also serves as the outcome of this activity. At the end of this activity, the knowledge on Process Mining is 100%. The knowledge on Process Mining is therefore verified.

4.6 Optimize

In the optimize phase, all activities of the 3 pillars are enhanced. As at this point, the knowledge on all pillars is developed, choices of prioritization can be made. This simplifies the project and enables more chances of success and of trust. Moreover, the data collection is enhanced to enable to increase the quality of event logs.

4.6.1 Use-case scope optimization

The goal of this activity is to **optimize the scope of the chosen use-case**. The input for this activity is the process mapped and the performance indicator. This activity should be performed by the project manager and the Process Mining expert.

Executing a Process Mining implementation is complicated because of the amount and the variety of information. The team can easily get lost between all the activities of the implementation, the data and the steps of the use-case. Some external constraints like time or budget might increase the difficulty. Therefore, it is a good advice to break down the use-case and prioritize. Thanks to the design of the performance indicators, the team should have a clearer overview of the most important part of the process. This guideline suggests starting Process Mining with only a very short part of the process and plan to include more and more steps in the future.

A concrete example of that is: in a order management use-case, one factory may order goods and services to multiple suppliers. The Process Mining implementation would become easier if the scope would be narrowed down to one supplier for the first analysis, and them more suppliers can be added. A prioritized plan for including more suppliers can then be created.

The outcome of this activity is the reduced use-case scope and the prioritized plan to include more parts of the process.

4.6.2 Use-case execution optimization

The goal of this activity is **to ensure that all the targeted data in the event log will be collected**. The input for this activity is the preliminary event log.

In this activity, the project manager and business process expert implement changes in the physical execution of the process. For example, the process execution can be optimized to integrate new data to information system, and therefore be able to mine them too. In other words, this ensures to record all the necessary blueprints and integrate them in the Process Mining analysis.

The outcome of this activity is the optimized process execution.

4.6.3 Data scope optimization

The goal of this activity is **to optimize the scope of the data in the event log**. The input for this activity is the event log and the data mapping. This activity should be executed by the information system expert, the Process Mining expert and the project manager.

Experts 3 warns that duplicating and extracting millions and millions of rows of data slows down the activity on the Information System. In other words, extracting so much data decreases the performance of operations. Therefore, some decisions can be taken to extract relevant data in a way that would not have negative impacts on the rest of the company. Some examples provided by this guideline are:

- Reducing the historicity. For example, looking back at data with a one-year perspective instead of 10 might still be enough for a relevant analysis
- Perhaps, the extraction cannot be done in real time. It could be done once every week or every month.
- The extraction should be done during the night whether than during the day
- The number of data fields to consider in the event log can be reduced to a dozen attributes, instead of hundreds.

The outcome of this activity is the optimized data scope.

4.6.4 Data collection optimization

The goal of this activity is to **ensure that all the targeted data in the event log will be collected**. The input for this activity is the preliminary event log. In this activity, the project manager, business process expert and information system expert implement changes in the information system to automatically collect relevant data. For example, a change in the process can be to automatically

record the timestamp when there is a change in the status of an order. This ensures to record all the necessary blueprints and integrate them in the Process Mining analysis.

The outcome of this activity is the optimized data collection.

4.6.5 Event log optimization

The goal of this activity is to **create an enhanced event log that will lead to a more thorough analysis of the use-case**. The input for this activity is the preliminary event logs and the performance indicator.

To perform this activity, the project manager, the Process Mining expert, and the information system expert work together to enhance the preliminary event log by adding some relevant attributes. Concretely, they link all characteristics of the performance indicator with the corresponding data fields. Those data fields are then added as attributes to the preliminary event log. For example, attributes like cost, resource, employee ID or authorization can be added to the event log if relevant. The outcome of this activity is the complete event log.

4.6.6 Ability

The goal of this activity is to o**ptimize all elements of the organizational pillar and obtain the full support from the sponsor**. The input for this activity is all activities related to change management (awareness, desire, knowledge), and the project governance and vision.

The project manager review and, if needed, bring corrective actions to optimize the ability of the team to implement Process Mining, and the governance choices.

The critical points to enable ability are: availability of required resources, especially time freed-up for the team, a clear vision and strategy, and great quality performance indicators and event logs enabling a future reliable analysis of the use-case.

The final outcome is reported to the sponsor. A sign-off moment is established where the sponsor validates the progress made and agrees to pursue the Process Mining implementation, considering settled resources, strategy, and future objectives. This outcome serves as the final milestone of this guideline.

5 DISCUSSIONS AND LIMITATIONS

5.1 Discussions and contributions

To the researcher's knowledge, the IPMI guideline is the first that has operationally explored the initiation of Process Mining implementation. In this research, several activities were identified that are consistent with existing literature (van der Aalst, 2016, p. 387-420, van Eck et al., 2015, Turetken, 2020). These include the choice of the process, the creation of the project team, identifying business problems or determining the scope of data to extract. However, this research presents unique characteristics that were ignored by previous literature. The three most relevant discoveries are presented from the broadest to the most specific one: the unique scope of this research focusing on the initialization phase, the identification of the 3 pillars of the guideline, and the emphasis on the governance and strategy activities. These findings will contribute to the literature by expending the existing theories. Moreover, the guideline developed in this thesis will contribute to the managerial world by providing operational guidance to companies that would like to initiate a continuous Process Mining implementation.

The first characteristic that demonstrates why this study is important is its unique scope. This research is the first one that focuses on the initiation phase of the implementation. Previous literature studied the implementation of Process Mining at a higher level and broader scope. Typically, research would combine a wide range of activities from the planning stage until the monitoring and sustaining of the Process Mining activity (van der Aalst, 2016, p. 387-420, van Eck et al., 2015, Turetken, 2020). This broader scope also implies that the activities identified previously were less detailed and operational. As a result, companies were limited in their ability to seek guidance from the literature when initiating a Process Mining implementation. Moreover, this pain point was experienced by Philips Domestic Appliances. Hence, to help companies facing their implementation issues, this guideline adopts a focused scope on the initiation phase. Therefore, this study is important because: by following this guideline, companies will benefit from the newly explored focus of the initiation phase, from its operational pieces of advice and from the resources provided. This implicates a successful Process Mining initiation.

The second interesting finding of this research is the identification of the 3 pillars that enable a continuous Process Mining implementation. The 3 identified pillars are: Data, Organization, and Process. Previous Process Mining implementation frameworks ignored those key enablers for a continuous implementation. In the L* life cycle model created by Van der Aalst (2016), most of

the importance was given to technical event log activities, while the 2 other pillars remained unaddressed. In the PM² methodology and the PPMS methodology, some dimensions of the Process pillar and People and Governance pillars are included. For example, activities such as *composing project team*, or *identify and select business processes* are treated. However, those components were treated as isolated activities and were not detailed nor completed by other related activities. As a response to this unaddressed perspective by the literature, this research identified the 3 pillars for initiating a continuous implementation. This finding is important because the pillar articulate the crucial activities in the guideline and help companies oversee all aspects of the implementation project. Moreover, those 3 pillars are interesting because they create synergies between several domains of the Information Management.

The third finding, making this research unique, is the consideration given to the Governance et strategic activities within the Organizational pillar. The governance has been ignored by previous research on Process Mining implementation. In this thesis, governance and strategy were unexpected findings that, ultimately, revealed themselves as essential. Originally, the Organizational pillar did not include the governance part. This is because this topic was never addressed by the existing literature. Eventually, this subject was mentioned several times during the user interviews and its importance was systematically confirmed during the validation interviews. While the governance activities were ignored by existing literature, this matter is of capital importance. The implication is the following: it enables a small-scale experimental pilot to grow into a fully operational and structured competency that can be applied company wide. Including the Organizational Pillar enables companies to technically and socially embed Process Mining as an organization-capability.

Therefore, Process Mining can be used as a tool that is part of bigger vision and strategy for the company.

5.2 Limitations and opportunities for future research

All pieces of research are imperfects. Therefore, number of limitations were found in this study. These limitations are opening a path for future research in the continuous Process Mining implementation field. The 4 identified limitations are:

(1) **Guideline application:** A continuous Process Mining implementation takes time. Considering the time constraint for this research, the guideline has not yet been fully applied in the business world. Therefore, it is not yet validated by its successful usage.

(2) **Activity's subjectivity:** In the guideline, a set of activities has been elaborated. Although great attention was provided to specifying metrics, assessment methods, practical examples and recommendations, this guideline could not offer absolute indications (for example by claiming "*Purchase to Pay is always the best process to initiate the implementation*"). This is because all companies are differently structured and are evolving in different contexts.

(3) **Researcher's bias:** Despite the measures taken to lower the researcher's bias (see 2.2.1.2 Data quality), some participants' interviews might still have been over-interpreted. This is because the researcher worked on the implementation of Process Mining at Philips DA, although, project's members from this company were interviewed. To reduce this limitation, the researcher took all precautions explained the conflictual situation to all interviewees, asking them to voice up all their experiences and behaviors as the researcher is not allowed to include any interpretations in the findings.

(4) **Triangulation:** To increase the credibility of the study, two types of interviews were realized (users and validation interviews). This study could have benefited from different/multiple methods of triangulation, such as another data collection method (for example, a survey). Moreover, this research contains data triangulation limitations, as the different sources of data were collected in a short amount of time.

Therefore, the identified limitations offer an opportunity for new pieces of research. Firstly, in future studies, the guideline should be applied in real-business settings to assess its successfulness in helping organizations to initiate the implementation. Eventually, the artifact may iteratively be enhanced by its outcomes. Secondly, new research should be conducted using the transparently described research steps to confirm this guideline, while not being involved in one of the interviewed company's implementations. Lastly, new research conducted on continuous initiation of Process Mining implementation should combine different data sources, at different time and space.

6 CONCLUSION

This research was originated to help companies initiate a continuous Process Mining implementation. As the Process Mining field is growing both in the business and the academical world, the technology is becoming mature. However, the managerial perspective was not provided with enough attention from the existing literature. From an academical standpoint, this study is the first to consider Process Mining as a capability to be developed, rather than a one-time implementation or analysis. Moreover, its focused scope on the initiation phase enabled to offer a new operational perspective with concrete resources, metrics, and examples. From a managerial standpoint, Philips DA was struggling with initiating the implementation. Designing an operational guideline contributes to facilitating and encouraging Process Mining adoption at Philips DA, as well as in other companies. The aim of this guideline is to act as a handbook to support companies from the business world in their Process Mining implementation journey.

To address the main research question, "*How to initiate a continuous Process Mining implementation?*" an artifact was created using the Design Science methodology. To face a problem in the real world, collecting data in real companies was essential. 10 user interviews were conducted to explore the possibilities and best practices for initiating the implementation. Furthermore, the guideline was approved by 4 Process Mining professionals during validation interviews. These experts also contributed to enrich the artifact with metrics, operational pieces of advice and examples. To consolidate the artifact, the literature was studied at all stages of the research. Eventually, the research question specifies *a continuous* implementation. This was partly enabled by adapting the DIDOV framework. The purpose of this framework is to enable a continuous implementation of new processes or technologies in organizations.

This thesis answers the main research question by designing a guideline that comprehend three pillars for initiating a successful Process Mining implementation: the Data-, Organizational-, and Process- Pillars. The combination and synergies between each pillar are key.

To answer the first sub-question "*which Data measures should an organization take to initiate a continuous Process Mining implementation?*", this research pointed out essential activities. Firstly, the company should assess the quality of their data for each use-case to define whether the company and the use-case are ready for process mining. Then, the data must be localized and identified in the Information System. The Business Process Expert should translate the type of data used by the executors of the process from a business language into a technical language in the Information System to identify the data fields. The data can be extracted, and the event logs can be designed. It is then important to verify that the data corresponds to the process model and

to the reality, and to verify if the event logs enable to address the performance indicators questions. Eventually, the scope of data collected may be optimized (for example, if the extraction is too heavy), or some modifications can be brought to the Information System to collect more adequate data and attributes. This may result in an event log optimization.

To answer the second sub-question *"Which Organizational measures should an organization take to initiate a continuous Process Mining implementation?",* the emphasis is on the governance. Defining a vision and a strategy for Process Mining will serve as an important guide during the implementation. A practical selection method was designed for the team creation, and operational pieces of advice were provided to help companies in defining who are the owners of the decision rights. The main governance choices were listed to help organizations plan their implementation. Moreover, importance was given to change management. This guideline delivers a communication plan to leverage awareness and desire accordingly to stakeholder's own characteristics. It also provides a list of concrete resources for developing knowledge in various manners. Finally, this guideline indicates the importance of a sign-off moment where the ability of team members and of the project is verified considering the resources, knowledge, and strategy. The team and the sponsor should then commit to pursue the implementation.

The last sub-question was "*which Process measures should an organization take to initiate a continuous Process Mining implementation?*". Choosing the right process to start with Process Mining can be challenging. Therefore, this guideline suggests breaking this problem down by defining the company's value-chains, processes, and corresponding use-cases. Each use-case should be linked to at least one benefit. Out of three designed assessment tables to choose the best use-case, one in particular was recommended because it offered the right balance between simplicity and completeness. Nevertheless, all methods are valid. Following the choice of the use-case, its scope must be identified: it was found that a narrower scope is more likely to succeed. Hence, the process knowledge should be exploited to facilitate the design of performance indicators and the data identification and localization. The performance indicators enable to define the characteristics of the use-case an organization aims to track during the analysis phase, and to define a way to measure it. Even if the analysis phase happens after the scope of this guideline, it was found interesting to design those metrics beforehand to be more efficient when performing the analysis, and not feel disoriented which would lead to breaking trust in Process Mining. Finally, it is crucial to verify that the process modeled by hand corresponds to the reality. It is also important to verify that the performance indicators are relevant and that they can indeed be measured based on the scope of the process. Eventually, the scope of the process and its operational execution may have to be optimized to collect appropriate data for the event logs.

REFERENCES

4: Trustworthiness and publishing. European Journal of General Practice, 24(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Armstrong, S. (2022, April 28). 80% Of Projects Fail Because Of 'People' Issues . . . Here Are 6 Things You Can Do To Reduce That Risk – Making Lives Better By Building Better Leaders. Steve Armstrong. Retrieved April 29, 2022, from https://stevenarmstrong.ca/uncategorized/6-things-you-can-do-to-reduce-80-of-the-reasons-your-projects-fail/

Bolt, A., et al. (2015). Scientific workflows for Process Mining: building blocks, scenarios, and implementation. *International Journal on Software Tools for Technology Transfer*, *18*(6), 607–628. https://doi.org/10.1007/s10009-015-0399-5

Bose, R. J. C., Mans, R. S., & van der Aalst, W. M. (2013). Wanna improve process mining results? 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 127–134. https://doi.org/10.1109/cidm.2013.6597227

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Calvanese, D., et al. (2017). OBDA for Log Extraction in Process Mining. *Reasoning Web. Semantic Interoperability on the Web*, 292–345. https://doi.org/10.1007/978-3-319-61033-7_9

Cambridge Advanced Learner's Dictionary & Thesaurus. (2022). Relevance. In Cambridge dictionary. https://dictionary.cambridge.org/dictionary/english/relevance

Celonis. (2022). Ingram Micro + Celonis | Process Mining. Retrieved March 22, 2022, from https://www.celonis.com/customer-success-stories/ingram-micro-celosphere-keynote-process-mining/

Celonis. (2022). Siemens + Celonis | Process Mining. Retrieved March 22, 2022, from https://www.celonis.com/customer-success-stories/siemens-digital-transformation-process-mining/

Creveling, C. M., Slutsky, J., & Antis, D. (2003). Design for Six Sigma in Technology and Product Development (13th, Illustrated ed. ed.). Prentice Hall. https://books.google.nl/books/about/Design_for_Six_Sigma_in_Technology_and_P.html?id=2TT9ngEACAAJ&redir_esc=y

de Oliveira, H., et al. (2020). "Bow-tie" optimal pathway discovery analysis of sepsis hospital admissions using the Hospital Episode Statistics database in England. *JAMIA Open*, *3*(3), 439–448. https://doi.org/10.1093/jamiaopen/ooaa039

Ebert, C., Vizcaino, A., & Manjavacas, A. (2020). IT Governance. IEEE Software, 37(6), 13–20. https://doi.org/10.1109/ms.2020.3016099

Fernandez-Llatas, C. (2020). Interactive Process Mining Challenge. *Interactive Process Mining in Healthcare*, 295–304. https://doi.org/10.1007/978-3-030-53993-1_17

Fortune Business Insights. (2021, November). *Market Research Report*.

Francisco, M. G., Canciglieri Junior, O., & Sant'Anna, N. M. O. (2020). Design for six sigma integrated product development reference model through systematic review. International Journal of Lean Six Sigma, 11(4), 767–795. https://doi.org/10.1108/ijlss-05-2019-0052

Gartner. (2020). *Interview with Prof.dr.ir. Wil van der Aalst* (Research Note G00733123). http://www.padsweb.rwth-aachen.de/wvdaalst/publications/p1107.pdf

George, M. L. (2002). Lean Six Sigma : combining Six Sigma quality with lean speed (10th ed., Vol. 36). McGraw-Hill Companies. https://www.proquest.com/magazines/lean-six-sigma-combining-quality-with-speed/docview/214760423/se-2?accountid=14774

Hand, D. J., Mannila, H., & Smyth, P. (2001). Principles of Data Mining. Amsterdam University Press.

Hiatt, J. M. (2022). ADKAR: A Model for Change in Business, Government and our Community (1st ed.). Prosci Research.

Hilbert, M., et al. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(6025), 60–65. https://doi.org/10.1126/science.1200970

Houghton, C., Casey, D., Shaw, D., & Murphy, K. (2013). Rigour in qualitative case-study research. Nurse Researcher, 20(4), 12–17. https://doi.org/10.7748/nr2013.03.20.4.12.e326

IEEE TASK FORCE ON PROCESS MINING. (2012). Process Mining Manifesto. *Business Process Management Workshops*, 169–194. https://doi.org/10.1007/978-3-642-28108-2_19

Introduction - Process Mining. (2022). Processmining.Org. Retrieved March 22, 2022, from http://www.processmining.org/introduction.html

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. Journal of Advanced Nursing, 72(12), 2954–2965. https://doi.org/10.1111/jan.13031

Kelly S.E. (2010) Qualitative interviewing techniques and styles. In The SAGE Handbook of Qualitative Methods in Health Research (Bourgeault I., Dingwall R. & De Vries R., eds), SAGE, London, pp. 307–327.

Kerremans, M., et al. (2021, November 11). *Market Guide for Process Mining*. Gartner. Retrieved March 14, 2022, from https://www.gartner.com/doc/reprints?id=1-29ALBRGK&ct=220302&st=sb&mkt_tok=OTk1LVhMVC04ODYAAAGDF2QPwTG1

KOpIC-HM8oB6IMxXdZNHqDAG1WIzno1pDbZZROd-v4gYt-LQNHNPqgOMSKU7hEZBgC43jZKz1RBIhacjF0DoweBeQGOKm_EcGlMYqA

Korstjens, I., & Moser, A. (2017). Series: Practical guidance to qualitative research. Part

Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., & Michael, J. (2019). Privacy-preserving process mining. Business & Information Systems Engineering, 61(5), 595-614. https://doi.org/10.1007/s12599-019-00613-3

Mans, R. S., et al. (2015). *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes (SpringerBriefs in Business Process Management)* (2015th ed.). Springer.

Ould, M. A. (2005). Business Process Management: A Rigorous Approach. British Computer Society.

Philips Domestic Appliances. (2021, November 29). *Global program manager digital transformation Consumer Engagement & Operations (Philips Domestic Appliances) (Freelance) - Philips*. Retrieved March 13, 2022, from https://philips.talent-pool.com/projects/global-program-manager-digital-transformation-consumer-engagement-%2526-operations-(philips-domestic-appliances)-(freelance)/15604

Priniski, S. J., et al. (2017). Making Learning Personally Meaningful: A New Framework for Relevance Research. *The Journal of Experimental Education*, *86*(1), 11–29. https://doi.org/10.1080/00220973.2017.1380589

Rebuge, L., et al. (2012). Business process analysis in healthcare environments: A methodology based on Process Mining. *Information Systems*, *37*(2), 99–116. https://doi.org/10.1016/j.is.2011.01.003

Reinkemeyer, L. (2021). The Global State of Operational Excellence: Critical Challenges & Future Trends. Business Transformation & Operational Excellence World Summit & Industry Awards (BTOES). https://insights.btoes.com/download-research-report-2018/19-the-global-state-of-operational-excellence-critical-challenges-future-trends

Robert Wood Johnson Foundation. (2008). RWJF - Qualitative Research Guidelines Project | Semi-structured Interviews | Semi-structured Interviews. Robert Wood Johnson Foundation, Qualitative Research Guideline Project. Retrieved March 17, 2022, from http://www.qualres.org/HomeSemi-3629.html

Rojas, E., et al. (2016). Process Mining in healthcare: A literature review. *Journal of Biomedical Informatics*, *61*, 224–236. https://doi.org/10.1016/j.jbi.2016.04.007

Rubin H.J. & Rubin I.S. (2005) Qualitative Interviewing: The Art of Hearing the Data, 2nd edn. SAGE, Thousand Oaks, CA.

Signavio GmbH. (2022, February 15). *SAP Signavio Process Intelligence - Let Your Data Tell a Story!* Signavio | The Only All-in-One Business Process Software. Retrieved March 14, 2022, from https://www.signavio.com/products/process-intelligence/

Sohail, S. A., et al. (2021). Multilevel Privacy Assurance Evaluation of Healthcare Metadata. *Applied Sciences*, *11*(22), 10686. https://doi.org/10.3390/app112210686

Steinke, I. (2004). Quality criteria in qualitative research. A companion to qualitative research. In U. Flick, E. Von Kardorff, & I. Steinke, A companion to qualitative research (Vol. 21, pp. 184-190). London: Sage.

Tjahjono, B., Ball, P., Vitanov, V., Scorzafave, C., Nogueira, J., Calleja, J., Minguet, M., Narasimha, L., Rivas, A., Srivastava, A., Srivastava, S., & Yadav, A. (2010). Six Sigma: a literature review. International Journal of Lean Six Sigma, 1(3), 216–233. https://doi.org/10.1108/20401461011075017

Turetken, O., et al. (2020). Process Mining for Six Sigma. *Business & Information Systems Engineering*, *63*(3), 277–300. https://doi.org/10.1007/s12599-020-00649-w

Turner V, Gantz JF, Reinsel D, Minton S. The digital universe of opportunities: Rich data and the increasing value of the internet of things. International Data Corporation; 2014.

van der Aalst, W. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes* (2011th ed.). Springer. https://doi.org/10.1007/978-3-642-19345-3

van der Aalst, W. (2012). Process Mining. *ACM Transactions on Management Information Systems*, *3*(2), 1–17. https://doi.org/10.1145/2229156.2229157

van der Aalst, W. (2013). Data Scientist: The Engineer of the Future. BPMcenter.org. http://bpmcenter.org/wp-content/uploads/reports/2013/BPM-13-30.pdf

van der Aalst, W. (2016). *Process Mining: Data Science in Action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4

van Eck, M., Lu, X., Leemans, S. J. J., & van der Aalst, W. (2015). PM2: A Process Mining Project Methodology. Advanced Information Systems Engineering : 27th International Conference, CAiSE 2015, Stockholm, Sweden, June 8–12, 2015., 9097, 297–313. https://doi.org/10.1007/978-3-319-19069-3_19

Verbeek, Buijs, J. C. A. M., van Dongen, B. F., & van der Aalst, W. M. P. (2011). XES, XESame, and ProM 6. In Information Systems Evolution (Vol. 72, pp. 60–75). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17722-4_5

Vipin, P., Sunil, A., & Izhak, P. (2013). A Review of DFSS: Methodology, Implementation and Future Research. International Journal of Innovations in Engineering and Technology

(IJIET), 2(1 February 2013), 369–375. https://www.idc-online.com/technical_references/pdfs/mechanical_engineering/A%20Review%20of%20DFSS.pdf

Weijters, A., et al. (2004). Process Mining: a research agenda. *Computers in Industry*, *53*(3), 231–244. https://doi.org/10.1016/j.compind.2003.10.001

Weill, P., & Ross, J. W. (2004). IT Governance. Reed Business Education.

Wengraf T. (2001) Qualitative Research Interviewing: Biographic Narrative and Semi-structured Methods. SAGE, London.

Wieringa, R. J. (2014). Design Science Methodology for Information Systems and Software Engineering. Design Science Methodology for Information Systems and Software Engineering. https://doi.org/10.1007/978-3-662-43839-8

Zhong, W. (2017, May). *The Integration of Process Mining and Lean Six Sigma*. https://www.utupub.fi/bitstream/handle/10024/146540/WeiZhong.pdf?sequence=1&isAllowed=y

Zomaya, A. Y., & Sakr, S. (2017). Handbook of Big Data Technologies (1st ed. 2017 ed.). Springer.

APPENDIX 1: USER INTERVIEWS SUMMARY

Interview 1: [A] Automation and Industry 4.0 Engineer at Philips Domestic Appliances

Part 1: Experience Sharing

Personal background: [A] is an Automation & Industry 4.0 engineer working at Philips DA for 5 months, their main task is to identify opportunities for improving business processes using new trends in technology and industry 4.0.

Organization situation: small scale projects have been initiated to see what the benefits of process mining could represent and how it works. However, there is presently no process that continuously uses process mining. The discipline is still not known by the larger part of the company. The aim of the process mining pilot(s) was that there is a conviction that process mining can help to improve the execution of the processes, but the implementation was not led by a specific question/ problem.

Experience with Process Mining: [A]'s very first step was to look for available tools to execute process mining, [A] quickly understood that the technology is easily accessible and therefore the challenge does not lay in the technology. At the same time, [A] trained themselves on process mining and on what type of data is required for starting the exercise.

Process pillar:

- Main activities:
    - Choosing a process
    - Running a small case pilot project can help better understand the potential benefits of process mining and can help better scope the chosen process
    - When implementing Process Mining, it might be mandatory to change part of the execution of the process in order to collect and store the data in a more suitable way.
- Challenge encountered: Choosing the right process to start the implementation is a challenge when the number of potential candidates is very large. Therefore, the team should choose a process:
    - that is known in the business to contain problems
    - that has a lot of manual tasks (creating frustration and errors)
    - that could have room for improvement or that could be automated.
    - that involves enthusiastic and self-learner stakeholders.
    - a small and scalable process that occurs frequently.

Data Pillar:

- Main activities:
    - Map what type of data do exists / are available in which information system before trying to extract. [A] also looked for master data to help in the mapping. Recommendation to interview with Business Process Expert for this activity. A benefit of involving Business Process Experts is that they have the competency to translate business requirements in Information System language.
    - Assess the quality of data: some criteria used by [A] are: trustworthiness (whether the data reflects the reality in a reliable way), the ease of availability (whether the access is easy or constrained), automatic generation of data (over manual generation), single source of data (when trying process mining for the first time, it is easier to choose a process with data coming from one source, by opposition to combining multiple sources of data). If there is no other choice than to use data from multiple sources, solutions like Power Automate can be used to combine the sources.

- Challenges: lack of data: in [A]'s experience, at Philips DA, the processes are not sufficiently integrated in the information systems to generate and collect data. It should be possible to monitor the processes before implementing process mining. As the Philips DA company is only 1 year old, data is unmatured and unstructured. Because of the Exuviate program in Philips DA, the current IT tools are being moved to a new IT landscape which prevents easy access to data and mapping of available data (as Information Systems are not available)

Organizational Pillar:

- Main activities:
    - Team selection: in [A]'s experience, the main criteria and functions for choosing a team are:
        - the enthusiasm of the stakeholder and their affinity with digital
        - ability to think in terms of processes
        - ability to communicate with all kinds of stakeholders (users, IT, business, top management...),
        - involving a sponsor
        - involving a process Mining knowledgeable person

        - involving the executor of the process
    - Change management
        - Awareness: It should be made clear from the beginning that the information collected will not be used to blame people but to improve the activities. If there are false pieces of information/rumors running around the project, stakeholders might get cold feet.
        - Desire: Running a very small-scale pilot project could be useful to showcase the benefits of process mining. It can leverage the desire of the stakeholders from other candidate processes for a larger scale implementation in a later stage.
        - Train knowledge on process mining: training should be received by the core team (internal or external depending on resources available). Make a clear distinction with data mining.
    - Building the competency: from the early stages of the implementation, it is important to decide whether the analysis is done in-house (for example, via self-service software) or through a consulting vendor.
- Challenge: Gather commitment from team members and agree on when/where/how to meet. In Philips DA's context, the commitment from the IT team was a challenge as they did not have any time resources available.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

General advice:

- Before thinking to implement process mining, the company should improve the quality of their data generation and storage. Those foundations must be of great maturity as they will define the quality of the results. Moreover, working on good quality data, that are well structured and easily accessible will save some precious time to the implementation team.
- Talk about process mining implementation to companies who are already past this step. The goal is to copy best practices and identify the pitfalls ahead to avoid falling into all of them.
- Importance of defining a strategy for the use of process mining, how will it be used when implemented. This should be considered since the beginning so it can be taken in consideration when implementing, it might impact decisions.

- The difficulty of implementing process mining should be considered, process mining is not just a black box that will immediately tell you what to do to improve processes in the company. Of course, the analysis phase is very important. However, even before that, the preparation for the implementation phase is crucial and is not trivial. The early stages of the process will determine the success or failure of the project.

Interview 2: [B] Regional Operational Excellence Lead at Philips Domestic Appliances

Part 1: Experience Sharing

Personal background:

Working at Philips DA for 1 year and 3months. [B] is responsible for Operational Excellence for 3 manufacturing sites, the factories are in Romania, Italy, and Brazil.

Organization situation:

There is a Process Mining initiative in Philips DA, but it is not yet operationalized. People in the organization are getting aware of Process Mining but multiple questions on how to proceed remain unanswered, preventing a larger scale implementation. This guideline is therefore relevant for Philips DA as it will help answer those questions and, create awareness and desire for the strategic colleagues who could participate in the implementation. For example, as Philips DA is a process-driven organization, Business Process Experts are identified as the "enablers" for initiating the implementation. The main motive for initiating the implementation is to use process mining as a tool to continuously improve the processes.

Experience with Process Mining:

The first few steps identified to initiate the implementation are: to gather knowledge on different potential processes, scope them, map the execution of the process, identify stakeholders, understand the information and data flowing, and when available, assess the quality of the data. This step took around 2 months.

In the implementation, [B] is knowledgeable on the processes: their execution, ownership, enabling continuous monitoring, making them more and more data driven.

Process pillar:

- Main activities:
    - Choose a process: well documented and explained in their execution by the stakeholders, potential for improvement (especially waste elimination), time-consuming and frequent process, a monitored process integrated into an information system (where data is generated, collected, stored, and easily accessible. If there is no data-driven process, the execution of the process might need to be updated to collect data), a process that is suspected to contain a lot of waste (non-lean).
    - exploit as much as possible the knowledge base on the process. The SIPOC tool can be used to help determine the Supplier, the Input, the Process, the Outcome, and the Customer of the potential processes.
- Challenges:

understanding the process: identifying and estimating the gap between the way the process is documented and the users think it is executed. This has been done by interviewing stakeholders on their execution of the process, their needs, and constraints.

Data Pillar:

- Challenge:
    - DA faces a wall regarding the data because there is a lack of integration into the information system. Therefore, for the integrated data, the quality is poor (incomplete, and incorrect because of human interventions). A lesson learned by DA is that the process must be, for a big part, integrated into the information system and should automatically generate data (rather than manually) before thinking about process mining.
    - To get access to the data required to run the process mining exercise, 2 steps are mentioned by [B]:
        - Interviewing the executors of the process to understand the data fields used.
        - Then, request from Business Process Expert (skilled in IT and knowledgeable about the information system) to translate from business terms to information system terms the data fields mentioned by the executors of the process.

Organizational Pillar:

- Main activities
    - Change management:

- Desire: Creating desire for each stakeholder is the first step when gathering a team, it is a key success factor. Considerable importance should be given to dealing with resistance to change. Since process mining will highlight the unhappy flows of the execution of the process, the implementation team should expect resistance to change.
- Team creation: The main skills and roles who should be part of the team are the executors of the process, business process expert of the process, the supplier of the process, the internal customer of the process, and IT ad information system capabilities, sponsor, project manager.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- [B] stresses that preparing for process mining is not easy, it requires a long preparation: the processes should be clarified and documented. The process should then be able to be monitored: users need to let blueprint in the information system when working on the process, or (in the best-case scenario), data is automatically logged and recorded in the information system. Moreover, the best potential process should have stakeholders who understand its execution and its supposed execution, and the data it generates.
- Recommendation: process mining needs stability to be implemented: with the knowledge of the processes, the data generation, the general IT landscape, and the willingness to continuously improve the business processes. As Philips DA disentangled from Royal Philips 1 year ago, this stability is not there yet and prevents the implementation of Process Mining.
- Another recommendation is to assume that the ideal process is executed as it should be, and therefore underestimate or deny the gap that will be found.

Interview 3: [C] Director Business Process Owner – Order to Cash- Order Management at Royal Philips

Part 1: Experience Sharing

<u>Personal background:</u>

[C] is working at Royal Philips since 2008 and is the Director Business Process Owner – Order to Cash – Order management. [C] has responsibilities in both the business side and the IT side of order management.

Company experience with Process Mining:

Back in the 2010s, the company wanted to exploit its data, but process mining was not known yet. Royal Philips developed a delivery performance matrix for customers, where next to their KPIs, they could investigate root cause analysis. This approach was led by specific questions: why our delivery is late, why did we not ship the right quantities… This was considered as a first step toward process mining.

Around 2018, the company wanted to re-design its performance matrix, they ended up adopting process mining as it was a more suitable tool. The implementation was driven by the availability of data. The general aim was to get more insights into the effectiveness of their processes and standardize the way of working, but no specific question triggered the use of process mining. A pilot was released before the operational implementation

Personal Experience with Process Mining:

[C] identifies themselves as the initiator of the project. [C]'s main missions were to supervise the translation of the data requested by the business/process side into an IT language and to supervise the selection of the adequate data field.

Process pillar:

- Main activities:
    - Choosing a process. It was restricted by the content packages that the vendor provided. Also, by the time the vendor was not able to support CRMs information systems, but only ERPs. Therefore, the choice of the process was restricted to a process laying in the ERP.
    - It was decided to run 2 pilots at the same time to maximize the chances of success, with one being prioritized over the other one.
- Challenge encountered: There were long discussions on choosing where to implement process mining
    - Choosing the right process, criteria:
        - Urgency of the situation to improve

        - The company wanted to improve customer satisfaction; therefore, they oriented their approach on a process that involved direct interactions with customers.
        - Started with a process where bottlenecks were suspected but the cause was unclear
        - Choosing a process that should be at the maximum integrated into the ERP.
        - Start by mining only a few steps and slowly add more

Data Pillar:

- Main activities:
    - System changes: the execution of the process did not change in order to collect data, but changes were made in the ERP in order to collect better quality data.
    - When the process is selected: how is it reflected in the ERP? Understand the data that are available, map it, then select which one to extract. To understand the data field required, [C] pulled out one order and followed its path through-out the system.
    - Defining a case ID and following it from the beginning to the end of the chosen process
    - When the process is identified in the ERP, it should be submitted to evaluation before extraction. The goal is to verify that the data are representing the reality and that all the required data field have been identified.
- Challenges encountered: lack of data: in [C]'s experience, in all processes, there were 'black boxes' with steps of the process that happen outside of the software (for example, phone calls). The challenge was to identify those black boxes and overcome the difficulty by finding sufficient data in other sources to still have a reliable and complete outcome of the root causes.

Organizational Pillar:

- Main activities:
    - Team selection: the core team was large. In [C]'s experience, the main criteria and functions for choosing a team are:
        - Implementation initiated by the transformation lead of the strategic program (Blue heart program is the name of the 5 years strategy

adopted by Philips at the time), adopted the role of sponsor of the project
        - vendor/consultant to facilitate implementation (enables faster implementation)
        - IT team is strongly involved in the project, conducted pilot, selecting the process, technical implementation, extraction
        - IT Business specialists: translate the business process language into technical data requirements in the ERP's language
        - Business Process Owners/ Processes team are heavily involved for selecting the process, when the process flow is identified in the ERP, they validate whether it makes sense
        - Management layer of the department executing the process (used only for interviews, not part of the core team)
    - Change management:
        - Knowledge: the vendor provided training on the discipline of process mining and on the usage of their software
        - Ability: even if the team's knowledge was trained, some members did not have the ability to support the project because they were busy with the main transformation program that was ongoing (Blue heart)
    - Defining needed resources (time from the team as part of their job and not on top of it, consultants, budget for the vendor)
    - Building the competency: a vendor company was hired. The competency was built externally for about 2-3 years before building the competency in-house.

- Challenge: The time resource from the key stakeholders is important. In this case, the Business Process Owners and Experts were only partly available which caused challenges in the implementation.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
    - A too large core team involved can reduce the speed of the implementation

  - One should not make the mistake of lowering the involvement of the IT team by only asking for a standard report to be extracted as it is too broad and confusing. It is a better solution to work hand in hand with the IT team and carefully map which data should be extracted.
  - Starting with a clear problem statement / defining pain points helps to prioritize the following steps: it would also help gather support from top management and willingness to help from the executors of the process.
  - Process mining should not be assumed as "something you set up get going" as the set-up part is already very tricky. Even if the help from a vendor with content packages is used, it is not "you just connect the systems and you get going". [C] insists on the point that "it is a huge effort to implement before you get going".
- Pitfalls: At Royal Philips, the initiative of implementing process mining came as an activity in the transformation program "blue heart". However, it turned out to be seen as a side-track of the main changes. Therefore, the implementation was not given the deserved attention, and especially, it was not given sufficient time resources from key stakeholders because they were busy with the main transformation program.

Interview 4: [D] Global Operational Excellence Leader at Philips Domestic Appliances

Part 1: Experience Sharing

Personal background:

[D] is working at Philips DA for 9 months, [D] is the Global Operational Excellence lead. [D] is supporting the manufacturing factories in their continuous improvement journey and also supports the supply chain in new product introduction. There is an important focus on standardization. [D] likes to unlock the potential of people and the potential of data.

Company experience with Process Mining:

Process mining has not yet been implemented at Philips DA. Currently, there is an exploration initiative. Process mining is not part of the Exuviate program (Exuviate is the name of the strategic program with a 4-year horizon, created for the disentanglement of Philips DA from Royal Philips). The initiative came bottom-up from employees who like technologies. The main motives are the

wish to improve processes, efficiency, standardization, and use of available data. But the project is currently lacking a vision or a strategy.

Personal Experience with Process Mining:

[D] identifies as a sponsor of the project pilot. [D] started by finding a process mining knowledgeable person to launch the pilot together. [D]'s main responsibility is to create desire and awareness around the project. Afterward, [D]'s responsibility was to take the final decision of which process to choose.

Process pillar:

- Main activities:
    - Pointing out some potential candidates for the process
    - Assessing out of all potential candidates, which is the most suitable one
    - Mapping the process as the team thinks it is happening, writing down and exploit all possible knowledge on the process
- Challenge: Choosing the right process. Some criteria used to choose the right process are:
    - Important/ core process for the company
    - Known to have problems, potential for improvement
    - A process that involves a team that wants to support, who is enthusiastic about the process
    - Strategic process for the company
    - Frequency of the process allows for maximizing the benefits
    - A process that happens in a similar way in other departments of the organization, to benefit from learned insights and apply them elsewhere

Data Pillar:

- Main activities:

- When the process is selected, map what data the team is looking for in the ERP before opening the ERP (in a business language)
- Understand what data fields are available (in the ERP language) and try to match it with the type of data the business side was looking for
- Define a unique case identifier, verify the availability of a timestamp
- Final choice of data field selection
- Assess if the quality of the selected data is sufficient for a reliable result
- Work hand in hand with ERP experts for the data extraction

- Challenge encountered: In the selected process, there were more steps happening outside of the ERP than inside. Even if it is possible to find a way to combine sources, or to gather more information in individual/ manual excel files, this is not the best situation to start process mining for the very first time. There were a lot of processes that would have room for improvement or involve an enthusiastic team, the challenge was more about what process allows Philips DA to implement process mining with regards to the quality of data

Organizational Pillar:

- Main activities:
  - First step was to assemble a team. Team setup:
    - Project manager/facilitator who organizes day to day projects, connecting people, pushing the project forward
    - Managers of the execution of the process (is the person that feels the pain, can explain the problem, points out that there is room for improvement, able to explain daily interactions with the ERP)
    - Business Process Expert
    - ERP experts (supports in data selection and extraction). It would a great advantage if this person also knows about the process (not only IT skills)
    - Get support from a sponsor who would enable resources, give advice
    - Transversal skills: ability to think in terms of processes, and knowledge of the execution of the process
    - Process mining knowledgeable person (this person can also take the role of the project manager)
    - Change management: the current goal is to create a proof of concept that will be used for awareness-raising in the future when there will be a real process mining implementation
    - Creating Desire, make the project evolve in the direction of people's enthusiasm

- Challenges encountered: Because of the Exuviate program, ERP experts and Business Process Experts did not have the availability of supporting the project. There were lacking the time resource as they were busy with the main transformation program.

This considerably prevented access to data and slowed down the proof of concept. To successfully implement process mining, it should be seen as part of the company's strategy, it should be seen as a priority and be given enough resources. Time should be made for the core team to implement process mining as part of their job, and not on top of all existing missions.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
  - Building the project hand in hand with ERP experts and the management layer of the executors of the process is a good practice as it enables to have access to a wide range of information both operational and technical.
  - Sense of urgency, keeping the agenda prioritized helps because it is very easy to get lost in detail or in the variety of stakeholders
  - Change management is key
  - Gathering commitment from team member's enables the ability of the project.

Interview 5: [E] Head of Supply Chain Operational Excellence at Whirlpool

Part 1: Experience Sharing

Personal background:

[E] is working at Whirlpool for more than 11 years. [E] is the Head of Supply Chain Operational Excellence. [E] is responsible for the supply chain strategy, improvement, and standardization. [E] has a strong focus on Digital Innovation.

Company experience with Process Mining:

Process Mining has been implemented in the full Order Management process across Europe, with the help of a vendor/consultant. The main motive was to improve their most strategic process: the order management one. The process was already identified as being very long and very inefficient. The company had specific questions, they wanted to understand where the inefficiencies were coming from, if the team executing the process was oversized, why were some deliveries late etc. Data was available, and the company wanted to use data more and stop using people's experiences and intuitions for taking decisions. This process was a great occasion to try process

mining as it was fully managed in the ERP. The implementation took around 9 months to implement.

Personal Experience with Process Mining:

[E] was the sponsor and the customer of the process. The first steps were to map the process, identify the pain points and assess whether the vendor could fit with Whirlpool's needs.

Process pillar:

- Main activities:
  - Choice of the process. Some criteria:
    - Start with a process that is known for being inefficient
    - Pick a process that is managed through the ERP. If the data generation is automatic, it is better. If there are only processes with manual data generation in the ERP it is also good. The importance is to have data generated in a single source of Information systems. It is easier for the first-time implementation, especially for following the case ID through each activity.
    - Potential for automation of the inputs. If for the moment the data are manually generated in the ERP, process mining is a good way to quantify it and enhance the process with automation
    - The frequency of the process is key. If the process occurs only 2 times a month, it is not worth going through the effort of process mining implementation.
  - Map the process, understand where the problems in the process are, what are the needs that would enable improvement
  - The process was naturally chosen but it was very long: the company had to prioritize and started with only a subset of the process and continuously added more and more steps in the process.
- Challenge encountered: No important challenge as the choice of the process was natural. There was a small deception when [E] understood that Whirlpool would have to break down the process and add steps slowly instead of doing a one-time very big implementation

Data Pillar:

- Main activities:
  - Mapping all the data available that is linked to the order number: all data was in SAP

- Understand the meaning of each data field. For that: take relevant information provided by the business and find the equivalent data fields in the ERP (match them).
- Discussions between the IT team and the vendor to scope the right data fields (select the relevant ones, exclude the others)
- Challenge encountered: There was no real challenge linked with the data as there were a lot of support from IT. Sometimes it was tricky to understand what each data in the ERP means in a business language. This slowed down the project but never blocked it.

Organizational Pillar

- Main activities:
  - Team setup: employees being enthusiastic about digital and new technologies were particularly encouraged to join the team
    - Sponsors to overcome the roadblocks and communicate on change management
    - Managers of the execution of the process + operational excellence OR Business process experts, that will be able to explain the way they think they execute the process
    - IT support who are knowledgeable on how the ERP is working, able to quickly tell what data are available, where to find the relevant ones, and what each data means in the ERP
    - Business Process Experts are the bridge between IT and Operational Excellence as they have knowledge in both domains
    - Help from someone who has experience with process mining: in this case it was the consultants
  - Change Management:
    - Awareness: the company benefited from having a test environment in their ERP. It enabled them to launch a pilot, understand the benefits of process mining and then communicate. They managed to create awareness before the official broad implementation. Also, a credible sponsor is key for communicating change.
    - Desire: Choosing a process that is frustrating for the employees operating it and communicating on how process mining will improve employees' daily missions help to create the desire to support the project. Scope the project so that it involves people who are enthusiastic about digital.

- Knowledge: training from the vendor on what process mining is, the benefits, and their software
- Ability: Remove some daily tasks from the core team in order to have time to spend on process mining. Process mining implementation is long and requires efforts and investments.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
  - The company should be an analytical company. This means that any statical monitoring of integrated processes is possible, process mining comes as a next step to performing dynamic analysis. Before introducing process mining, most of the processes should happen in an integrated way in the Information System. This is mandatory and this should be the first effort of the company.
  - IT and Business Process Expert support are crucial in this kind of project.
  - The project should be connected to real need.

Interview 6: [F] Business Process Owner - Source to Pay process at Royal Philips

Part 1: Experience Sharing

Personal background:

[F] has been working at Royal Philips for almost 18 years. [F] is a Business Process owner on the Source to Pay process. [F] is designing, testing, and deploying processes.

Company experience with Process Mining:

Process Mining has been implemented at Royal Philips because the company had the goal of being more cost-efficient and more standardized and compliant. The implementation started in 2019 and took a bit less than a year. A pilot has been implemented before the broader implementation. Royal Philips partnered up with a vendor who estimated/ quantified the benefits of cost savings and time savings for Royal Philips.

Personal Experience with Process Mining:

[F]'s role in the implementation was to determine and then confirm which data field should be used. Also, [F] helped in the definition of the chosen process: where it should start and finish and helped in mapping how the execution of the process looks like.

Process pillar:

- Main activities:
  - Choosing the process:
    - A process that has a direct impact on all the operations. (Strategic process)
    - Process already monitored in the ERP (integrated). There were manual creation of data as the data was not automatically generated. But at least, at each step there were data in the ERP.
    - As Royal Philips is a big organization, some of its processes occurs in multiple locations. This enables quick scalability (for example, the order management process happens in each country where the firm is operating)
  - Define/ adjust the beginning and end of the process
  - Identify in detail the execution of the process and its stakeholders. The team should not only look at how the process is supposed to happen, but also understand the specificities of the execution. To do so, it is a good idea to interview with the managers of the execution of the process and with the Business Process Exerts. It is also a good idea to take one order (the order number was the unique case identifier) and to follow its path in the ERP.
  - Started with 2 processes at the same time
  - Started with a small process, added steps afterward
- Challenges Encountered: According to [F], the implementation was slowed by the fact that they did not pay enough attention to the execution of the process in the first place.

Data Pillar:

- Main activities:
  - By studying the process, define which type of data is flowing (interviews with executors of process, BPE, and also follow the path of one order in the ERP)
  - Connect those ideas with the ERP data (this job has been done by more technical Business Process Experts and by consultants)
- Challenge encountered:

- When the process was selected, it was a challenge to validate the data connectors field by field. Meaning that the team knew what activity was performed and had a general idea of which data they should extract. However, it was challenging to understand the technical data field in the ERP, and therefore connect what they were looking for, to the ERP data fields.
- Volume of transactions / of data that needed to be synchronized between the ERP and the process mining software caused a lot of problems. The frequency of the refresh was also challenging. The volume of data was so high that it caused bugs and slowed down the process mining software.

Organizational Pillar

- Main activities:
  - Team selection:
    - Business transformation leader who was steering the program and its team doing the reporting to the CIO
    - Business transformation manager working in the different regions
    - The technical team of the consultants
  - Change management:
- Desire: Triggered desire by comparing the KPIs of each location of the execution of the process. The employees of each location wanted to do "better" than the other locations and therefore supported the project.
- Knowledge: 5 days of training received on the process mining discipline and the software
  - Weekly calls with vendors to align and regular meetings with the Royal Philips' team to support each other.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
  - Exploit the current knowledge to a maximum – the process execution and its stakeholders

- Gather commitment from team involved, make sure that this implementation project is perceived as a top priority initiative.
- The human side is very important: having the right skillset in the team, the right communication structure (both in the team and outside the team for advertising the project), this creates curiosity and awareness.
- The resources needed, especially the time the team can dedicate to the project and the budget should be discussed beforehand
- At the beginning, create performance indicators to quantify the value the company can create (how much money and time will be saved, how many steps will be automated etc…) This enables to monitor the advancement of the project and communicate around it.

- Pitfalls:

- One should not start with process mining if there is no desire from colleagues or no ability because lack of time. The project will run into walls
- One should not start if the organization is not willing to change the execution of processes based on findings. Otherwise, efforts will be put into the implementation while it will only stay "paper"
- Don't set too high expectations at the beginning, it can be disappointing.

Interview 7: [G] Center of Excellence Lead for Insights and Analytics at Royal Philips

Part 1: Experience Sharing

Personal background:

[G] is working at Royal Philips for 11 years. [G] is the Center of Excellence Lead for Insights and Analytics. [G] is leading a team of business information analysts. [G] is also leading the capability for process mining, data visualization, and data quality.

Company experience with Process Mining:

Process Mining has been used at Philips for about 3 years at a company level. The implementation stage lasted 1/ 1.5 years. The main motives were that the company integrated all its IT landscape, therefore process mining was an initiative as part of the program. The initiative came top-down. The main goals were standardization of work and enhancement of the process.

Personal Experience with Process Mining:

[G] was responsible for the delivery of the technical pieces of the process mining (verify that the infrastructure is put in place). [G] first few steps were accompanying the vendor in making a proof of concept.

Process pillar:

- Main activities:
  - Together with the consultant, [G] realized a proof of value on 2 processes to estimate how much they could save if the process mining initiative was successful, and automation of processes could be made. The estimation was too high which afterward caused disappointment.
  - Choice of the process (started with 2 processes)
    - Fist selection: use a core process
    - 2nd round of selection: Balance: complexity vs impact (the value that can be returned when improving the process)
    - 3rd round: assessment: Is data available for this process
  - After the proof of value, a problem was selected
  - The problem was turned into research questions
  - Adjustment of the scope of the process, took a smaller part and added parts continuously
  - Creation of subsequent team

Challenge encountered: overestimation of possible savings

Data Pillar:

- Main activities:
  - The data was available most of the time in the ERP. Data was very complex and large amount. Good quality of data when it came from 1 source, lack of consistency when it came from 2 sources.
  - Understanding in business terms of required data
  - Mapping of the data as presented on the front-end of the Information System
  - Understanding of the connexion between front hand data and how it is stored in the back end
  - Validation part: all the data is complete, no data field is missing
  - Data extraction by IT team

- Decision to be made: refresh time of the data. Impossible to do it in real-time because the company did not have a real-time connection to its data lake. Also, extracting the data and connecting it to the process mining software weekly would slow down the whole Information system of the company. [G] decided to refresh data monthly.

- Challenge encountered:
- Most of the time, processes were managed through the Information System, so data availability was not a problem. The challenge was to understand the data, as the amount of data and data field was enormous.
- Lack of consistency when merging data from 2 sources.

Organizational Pillar

- Main activities:
- top-down support is very important
- culture to build the project externally and once it is safe, build competency in house.
- Decision to adopt a hybrid mode for implementing process mining. Hybrid mode= consultants form vendors + consultants from Philips. Not in house because no knowledge and they wanted the project to be fast and they wanted help because Philips is a bit organization so if they would hire someone with process mining knowledge, the hiring process+ finding someone with the right expertise + onboarding would take too long. They could not implement process mining only by the help of consultants because of the price.
- Team: building the team and building the knowledge took longer than expected. Large team of 20 people in the core team and 30 more supporting (not full time)
  - sponsor (with IT background)
  - project manager (needs to have business intelligence experience)
  - enterprise information management and IT team: knows how the Information system is working and can understand how processes are working (was difficult to find that knowledge)
  - industry consulting (taking care of program management and value realization reporting)
  - business process owners: define how the process is running on paper

- business analyst/managers of the execution of the process: explain how the process is running in the system
- vendors/consultants: knowledge on process mining

o change management
  o knowledge: read reference literature from both the consultant and the books from Wil van der Aalst + 5 days of training on both the process mining discipline and the process mining software for the whole core team, done at the beginning
  o Ability: not enough time was freed up
o Resources:
  o Budget for consultants
  o Time (especially required from ERP experts)
  o Knowledge in the ERP and Process Mining

- Challenge:

o communication between 3 key roles: the ones who understand how the process works, the ones who understand how the ERP works, and the ones who understand how process mining works.
o Time resource, process mining implementation takes a lot of time, while time was not freed up for the project. Desire was easy to create because everyone could understand the potential of process mining, but they did not have the ability (time) to support the initiative as there were multiple transformations going on at the same time in the company.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:

- Start small with experienced people and motivated users, that have time.
- Scope a small amount of data because a larger amount slows down the connection with the process mining software and the user experience is unpleasant

- Pitfalls:

- Choose a project manager who does not have experience with business intelligence
- When there is not enough trust between the different teams or when the IT governance decisions are unclear and therefore there is fights to know who does what

Interview 8: [H] Product Management Officer Supply Chain at Whirlpool

Part 1: Experience Sharing

Personal background:

[H] is at Whirlpool for 5 years and is a Product Management Officer.

Company experience with Process Mining:

Process Mining was implemented between 2019 and 2020. The main motive was that the Order Management process was unnecessarily long and complicated, relying on human work and decisions. The company wanted to understand why the outcome of the process was not as good as desired, and where could they automate tasks. Moreover, the whole company was well integrated into the ERP, which enables process mining.

Personal Experience with Process Mining:

[H] participated in the implementation by providing their knowledge in the execution of the process, and by identifying its stakeholders.

Process pillar:

Main activities:

- Choice of the process: the process was naturally chosen as there was a clear problem existing in the company in the Order Management Process.
  - The choice of the process was not limited by the availability of data, as the whole company was analytical.
  - The process that is known for being frustrating for the employees executing it and has the potential to improve
  - Strategic process for the company as it is the one that influences the most the customer service level.

- Exploit and use all available knowledge on the process: map the execution of the process, especially its pain points, and try to understand the problems in it.

Data Pillar:

- Main activities:
  - Data was available in the information systems for all information systems and was of great quality
  - Map available data linked to the chosen process
  - Understand the meaning of each data
  - Connect the business terms for the data to ERP technical terms

Organizational Pillar

- Main activities:
  - Team setup:
    - Operational Excellence, act as the project manager
    - Executors of the process, help to map the process and gather knowledge
    - Consultants, bring the process mining knowledge, gave training
    - Sponsor, helping with the change management, defining resources
    - Resources used:
    - A lot of time from internal IT support
    - Time freed up for the core team to implement the project
    - The vendor acted as a consultant
    - Budget for the consultants
    - Trainings on process mining

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
  - Everyone in the core team should have the knowledge and an aligned understanding on what process mining is. There should also be some team members that are experts in process mining.

- The company needs to already be analytical before process mining (already have all processes managed in the information systems with data of good quality)
- To be successful, the project needs to have full commitment from all team members, especially from IT support team
- The sponsorship from top management is key. Because a project like process mining will encounter difficulties that only a sponsor can help overcome.

Interview 9: [I] IT Project Manager at Heineken

Part 1: Experience Sharing

Personal background:

[H] is working at Heineken for 13 years, mainly as a IT and business process project manager. Currently, [H] is working on a program called Base: the goal is to have a standard ERP platform and standard way of working implemented in all the breweries. [H]'s role is to set a business process management program. [H] has experience with Business Intelligence.

Company experience with Process Mining:

Process Mining has been a topic for almost a decade in the company. At the time the company was not ready for it yet. This is because, the processes need some maturity: have insights on KPIs, monitor the process, and there must be a need / a desire/ a problem that leads to the use of process mining while there was no urge for standardization before. Today, the company understands that analysing data in the company is the way to win. As part of its Business Transformation program, the company started the process mining implementation in the order to cash and purchase to pay processes. There is also an initiative to implement process mining in the production order process. Before implementing process mining, there was a 7 year-long initiative of merging all the ERPs into a single one and building a great quality data architecture. At the time, the main motives for implementing process mining are process improvement and trying to standardize. The scope was very broad and there was no specific question, while today there are specific questions leading to the implementation of process mining on a new process.

Personal Experience with Process Mining:

[H] is the project manager, [H] is the one who is accountable for making the project happen. The main tasks are planning, resources, budget, and scope. At Heineken the first steps were: some

colleagues heard about process mining during a conference, championed and wanted to try it at Heineken.

Process pillar:

- Main activities:
  - Runed project pilot
  - Choice of the process:
    - A process that is supposed to be easy and standardized, such as procurement
    - Take a process that runs in one information system
    - A process that has sequential activities (following each other) rather than independent activities.
    - Scalable process, add steps continuously. If you try to make it too perfect from the beginning, you will never make it live.
    - Data availability
  - Map the process
  - When the process is mapped, ask oneself "what can we show" what are the performance indicators: so, the situation at the beginning of the analysis can be compared to after the enhancement of the process.

Data Pillar:

- Main activities:
  - Standardization of the main ERP across the whole organization, generating high-quality and standardized data (mandatory step before process mining)
  - Because each Information system has a different purpose, it is not possible to use only one information system in the company: Heineken is implementing reporting tables in their data lake that merge different backends together to visualize all items (ex: orders, invoices, etc) running on all information systems. This will enable Heineken to start process mining on processes that involve more than one information system. This is a great and structured way of working.
  - Together with Business process experts: data mapping and localization in the ERP
  - Extract data and connect process mining software to the test environment of the ERP
- Challenge encountered: the complexity of data caused by the enormous amount of it. The amount of data to extract is a challenge because data must be pulled all throughout

the company's network which can degrade the performance of other reports and the overall data structure. Therefore, it is only possible to load for a maximum of 24 months. You need to be selective in the data you want to extract. An alternative is to increase the capability of the data lake but for a large company like Heineken, it costs about 200 000 euros.

Organizational Pillar

- Main activities:
  - Support from top management came easily as they wanted to find a solution to measure compliance. Then [H] proposed process mining, which was accepted.
  - Weekly touchpoints to update the project plan
  - Set ambitions
  - Team setup:
    - Sponsor
    - Project manager
    - Product owner (responsible for all process mining in Heineken)
    - Delivery team (technical project manager + data architect: connecting to the database, building the data model) who is supported by a consultancy firm in the backend for the most technical and difficult questions
    - During the pilot: the managers executing the process are not yet informed of the initiative, only the business process experts who are providing content information to map the process and explain where to find the right information in the ERP (is familiar with both the ERP and the database structure)
  - Change management:
    - Awareness and desire are easy. In this kind of technological project, the benefits are easy to see even when not quantified.
    - Ability: having time freed up for process mining implementation is still a difficulty, even for the core team. They try to free up time themselves as they know the outcome will benefit them, but for the moment process mining is on top of their daily job, therefore the process is slowed down.

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
  - Think about the governance model. What kind of information process mining is going to bring, and who is going to act on that information? The usage of process mining is complex and therefore involves the right training for users… decisions to be made on how to use process mining in combination with existing tools / how to integrate process mining tools in the current IT landscape, without using so many tools for answering closely related questions? Who makes budget decisions? How many licenses do we open, providing the cost of training and licenses? Are there cheaper alternative solutions, who will assess the opportunity/ challenges of cheaper alternatives? Do we use consultants or implement in-house? Do we use dedicated software or do we repurpose and re-use one we already have? Who provides insights based on the analysis, is it a centralized analytics team, or is it each end-user in each factory/ business unit/department? [H] point of view: the role of the project manager is to show all possible options, ultimately It is the project sponsor who makes the final choice. This type of decision should be opened from the pilot part, this is because those decisions are taking time. Is process mining the right tool for the questions asked/ the problem?
  - A standard ERP and standard data are mandatory before process mining
  - The ideal steps are
    - Is process mining the right tool? Define the to be situation
    - Are we ready for it (ERP and data quality)
    - Technical questions on data extraction, connection, analysis …

Interview 10: [J] Operational Excellence Supply Chain Manager

Part 1: Experience Sharing

Personal background:

[J] worked at Whirlpool for 2 years, in the supply chain operational excellence. [J] focused on the forecasting, order management, and Sales & Operations processes.

Company experience with Process Mining:

Process mining was a discussion subject for multiple months before [J] joined the company but the implementation did not start yet. [J] worked on the implementation but when [J] left the company, there were issues to operationalize the project before the governance (especially the strategy to operationalize the project) did not receive sufficient attention during the implementation phase. The main goal for implementing process mining was general cost savings, productivity, and transparency of the process.

Personal Experience with Process Mining:

[J] mainly participated in the project by helping to design the performance indicators.

Process pillar:

- Main activities:
  - Choice of the process
    - The process was monitored through the ERP
    - A business problem was already known in the company: the chosen process was supposed to be executed in a simple way, but in reality, it was a very long process, with a lot of manual inputs, and too many employees working on the process
    - It is a scalable process because it happens in a similar way across different markets. It is possible to run the process mining on the same process but different markets and then compare them in an objective way: "*if the chosen process on market A can perform this way, why market B cannot?*"
  - Mapping of the process and confirmation from the operational executors of the process. Confirmation on both the physical execution of the process, and how it was executed in the ERP for each field.
  - Design of performance indicators: for example [J] designed a matrix that would help track how many orders were fully automatically generated through EDI (electronic data interexchange)
  - Create planning for the process mining implementation: 3 waves
    - 1st wave: monitor and increase the number of orders fully automatically generated
    - 2nd wave: root cause analysis in the chosen process (changes in orders, enhancement of invoicing and logistics process)

- 3rd wave: internal employees build the process mining competency in house and take over on external consultants

Data Pillar:

- Main activities:
- Preliminary check that the quality of data collected is good, and that there is sufficient master data
- Mapping of all the data. Interviews with the executors of the process or business process experts to understand and map all the flow of data in the digital execution of the process
- Validation of understanding of data
- Optimization/ changes in the way data was collected because some relevant data fields were not collected and stored.
- The implementation team had to exclude the United Kingdom market because they were not using the same version of SAP which would have made the data connection too complicated.
- Challenge encountered: Whirlpool chose a process that was monitored through SAP. SAP has thousands of tables, so in this case, the data was generated. The challenge was to localize the data from different repositories in the ERP.

Organizational Pillar

- Main activities:
- Team creation:
  - External consultant with process mining knowledge
  - A steering committee with IT senior director, operational excellence senior manager
  - IT / ERP team for data extraction
  - Project manager coming from the business domain
  - People working on the operational project who could explain the process
- Challenge: governance/ business use of the process mining tool was not given enough attention. Also, change management was not given enough importance. Consequently, the designated users of the process mining tool refused to use it.
-

Part 2: Reflect on implementation: let's take a step back and reflect on general success factors and pitfalls and general advice to initiate a successful process mining implementation.

- General advice:
  - Scalability of the process: easy implementation of process mining in another area than the original scope because they could re-use knowledge and chose a scalable process
- Pitfalls:
  - Too much of a top-down project and not involving enough the future user of the tool can lead to resistance in the usage of the process mining. Defining the user and involving them in defining the purpose of process mining is a good advice.

APPENDIX 2: RESEARCH DATA MANAGEMENT PLAN

Research data

Research data refers to all the material with which the analysis and results of the research can be verified and reproduced. It may be, for example, various measurement results, data from surveys or interviews, recordings or videos, notes, software, source codes, biological samples, text samples, or collection data.
In the table below, list all the research data you use in your research. Note that the data may consist of several different types of data, so please remember to list all the different data types. List both digital and physical research data.

Research data type	Contains personal details/information*	I will gather/produce the data myself	Someone else has gathered/produced the data	Other notes
User interviews		x		
Validation interview		x		
Literature review			x	

* Personal details/information are all information based on which a person can be identified directly or indirectly, for example by connecting a specific piece of data to another, which makes identification possible. For more information about what data is considered personal go to the Office of the Finnish Data Protection Ombudsman's website

Processing personal data in research

If your data contains personal details/information, you are obliged to comply with the EU's General Data Protection Regulation (GDPR) and the Finnish Data Protection Act. For data that contains personal details, you must prepare a Data Protection Notice for your research participants and determine who is the controller for the research data.

I will prepare a Data Protection Notice** and give it to the research participants before collecting data ☐

The controller** for the personal details is the student themself ☐ the university ☐

My data does not contain any personal data ☒

** More information at the university's intranet page, Data Protection Guideline for Thesis Research

Permissions and rights related to the use of data

Find out what permissions and rights are involved in the use of the data. Consult your thesis supervisor, if necessary. Describe the use permissions and rights for each data type. You can add more data types to the list, if necessary.

Self-collected data

You may need separate permissions to use the data you collect or produce, both in research and in publishing the results. If you are archiving your data, remember to ask the research participants for the necessary permissions for archiving and further use of the data. Also, find out if the repository/archive you have selected requires written permissions from the participants.

Necessary permissions and how they are acquired

User Interview: Asked consent too interviewees to store audio recording of interview on researcher's personal laptop for 5 years (no personal information, anonymized), recording will be deleted after

Validation Interview 2: Asked consent too interviewees to store audio recording of interview on researcher's personal laptop for 5 years (no personal information, anonymized), recording will be deleted after

Data collected by someone else

Do you have the necessary permissions to use the data in your research and to publish the results? Are there copyright or licencing issues involved in the use of the data? Note, for example, that you may need permission to use the images or graphs you have found in publications.

Rights and licences related to the data

Literature review: data collected through literature review is publicly available

Storing the data during the research process

Where will you store your data during the research process?

In the university's network drive ☐

In the university-provided Seafile Cloud Service ☐

Other location, please specify: ☒ On the researcher's personal laptop, and a copy of it on the researcher's external hard drive

The university's data storage services will take care of data security and backup files automatically. If you choose to store your data somewhere other than in the services provided by the university, please specify how you will ensure data security and file backups. Remember to make sure you know every time where you are saving the edited/modified data.

If you are using a smartphone to record anything, please check in advance where the audio or video will be saved. If you are using commercial cloud services (iCloud, Dropbox, Google Drive, etc.) and your data contains personal data, make sure the information you provide in the Data Protection Notice about data migration matches your device settings. The use of commercial cloud services means the data will be transferred to third countries outside the EU.

Documenting the data and metadata

How would you describe your research data so that even an outsider or a person unfamiliar with it will understand what the data is? How would you help yourself recall years later what your data consists of?

Data documentation

Can you describe what has happened to your research data during the research process? Data documentation is essential when you try to track any changes made to the data.
To document the data, I will use:

A field/research journal ☐

A separate document where I will record the main points of the data, such as changes made, phases of analysis, and significance of variables ☒

A readme file linked to the data that describes the main points of the data ☐

Other, please specify: ☐

Data arrangement and integrity

How will you keep your data in order and intact, as well as prevent any accidental changes to it?

I will keep the original data files separate from the data I am using in the research process, so that I can always revert back to the original, if need be. ☒

Version control: I will plan before starting the research how I will name the different data versions and I will adhere to the plan consistently. ☒

I recognise the life span of the data from the beginning of the research and am already prepared for situations, where the data can alter unnoticed, for example while recording, transcribing, downloading, or in data conversions from one file format to another, etc. ☒

Metadata

Metadata is a description of you research data. Based on metadata someone unfamiliar with your data will understand what it consists of. Metadata should include, among others, the file name, location, file size, and information about the producer of the data. Will you require metadata?

I will save my data into an archive or a repository that will take care of the metadata for me. ☐

I will have to create the metadata myself, because the archive/repository where I am uploading the data requires it. ☐

I will not store my data into a public archive/repository, and therefore I will not need to create any metadata. ☒

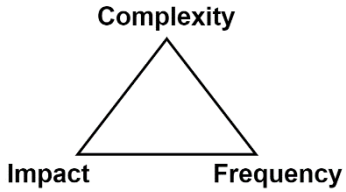
Data after completing the research

You are responsible for the data even after the research process has ended. Make sure you will handle the data according to the agreements you have made. The university recommends a general retention period of five (5) years, with an exception for medical research data, where the retention period is 15 years. Personal data can only be stored as long as it is necessary. If you have agreed to destroy the data after a set time period, you are responsible for destroying the data, even if you no longer are a student at the university. Likewise, when using the university's online storage services, destroying the data is your responsibility.
What happens to your research data, when the research is completed?

I will store all data for 05 years.

If you will store the data, please identify where: on the researcher's personal laptop and a copy of it on the researcher's external hard drive.

APPENDIX 3: USE-CASE CHOICE: METHOD 2, BALANCE BETWEEN FREQUENCY, IMPACT, AND COMPLEXITY (EXPERT: 2)



This second method suggests choosing the use-case by balancing 3 criteria: the complexity, the impact and the frequency (Figure 19). This method has been introduced by Expert 2 during a validation interview.

Figure 19: Use-case choice, method 2

For this method, the Table 13: Prioritization of Potential Use-Cases created in 4.2.3, should be used as an input and will be enhanced in this activity. Each potential use-case should be assessed on a scale from 0/5 to 5/5 with the following criteria: complexity, impact, frequency.

The complexity can be defined by answering the following questions: is data generated at all steps of the process (either manually or automatically)? Is the data easily accessible? Is there documentation available on the process? Are the stakeholders involved in the process ready to free-up time to help in the implementation? How long is the process? Is data laying in a single source (not complex) or multiple (complex)?

The impact can be defined by defined by answering the following questions: How much money or resources or time can be saved if the implementation is successful?

The frequency is defined by the number of times the use-case is happening in a given amount of time.

Unfortunately, Expert 2 mentioned that it is not possible to provide any metric that would answer to the following: *if the chosen use-case happens X times a month / if the revenue saved is superior to X/ if the complexity is lower than X then this use-case is better than another one.* Expert 2 explains *"you can have a process with 3 steps (so a simple process) that happens 1 million times, but its improvement will save you a few minutes and a few thousand euros, or you can have a complex process of 25 steps that happens only 100 times but that will have a bigger impact on the resources saved by saving days and millions of euros".* Expert 2 exemplifies: *"it all depends on what is important for the company, for example, a company like ASML wants to be fast and have high quality, but the revenue saved is not the most important criteria. On the other hand, a supermarket will want to have a low cost and fast processes, but the quality will not be the main criteria"*

At the end, in this methodology, the importance is on the balance between the criteria. Optionally, the importance of each criterion can be pondered by the companies using this guideline, according to the characteristics that are important to them.

To complete this activity, and choose one use-case, the following template should be completed by the companies using this guideline (Table 21: template for choosing the use-case, method 2). This table was originated in section 4.2.3. The use-case with the highest grade should be chosen.

Value Chain	**Process**	**Use Case**	**Benefit**	**Complexity**	**Impact**	**Frequency**
Order to Cash	Process 1	Use-Case A	Risk and Control	/5	/5	/5
		Use-Case B	Process improvement	/5	/5	/5
	Process 2	Use-Case C	Automation Opportunity	…	…	…
		Use-Case D	Risk and Control			
		Use-Case P2 E	Digital Transformation			
…	…	…	…			

Table 21: template for choosing the use-case, method 2

The advantage of this method is that it is simple to put into practice and is the quickest of all to perform. The downside is that it might be too high level and does not take important criteria into consideration (for example, scalability).

APPENDIX 4: USE-CASE CHOICE: METHOD 3, THREE ROUNDS ASSESSMENT

This three rounds assessment toll has been created by merging the 10 user interviews recommendations and experiences. 3 rounds of assessment should be performed until there is only one or two processes left (Table 22). The assessment differentiates the mandatory characteristics of the use-case, the preferred ones, and the optional ones. The chosen use-case must absolutely respect the mandatory characteristics, if those conditions are not met, the project will be more likely to fail. The preferred characteristic enables to choose a process that has a lot of potential for the company, while if the optional characteristics are met, the implementation becomes easier. This activity should be performed by the project manager.

Mandatory characteristics of the use-case	Data is generated in the information system at all steps of the process (either manually or automatically inputted)
	Data comes from one single information system
	The process happens frequently (a metric cannot be provided as this criterion should be put into relation with the next one)
	The process is strategic for the activity of the company (the benefit of implementing process mining on a specific use-case supports the company's general vision and strategy)
Preferred characteristics of the use-case	The process is not supposed to be complex. It is supposed to be simple and straightforward. There is room for improvement and simplification. For example, an invoicing or a customer payment process is in general supposed to be more straightforward than a new product development process.
	The process is scalable in both ways: - It is easily possible to focus on a small part of the process and then look at the bigger picture. For example, it is possible to start with a few steps of the process, and then add more steps to capture the end-to-end process. - The process happens in a similar manner in another location. For example, in another factory, in another market… It is therefore possible to re-use most of the knowledge and amortize efforts put into the implementation to have a bigger return on investment.
	The process involves an enthusiastic team of self-learners and who are ready to go the extra-mile for this project.

Optional character-istics	The process involves people who can free-up time for the project. (at least a few days per week for the core team)
	The process is very well documented (in a business process management soft-ware, through documentations on improvement initiatives…).

Table 22: assessment table for the choice of the use-case

The output of this activity is the chosen use-case.