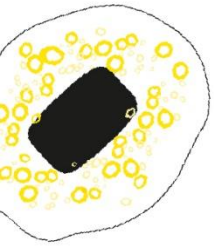
UNIVERSITY OF TWENTE.



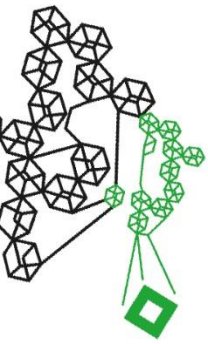
# PROMISE

Process Mining for SMEs: a Methodology


Carlijn Kokkeler

s2148447

M-BIT


UT Supervisor (1): Dr. Faiza Bukhsh

UT Supervisor (2): Dr. Lucas Meertens

Company Supervisor: Samet Kaya

# Preface

This thesis marks the end of my student life and the beginning of my career. I am very grateful to the people who have helped me to deliver this piece of work.

First of all, I want to express my gratitude to my university supervisors Faiza Bukhsh and Lucas Meertens. Thank you for guiding me through this process, making time for me and providing me with valuable feedback. Faiza, thank you for helping me understand the topic of process mining, providing me with the right materials, and offering your time for our useful and reassuring weekly meetings. Lucas, thank you for your academic insights and feedback, for helping me choose my topic, and for answering my questions.

Secondly, I would like to thank eMagiz for providing me with the possibility to conduct my research at the company. Special thanks to my company supervisor Samet Kaya for his support and feedback and getting me into contact with people who could provide valuable input.

Thirdly, thanks to the participants that took the time and effort to help me with my evaluation. Special thanks to Evelyn Tempel for providing me with extensive feedback and helping me to understand important process mining elements.

Lastly, I would like to thank my family and friends for their love and support.

Carlijn Kokkeler

Enschede, June 2023

# Abstract

Process mining is an emerging scientific discipline that focusses on understanding and improving business processes based on data. It relates the fields of process science and data science, with the goal of transforming insights into actions. Process mining can bring benefits to any organization in terms of process efficiency, monetary values, and non-monetary values. However, it appears that practitioners miss guidance when applying process mining. Specifically, research on the use of process mining in small to medium sized organizations is limited. This research bridges this gap by providing PROMISE, a methodology on how to start with process mining in SMEs. It includes a visualization that shows all phases of the methodology, pillars including activities and deliverables for each of the phases, and specific steps that define the required process mining activities. The methodology was developed based on existing literature, refined in one case study, and validated in another case study. Additionally, the methodology was validated through expert and practitioner evaluations. It was concluded that the methodology includes the necessary activities in a process mining project, is clear and easy to understand, and is expected to be useful for practitioners who have a basic level of knowledge, experience, and skills in the field of data and business processes.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | | |
|---|---|---|
| PM | - | Process Mining |
| SME | - | Small to Medium Sized Organization |
| KPI | - | Key Performance Indicator |
| RQ | - | Research Question |
| SLR | - | Systematic Literature Review |
| IC | - | Inclusion Criterium |
| EC | - | Exclusion Criterium |
| QC | - | Quality Criterium |
| DSM | - | Design Science Methodology |
| LG | - | Literature Guideline |
| LF | - | Literature Finding |
| LC | - | Literature Challenge |
| R | - | Requirement |
| G | - | Guideline |
| ILM | - | Integration Lifecycle Management |
| TRQ | - | Treatment Refinement Question |
| TVQ | - | Treatment Validation Question |
| S | - | Step |
| E | - | Expert |
| P | - | Practitioner |
| PU | - | Perceived Usefulness |
| PEU | - | Perceived Ease of Use |
| ITU | - | Intention To Use |

# 1 Introduction

In this chapter, some background information on process mining will be given first. Then, the problem statement will be presented, followed by the research objective. After this, the research questions will be introduced, and, to conclude, the report structure will be outlined.

## 1.1 Background

Process mining (PM) is an emerging scientific discipline [5] that focusses on understanding and improving business processes [41] in a variety of application domains [31]. PM combines the strengths of process modelling and data mining [31][3], and is positioned between process science and data science [5]. It provides methods, techniques, and tools to extract knowledge from event logs [58] to discover, monitor, and improve processes [3]. Apart from improving processes, PM may help to improve process transparency and reduce costs [5]. PM can be applied in any organization [5].

Many tools are available that support PM functionality [3]. Such tools generate a process model [2], such as a BPMN model, or UML diagram, based on event logs [3]. For this, most PM tools and approaches require at least the following three attributes: case ID, activity, and timestamp [5]. The case ID refers to a process instance, activity refers to a task or operation, and timestamp refers to the time of the activity or event [5]. Examples of prominent PM tools are ProM and Disco [31].

Next to process models, PM dashboards with e.g., Key Performance Indicator (KPI) monitoring can be created to gain more insight [29]. Although the creation of a PM dashboard is not specifically mentioned in PM methodologies, it is important for PM activities in organizations [60]. Moreover, the analysis of dashboards is included in PM methodologies such as the L*Life-cycle and PMPM [60].

While much PM research has been performed on technical topics such as the development of algorithms [41], less attention has been paid to understanding how PM is used in practice [58]. Limited research exists on the use of PM in organizations [41][60][56], specifically literature on the use of PM in small to medium sized organizations (SMEs) [26]. SMEs can be distinguished from large organizations mainly by their size and annual turnover. According to the European Commission recommendation 2003/361/EC [62], SMEs are enterprises that have less than 250 employees and an annual turnover of up to EUR 50 million [45]. Results on the use of PM in large organizations cannot automatically be generalized to SMEs [46].

## 1.2 Problem Statement

PM is important for the overall health and well-being of an organization [48]. It has proven its value in many organizations [6] and is expected to grow exponentially in usage [6][30]. This growing interest can be justified by the constant increase in the amount of data that is recorded in information systems [19], as well as the growing complexity of business processes [30]. The increasing significance of PM has also prompted the establishment of the IEEE Task Force on Process Mining [2]. The goal of this task force is to promote the application of PM, provide PM guidelines, and stimulate research on PM [3].

Although PM has proven to be valuable for organizations, it is unclear how organizations should apply PM to generate business values [30][59][12]. In [12], three main business values that PM can bring to organizations are identified: (i) process efficiency, (ii) monetary values, and (iii) non-monetary values. The study provides a model on how these business values can be created but does not provide a guide on how to start with PM in organizations. In particular, knowledge on the application of PM in SMEs is very limited [26].

The study from [30] revealed that process managers miss guidance in their application of PM. Moreover, [12] acknowledges that PM does not only rely on algorithms and techniques, but that human capabilities and goals, as well as organizational factors influence the value that can be generated through PM. It was also found by [60] that existing PM methodologies lack significant PM project elements, and that more research on PM experiences is needed to add to the completeness and relevancy of PM methodologies. Thus, it is clear that research is needed to understand how PM should be used in practice.

## 1.3 Research Objective

The objective of this research is to address the aforementioned gap by proposing a methodology on how to start with PM in SMEs. A methodology is "concerned with revealing in a systematic manner the practices of researchers and the ideas and presuppositions that lie behind those practices" [14], p. 167. Most commonly, a methodology is defined as an overall approach to research, while a method refers to systematic modes, procedures or tools for data collection and analysis [40]. Since the goal of this research is to provide an overall approach on how to start with PM in SMEs, it can be classified as a methodology. The methodology should also include practical guidelines on the use of PM in SMEs.

While some PM methodologies exist, e.g., the $PM^2$ methodology [23] and the L*Lifecycle methodology [4], to the best of our knowledge, no PM methodologies that provide practical guidelines on how to start with PM in SMEs exist yet.

## 1.4 Research Questions

This research follows the Design Science Methodology from [57], as will be elaborated on in Chapter *3 Research Methodology*. A template to define the research problem is proposed in [57], which addresses the problem context, the artefact that is to be designed, the requirements, and the goals of the stakeholders. The template is the following:

*How to <(re)design an artefact>*

*that satisfies <requirements>*

*so that <stakeholder goals can be achieved>*

*in <problem context>?*

Applying this template to the research objective described in Section *1.3 Research Objective*, the main research question of this thesis can be defined as follows:

*How to design a methodology on the use of PM*

*that gives practical guidelines*

*so that PM can be implemented optimally*

*in SMEs?*

To answer this main research question, several sub-questions have been defined:

**RQ1**: What empirical evidence on PM in SMEs is available?

> **RQ1.1**: What methods/techniques/approaches/findings regarding the application of PM in SMEs have been published in the last decade (2012 to 2022)?
>
> **RQ1.2**: What empirical evidence has been produced in the scientific literature about methods/techniques/approaches/findings regarding the application of PM in SMEs that were published after 2012?

**RQ1.3**: What evaluation approaches have been used in empirical studies to validate the proposed methods/techniques/approaches/findings regarding the application of PM in SMEs?

**RQ2**: How does PM in SMEs differ from PM in large organizations?

**RQ3**: How useful are existing PM methodologies when starting with PM in an SME?

**RQ4**: What are requirements for PM methodologies to be effective?

**RQ5**: What elements should a methodology on the use of PM in SMEs address?

**RQ6**: To what extent can the proposed PM methodology be validated against the requirements?

## 1.5 Report Structure

The structure of this report is as follows:

- Chapter *1 Introduction* introduces the background, problem context and research questions.
- Chapter *2 Systematic Literature Review* presents a systematic literature review on the topic of PM in SMEs.
- Chapter *3 Research Methodology* describes the research methodology that will be followed in this study.
- Chapter *4 Problem Investigation* outlines the problem through investigating results from the literature, and defines the stakeholders and goals for this study.
- Chapter *5 Treatment Design* discusses the requirements for the methodology, and provides a first version of the PM methodology.
- Chapter *6 Treatment Refinement* presents the results of applying the methodology in a case study, and involves a refinement of the methodology.
- Chapter *7 Treatment Validation* describes the results of applying the methodology in a second case study, the results of expert and practitioner evaluations, and involves a validation of the methodology.
- Chapter *8 Discussion* provides the key takeaways, contributions and recommendations.
- Chapter *9 Conclusion* gives the main conclusions, limitations and suggestions for future research.

## 1.6 Summary

An overview of the report structure is given in *Figure 1 Report Structure*. The figure shows where the formulated RQs will be answered and what techniques will be used.

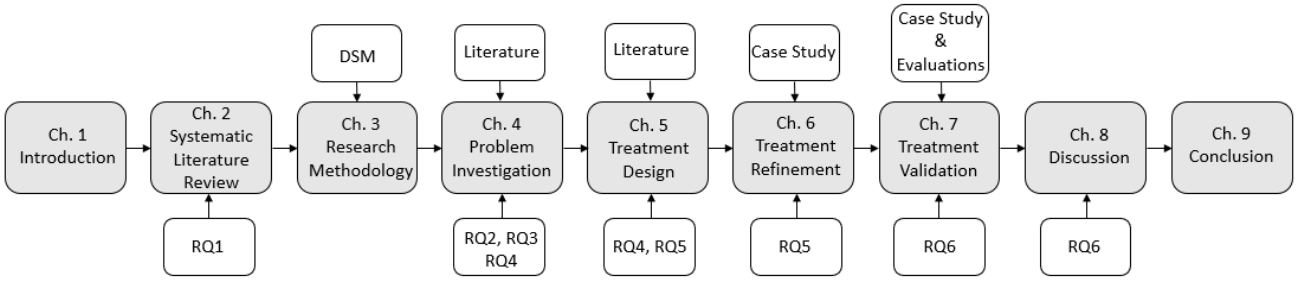



*Figure 1 Report Structure*

# 2 Systematic Literature Review

To fully understand the research gap identified in Section *1.2 Problem Statement*, a systematic literature review (SLR) is executed. The goal of this SLR is to evaluate the empirical evidence that is available on PM in SMEs. To the best of our knowledge, such an SLR on this topic has not been performed before. The SLR takes the approach from [15], and covers studies from 2012 to 2022.

## 2.1 Research Methods

This section provides an overview of the approach that is taken to execute this SLR, based on [15]. First, the research questions are formulated to address the exact purpose of the SLR. After that, the search strategy is defined, including the inclusion and exclusion criteria. Thereafter, a critical appraisal of the collected studies is executed, where quality criteria are established. Lastly, the execution of the search is described.

### 2.1.1 Research Questions

Three Research Questions (RQs) were formulated to analyse the empirical studies on the use of PM in SMEs from 2012 to 2022. These are the following:

**RQ1.1**: What methods/techniques/approaches/findings regarding the application of PM in SMEs have been published in the last decade (2012 to 2022)?

**RQ1.2**: What empirical evidence has been produced in the scientific literature about methods/techniques/approaches/findings regarding the application of PM in SMEs that were published after 2012?

**RQ1.3**: What evaluation approaches have been used in empirical studies to validate the proposed methods/techniques/approaches/findings regarding the application of PM in SMEs?

### 2.1.2 Search Strategy

The search strategy was designed based on the formulated research questions. The first search was executed in Scopus, since this digital library is the most comprehensive and user-friendly database [15].

The search in Scopus was executed on December 14th, 2022, and used the following search string in the article title, abstract, or keywords:

**““process mining” AND (((small OR medium) AND (company OR organization OR enterprise)) OR (SME OR SMEs))”**

The following restrictions to define the boundaries of this study have been applied:

(i) limit by document type (i.e., conference papers and journal articles),

(ii) limit to English language, and

(iii) limit by publication year, starting from 2012.

This search was complemented with a second search in four other digital libraries: ACM, IEEE Xplore, ScienceDirect, and Web of Science to also include materials that could possibly be not within the set of Scopus.

It was not possible to use the full query to search in IEEE Xplore, so the query was split in two:

“” process mining” AND (small OR medium) AND (company OR organization OR enterprise)”, and “”process mining” AND (SME OR SMEs)”, where duplicate papers were removed.

### 2.1.3 Inclusion and Exclusion Criteria

To select the studies that are relevant for our research, inclusion and exclusion criteria were formulated.

The inclusion criteria are:

- **IC1** The paper directly relates to research about PM in SMEs. This means that papers that explicitly propose challenges, experiences, expectations or other findings regarding the use of PM in SMEs will be included. In addition, papers that judge the effectiveness of PM guidelines in SMEs by means of comparative studies, case studies, and experiments, will be included.
- **IC2** The paper addresses the research questions.
- **IC3** The paper is published in a peer-reviewed journal, conference or a workshop.
- **IC4** The paper is in English.
- **IC5** The paper is available for download.

The exclusion criteria are:

- **EC1** The paper is about PM algorithms or other technical details. I.e., papers about PM that do not address its value to organizations/enterprises/companies will be excluded.
- **EC2** The paper is about PM and its application in large organizations/enterprises/companies.
- **EC3** The paper is not peer-reviewed.

These criteria were applied while reading the abstracts of the collected studies.

### 2.1.4 Critical Appraisal of Collected Studies

To ensure that only studies of good quality are used to answer the three RQs, the quality of the found studies was assessed. To assess the quality of each study, the following quality assessment questions with respect to the RQs are formulated and applied on each individual study. (For detailed quality scores see Appendix *B Quality Scores*).

For **RQ1.1**,

- **QC1** Does the paper propose a new method/approach/technique/finding regarding the application of PM in SMEs clearly?
- **QC2** Is the proposed method/approach/technique based on previously defined methods, or are the findings based on previously presented results?
- **QC3** Is the proposed method/approach/technique/finding empirically evaluated or validated (e.g., by using a realistic example, a case study in a real-life setting, an experiment or another empirical research method?)

For **RQ1.2** & **RQ1.3**,

- **QC4** Does the concluded result in the selected empirical paper match the purpose of the empirical study presented in that paper, w.r.t RQ1.2?
- **QC5** Does the empirical study explicitly state its evaluation method w.r.t RQ1.2 and RQ1.3?

- **QC6** Has the method/approach/technique/finding been supported by a specific validation method w.r.t RQ1.2 and RQ1.3?
- **QC7** Is the purpose of the empirical study clearly defined w.r.t RQ1.2?

These questions have been formulated based on [15]. The scoring of the quality assessment questions is applied similarly to [15]. For the critical appraisal of each study, the ordinal range of ratings from 0 to 4 is used, where each of the quality assessment questions holds 1 point. The scoring is independent, meaning that e.g., if a paper scores 2 for RQ1.1 and does not score any point for RQ1.2&RQ1.3, it is included. All quality scores can be found in Appendix *B Quality Scores*.

### 2.1.5 Execution of the Data Extraction Process and Synthesis Strategy

This section shows the study selection process and addresses the need to include additional studies by deriving references. *Figure 2 Study Selection Process* gives a detailed overview of the research process that was used to gradually exclude studies.

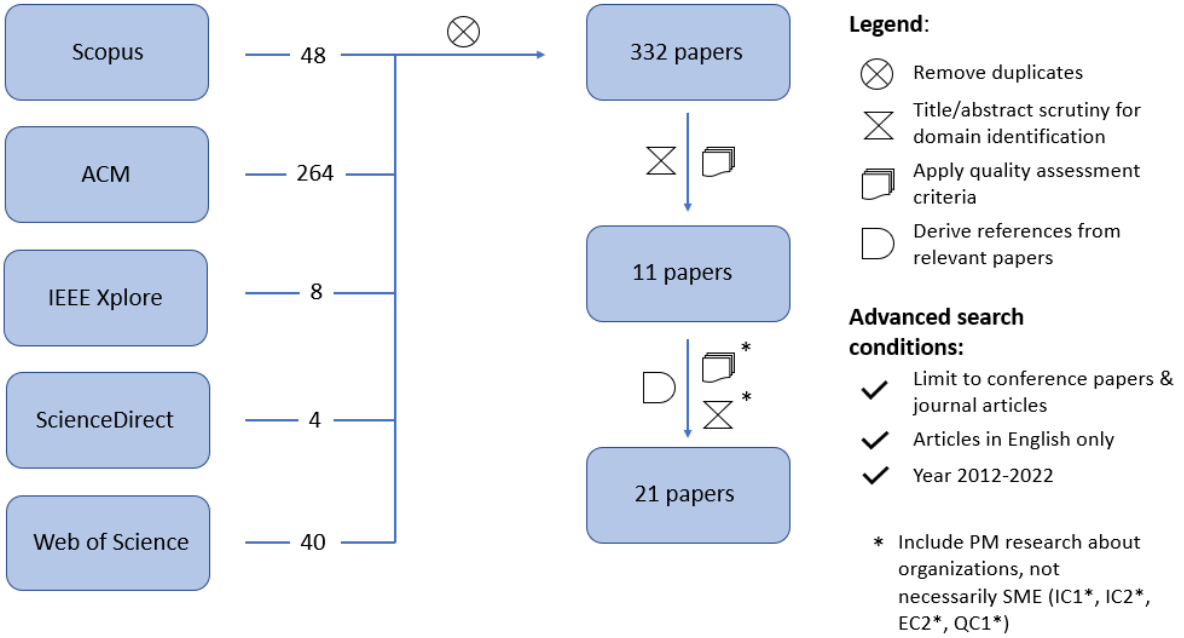



*Figure 2 Study Selection Process*

In total, 11 studies, obtained from the search through the digital libraries, passed the quality assessment procedure by attaining at least 2 points. For this, the abstracts of all the 332 papers were read to apply the inclusion and exclusion criteria and assign them a point. In case the abstract provided insufficient information to assess the quality, the full text was skimmed for a better assessment.

A literature search about the use of PM in SMEs had also been executed as part of the research from [26], where it was found that solely two recently conducted case studies regarding PM in SMEs are available. One of these case studies was already included, namely [54], the other study did not satisfy IC5.

### 2.1.6 Review of Additional Papers

Since the total number of studies that passed the quality assessment procedure is quite low, it seems useful to also include papers that are not solely focused on SMEs, i.e., papers focused on large organizations, or papers that address organizational aspects. The reason for this, is that such papers might provide information on PM that could also be applicable for

SMEs. The search query used in [26] includes such studies, and since this search contained recent papers (2011-2021), the results from this search are used to include more studies. Thus, instead of extending our search query, references from [26] are used to include papers that are not specifically focussed on SMEs. For the inclusion of these papers, some inclusion and exclusion criteria, as well as quality criteria had to be adapted. The adapted criteria, marked with an asterisk, are the following:

**IC1*** The paper directly relates to research about PM in organizations. This means that papers that explicitly propose challenges, experiences, expectations or other findings regarding the use of PM in organizations will be included. In addition, papers that judge the effectiveness of PM guidelines in organizations by means of comparative studies, case studies, and experiments, will be included.

**IC2*** The paper addresses the research questions, where 'SME' is replaced by 'organizations'.

**EC2*** The paper is about PM and its application in organizations/enterprises/companies.

**QC1*** Does the paper propose a new method/approach/technique/finding regarding the application of PM in organizations clearly?

All other inclusion, exclusion, and quality criteria remain the same.

In total, 19 papers were reviewed. After having applied the quality assessment procedure with IC1*, IC2*, EC2*, and QC1*, the number of papers that remained was 10. The quality scores can be found in Appendix *B Quality Scores*, where the added studies are marked with an asterisk. Since the number of papers that remained after having applied the quality assessment procedure was 10, the total number of selected studies is 11 + 10 = 21.

*Figure 3 Selected Studies with respect to Year of Publication* shows the distribution of the papers per year of publication. It is clear that more research has been done in recent years, especially in 2021. A reason for the number of papers in 2022 being lower than the number of papers in 2021, could be that not all papers from 2022 might have been published online when the search was executed. The search was executed in Dec 2022, and it can take several months before research is published, so it could be the case that more research will be published in 2023. Moreover, most research from 2021 was published towards the end of the year (Aug – Dec).

From the 6 studies that were published in 2021, 3 studies focussed on PM in SMEs, and 3 studies focussed on PM and its application in organizations (not specifically SME). The two studies that were published in 2022 were both focussed on PM in SMEs.

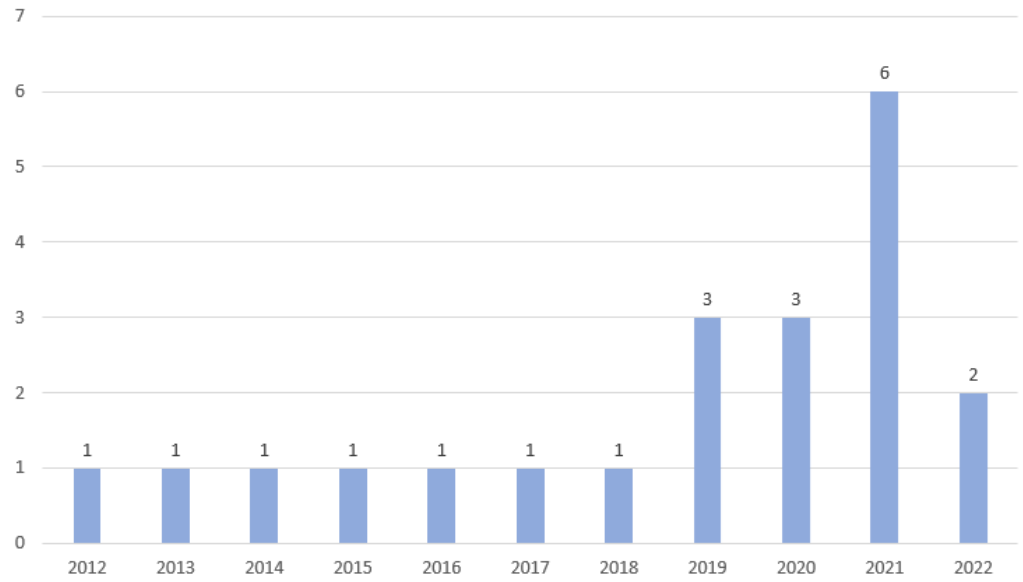



*Figure 3 Selected Studies with respect to Year of Publication*

Once the 21 relevant studies were selected, the following data extraction strategy was used. For RQ1.1, the type of result was identified, e.g., method, framework, or analysis. Studies in which advantages, disadvantages, challenges, or guidelines are defined are categorized as analysis studies. There were no studies that existed of a series of publications. For RQ1.2, the specific methods, their limitations and the evidence that the application of the method produced, were analysed and categorized [15]. For RQ1.3, the validation and evaluation technique of each study was identified [15]. An overview can be found in Appendix *C Overview of Results*.

## 2.2 Results

This section presents the findings for the RQs defined in Section *2.1.1 Research Questions*. Section *2.2.1 Type of Result* explores the type of result that is obtained from each study. Section *2.2.2 Empirical Evidence* identifies and discusses the empirical evidence that each study produced. Section *2.2.3 Validation Techniques* addresses the validation techniques of each study. An overview of the results can be found in Appendix *C Overview of Results*. The demographics of the studies are evaluated in Section *2.2.4 Demographics of the Studies*.

### 2.2.1 Type of Result

For each of the 21 selected studies, the type of result was determined. In total, 9 types of results were found, namely (i) analysis, (ii) implementation, (iii) method, (iv) reflection, (v) framework, (vi) system, (vii) procedure, (viii) methodology, and (ix) system. In *Figure 4 Selected Studies with respect to Type of Results*, the types of results of the studies can be found. The results from the studies are elaborated on below.

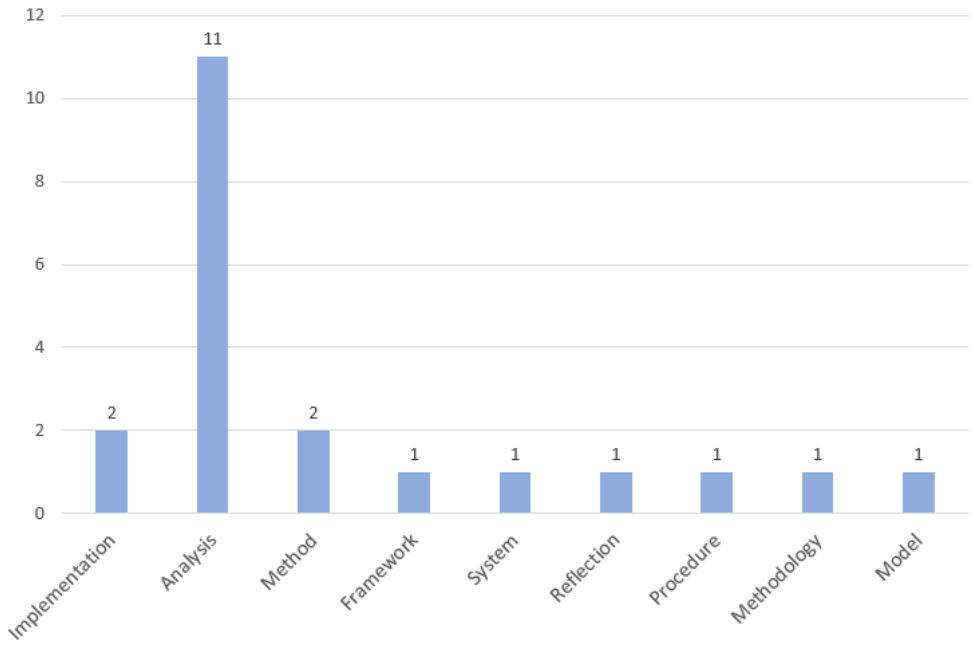



*Figure 4 Selected Studies with respect to Type of Results*

#### 2.2.1.1 Analysis

Among the 21 selected studies, 11 studies were categorized as analysis studies. These studies may address advantages, disadvantages, challenges, guidelines, and other types of analysis results.

From the 11 studies, 4 studies were SME specific. From these studies, the studies [51] and [11] report on observations. More specifically, [51] reports on the observation of a group of small enterprises after the installation of an ERP software. With the use of PM tools, events logged by the ERP are analysed. The study [11] reports on the observation of the interaction between users and an internally developed software, focusing on usage processes. This user behaviour analysis is performed using PM techniques. Both studies report on findings from the analysis, as well as how PM techniques are applied. These results could be used for the development of guidelines in our study.

The studies [54] and [26] present challenges and guidelines on the use of PM. The study [54] focusses on PM in manufacturing companies, while [26] focusses on PM in IT companies. Thus, some findings may be domain specific, in which case the findings will not be included in our study. Moreover, [54] does not report on findings from literature, while [26] presents a list with PM challenges found from the literature, where [54] is included in their review. Furthermore, the study [54] comes up with only 4 main challenges and guidelines, and other than that, they mainly describe the expectations and experiences. On the other hand, [26] comes up with 13 PM challenges in SMEs and 7 guidelines to address them. Thus, [26] seems most useful for our study, where [54] could help to clarify some of the challenges or guidelines.

With respect to the studies that are not necessarily SME specific, the studies [30][55][41] focus on the adoption and use of PM, through focus group studies and interviews, respectively. Thus, they do not apply PM themselves, rather they analyse results. The studies all present challenges regarding PM in organizations. Next to challenges, [30] presents perceived benefits, [55] identifies enablers to overcome some challenges, and [41]

comes up with a list of opportunities. From these studies, [41] presents the most extensive list of opportunities and challenges.

The studies [24][47][52] present an application of PM and report on their findings. Thus, these studies investigate the use of PM through a case study, where they apply PM to an organization themselves. This is useful to understand PM techniques that may be applied. Moreover, [52] clearly identifies success factors and remaining challenges that were found.

Lastly, [25] investigates how business value can be realized through PM, by performing a literature review. The study mainly makes contributions for PM research, no practical findings such as challenges and opportunities for practitioners are presented. Nevertheless, the paper may be useful to understand how organizational practices may contribute to the creation of business value.

#### *2.2.1.2 Implementation*

Two studies [53][37] present an PM implementation approach. These studies are both focussed on SMEs and present a PM technique in combination with another technique. The study [53] investigates how PM can be effectively combined with journal entry tests. They do not present specific guidelines or challenges regarding the application of PM, but their implementation approach can be useful to check which activities need to be performed to apply PM. The study [37] implements an approach using PM and social network analysis. They focus on the selection and allocation of human resources and provides several formulas for this. While the formulas are not relevant for our research, the application of their approach may give some useful insights as to how PM may be applied.

#### *2.2.1.3 Method*

Two studies, both SME focussed, develop a method based on PM techniques, both addressing different aspects. The study [45] proposes RMV, a Recommendation Method for Virtual organizations. This method is supposed to support a collaborative process between SMEs based on PM techniques. The other study [18] presents CEFOP, a method for the Continual Evolution for Organisational Processes. This method describes how to analyse, diagnose, and evolve organisational processes. It takes into consideration the needed continuity of PM. Both studies demonstrate how to apply their method, and they might include some essential PM steps.

#### *2.2.1.4 Reflection*

One study [6] performs a short reflection, including one general guideline for future research. The study mainly provides some insights for SMEs, but these are also applicable for large organizations. They reflect on the current status of the PM discipline and provide an outlook on upcoming developments and challenges.

#### *2.2.1.5 Framework*

The study [13] develops a framework on the application of PM in software process lines, which are families of processes within the software development domain. The method is developed for small software enterprises, and the framework addresses discovery, conformance, and enhancement PM techniques. The framework is still work in progress, so the insights that can be taken from it may be limited.

#### *2.2.1.6 System*

In one study [43], a system is developed, especially for SMEs, to support customer journey management. PM techniques were applied to implement the system in the real world. While the internal functionalities of the system are not likely to be relevant, the design of it may give some information as to how PM can be applied in organizations.

#### 2.2.1.7 Procedure

One study [39] proposes a data-driven procedure to improve productivity in make-to-stock manufacturing companies. PM is used to automatically map and analyse manufacturing processes. The study is not specifically focussed on SMEs, but some aspects may be useful for SMEs as well, such as designing the as-is and to-be process models, which is also discussed in [18].

#### 2.2.1.8 Methodology

The study [20] develops a methodology offering general guidelines and activities that should be followed when applying PM in an organization. The study conducts a comparison of two most prominent PM tools, checking their technical and performance features, and identifies the desired tool for each step. While the study is not necessarily SME focussed, the methodology is likely to be useful for the development of our methodology, because it includes several essential steps to apply PM. For example, data loading has to be achieved in every organization, regardless of their size, and the same holds for the processing of data, which are both described in the methodology.

#### 2.2.1.9 Model

One study [49] develops a model tailored to the characteristics of PM projects. The study identifies and relates PM success factors and measures, which have been evaluated in a multiple case study. The study is not SME specific, so not all success factors and measures may be generalizable to SMEs. Nevertheless, some relevant insights may be gained.

#### 2.2.1.10 Type of Company and/or Dataset

In many studies, a particular company was investigated to evaluate a finding or to report on findings. In some cases, multiple companies were investigated, or a dataset was investigated. It seems useful to check which types of companies or datasets are generally analysed. Below, it is described for each study which company or dataset is investigated. The non-SME specific studies are marked with an asterisk.

- [53] studies a dataset provided by a German medium-sized audit firm.
- [51] investigates a group of 6 small Italian enterprises (some dealers of various products and one manufacturer).
- [45] does not specify the company. It makes use of a dataset that is not publicly available.
- [11] studies Schwer Präzision, a small company located in Italy. The company has 11 collaborators and manufactures complex precision turned parts, milled parts, and technical products.
- [6] does not study a particular company or dataset.
- [13] evaluates their approach within 5 Chilean SSE (small software enterprises).
- [37] does not specify the type of company or dataset. It uses a dataset containing 3880 events and 608 cases or instances.
- [43] makes use of the Google Merchandise Store dataset.
- [18] studies Net Invaders, a juvenile French Start-up, specialized in the development and maintenance of e-commerce sites.
- [54] investigates 2 SMMCs (small and medium sized manufacturing companies).
- [26] studies an SME IT vendor of ERP systems in Germany.
- [47]* studies a naval and ship parts manufacturing company in Korea that is producing steel structures, engine tools, cell guides and peripheral apparatuses for shipbuilding and marine processes. No size information.
- [39]* investigates one of the factories of Geberit AG, a leading manufacturer of sanitary products headquartered in Switzerland.

- [52]* examines three process mining projects performed at the largest rail organization in the Netherlands (NS Stations).
- [20]* uses a dataset that can be found in the 4TU Centre for Research Data repository. The dataset contains manufacturing data extracted from an ERP system.
- [30]* does a focus group study with 22 participants that were part of a workshop organized by Signavio, a commercial provider of BPM software and process mining applications. The 22 participants were representatives from organizations from different industries, such as healthcare and financial services. The profiles of these participants largely cover the tasks that are typically assigned to process managers.
- [41]* conducts a Delphi study with 40 international PM experts from academia and industry.
- [49]* conducts four projects; (i) ASML: large manufacturer of advanced technology systems for semiconductor industry, (ii) IT auditing department of a large German multinational company active in the energy, healthcare and manufacturing industries, (iii) Verbeeten institute: specialized hospital with high expertise in radiotherapy and nuclear care, and (iv) Dutch branch of T-Mobile.
- [55]* studies a large pension fund in the Netherlands, Algemene Pension Groep (APG).
- [25]* does not study a particular company or dataset, rather they execute a literature review on a set of 58 research articles published between 2005 and 2019.
- [24]* studies Algemene Pension Groep (APG), a large provider of services to pension funds in the Netherlands.

It is clear that mainly manufacturing companies have been investigated. From the 21 studies, 7 studies analysed a manufacturing company. A reason for this may be that manufacturing clearly includes a process, namely the establishment of a product through components. Next to manufacturing firms, IT and software companies were investigated most. From the selected studies, 5 studies focussed on such companies. A reason for this may be that it is expected that IT and software companies store much data and have knowledge on how to obtain that data, since their employees are IT experts, which is needed for PM. From the papers it was not clear why certain types of companies were chosen.

The remaining studies focussed on auditing firms, pension firms and a rail organization. In four studies, the type of company or dataset is not specified.

#### *2.2.1.11 Types of Processes*

In most studies, the type of process that is investigated is discussed. The reason for studying a particular process was mainly dependent on the type of company. It seems interesting to see which types of processes are investigated most. Below, the type of process that is investigated in each study, if this was described, is given.

- [53] studies the purchase-to-pay-process.
- [51] analyses the events from the sales, the purchases and the manufacture cycles.
- [45] investigates the partner selection process.
- [11] studies the process of starting the program and creating a new order entry with an associated contract.
- [6] does not study a particular process.
- [11] focusses on usage processes where users interact with an internally developed software.
- [13] investigates the set of predefined software processes that a company follows.
- [37] analyses the purchasing process.

- [43] analyses the customer journey; a set of process-based interactions between customers and company's products or services that take place across various communication channels such as social media, websites, emails, and face-to-face meetings.
- [18] studies the ticket support providing process.
- [54] investigates the electroplating process; a bath for surface treatment of parts has to be refilled after use or time.
- [26] investigates three processes: (i) the consulting request process, (ii) the CIM project lifecycle process, and (iii) the circulation checklist process.
- [47]* analyses the material purchasing process.
- [39]* investigates five sequential processes of a plastic actuator plate: (i) moulding, (ii) assembly 1, (iii) assembly 2. (iv) sorting, and (v) packaging.
- [52]* investigates three processes: (i) locker retention, (ii) service desk, and (iii) wheelset overhaul.
- [20]* analyses the manufacturing process of a product.
- [30]* employs a focus group study, where the participants focus on the following processes: business process improvement, auditing and compliance, digital transformation, and IT operations.
- [41]* does not focus on one particular process.
- [49]* investigates four projects with different processes; (i) testing of wafer scanners before they are delivered to customers, (ii) purchase-to-pay process, from purchase requisitions to outgoing payments, (iii) radiotherapeutic treatment of cancer, and (iv) activation of customer services for existing customers when they initiated a new iPhone subscription.
- [55]* investigates data from several processes that are executed at a large pension funds provider.
- [25]* does not study a particular process, it is a literature review on process mining.
- [24]* investigates three different processes; (i) customer journey: administrative processes such as clients starting retirement, starting a new job, and other life events, (ii) pension-related processes, (iii) specific financial process.

It appears that mainly manufacturing processes were investigated, namely 7, which is logical since mostly manufacturing companies were investigated. Furthermore, mostly purchase-to-pay processes were analysed, namely in 5 studies. Three studies do not specify the type of process being studied.

The identification of the type of process may help to identify potential process to be analysed in our study. Moreover, if an approach is specific to a type of process, the approach may be less useful for the development of our methodology, so it is important to be critical about this.

#### *2.2.1.12 Type of PM Tool*

Several studies specify the PM tool that is used in their research. It seems useful to identify the type of PM tool that is used in each study, so that information about different PM tools can be obtained, which can be used in our methodology. This way, it may be possible to give guidelines regarding the type of tool that should be used. In case the type of PM tool that is used in the study is specified, this is mentioned below.

- [51] makes use of ProM rel. 6.4.
- [11] uses the Disco tool.
- [13] makes use of the ProM tool.
- [26] uses the ProM tool in the version 6.10.

- [47]* uses the Disco tool.
- [39]* uses VSM.
- [20]* uses the ProM 6.8 tool and the Disco 2.2.1. tool.
- [30]* employs a focus group study, where the participants use the following tools: Lana Labs, Fluxicon, Celonis, Process Gold, Stereo LOGIC, and Software AG ARIS.
- [55]* uses the Celonis tool.
- [24]* investigates the Celonis tool.

The following studies make use of a PM tool, but do not mention the specific tool that is being used: [53][18][54][45][52]*. The studies [41]* and [49]* mention some PM tools but do not elaborate on them. All other studies do not make use of a PM tool.

#### *2.2.1.13 Type of Discovery Algorithm*

Some PM tools provide several algorithms to mine process models. To see which types of algorithms are generally used, the papers were checked for the type of PM algorithm used to mine the process models. These results could be used for the development of guidelines regarding the desired PM algorithm(s).

- [53] makes use of the Alpha Miner, Inductive Miner, and Heuristic Miner, where the Heuristic Minder provided the best results.
- [11] uses the Fuzzy Mining algorithm.
- [13] makes use of the v-algorithm; a SPrL discovery algorithm.
- [37] specifies some mechanisms; DoS, role-based, and random-based.
- [43] makes use of four different discovery algorithms: Fuzzy Miner, Alpha Miner, Heuristic Miner, and Inductive Miner, where the Inductive Miner provided the best results, and thereafter the Heuristic Miner.
- [26] uses the algorithms Inductive Miner, ETMd, Heuristic Miner, Fuzzy Miner, and the DFG miner, where it differed per process which algorithm provided the best results.
- [47]* uses the fuzzy miner algorithm.
- [20]* uses the inductive miner, heuristic miner, genetic miner, alpha algorithm, multi-phase miner, and the fuzzy miner.

All remaining studies either do not make use of a discovery algorithm or do not specify the type of algorithm that is used.

#### *2.2.1.14 PM Methodology*

Since a methodology will be developed in this study, it seems useful to check whether PM methodologies are followed in the studies, and if so, which type of PM methodology. This is given below.

- [26] uses the L*-Lifecycle-Model from van der Aalst (2011).
- [39]* mentions possible methodologies but does state which methodology is used to develop their procedure.
- [52]* adheres to the business process lifecycle framework of Weske (2012).
- [20]* mentions the PM^2 methodology but does not state whether it makes use of this methodology.
- [30]* mentions that PM methodologies exist but does not specify them.
- [41]* mentions that PM methodologies exist but does not specify them.

It is clear that not many studies mention a PM methodology. This raises the question as to whether the type of methodology is simply not mentioned, or whether no particular methodology has been followed. It does imply that existing methodologies do not give such

guidance that they are worth mentioning. This emphasizes the need to develop a methodology with clear guidance.

### 2.2.2 Empirical Evidence

For each of the 21 selected studies, the type of empirical evidence was determined. In total, 5 types of results were found, namely (i) case study, (ii) illustrative example, (iii) focus group study, (iv) simulation, and (v) Delphi study. The types of empirical evidence of the studies can be found in *Figure 5 Selected Studies with respect to Type of Empirical Evidence*. Note that one study contained a case study as well as a simulation.

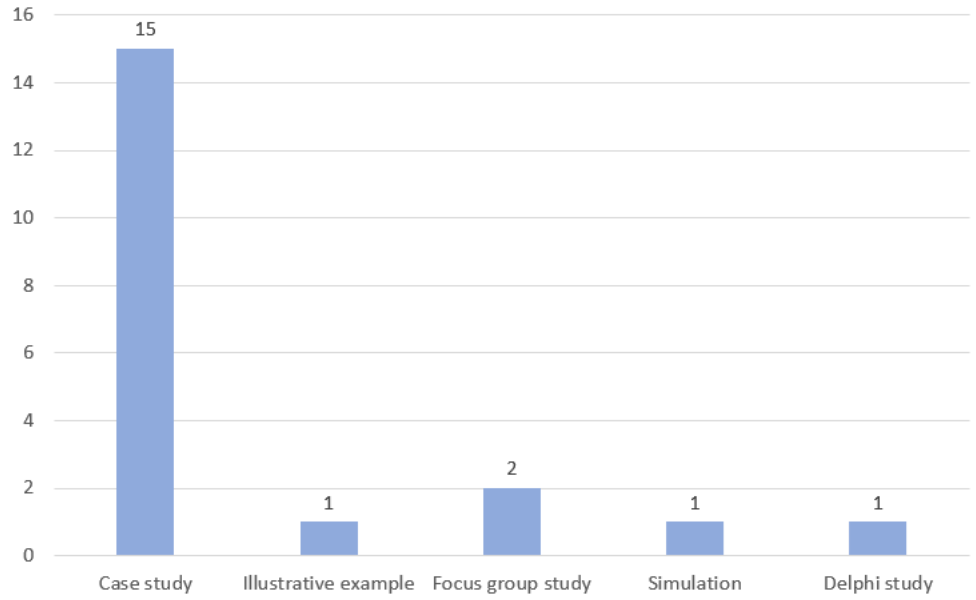



*Figure 5 Selected Studies with respect to Type of Empirical Evidence. Note that one study contained a case study as well as a simulation.*

It is clear that mainly case studies have been conducted. These case studies were often conducted at a particular company. The companies that were investigated are identified in Section *2.2.1.10 Type of Company and/or Dataset*. It seems logical that mainly case studies have been performed, since the search concerned PM in organizations. To demonstrate the use of PM in organizations, case studies are helpful, because they allow for an analysis of a particular PM approach at a company. The same holds for deriving findings on the use of PM in organizations.

Two focus groups studies were conducted. One of the two focus group studies [54] is focussed on SMEs, the other [30] does not specify company sizes. The study [54] reports on expectations and experiences on PM in small to medium sized manufacturing companies. The study [30] explores the adoption, use and management of PM through a focus group study with participants that are representatives from organizations from different industries, such as healthcare and financial services.

The illustrative example [45] and the simulation [37] were obtained from the studies focussing on SMEs. The simulation model was created based on event logs from a case study performed in the study. Thus, note that this study was categorized as both a case study, as well as a simulation. The Delphi study [41] was performed with 40 international experts from academia and industry and consisted of 6 rounds. Such a Delphi studies yields results that are more thoroughly validated and better generalizable as compared to a case

study conducted at one particular company. The study provides an extensive list of opportunities as well as challenges, so these may be taken into account for our study.

Two studies [6][25] do not provide empirical evidence. The study [6] summarizes findings, and [25] performs a literature review.

### 2.2.3 Validation Techniques

Four studies explicitly provided validations for their research. These studies including their validations are given below.

- [37] creates a simulation model based on event logs from their case study and ran the model 1000 times. Then, descriptive statistics were obtained, Welch's t test was used, and multiple comparisons using Games-Howell post hoc test were conducted.
- [39]* proposes a procedure which is afterwards validated in a single case study.
- [49]* adapts Bandara's original modelling success model with success factors specific for PM. They make use of a re-specification phase to confirm the validity of the success measures. In this final part of their study, a case study was applied to validate the findings and, if required, to re-specify the a priori model. This was achieved through a cross-case analysis.
- [41]* validates the findings from their Delphi study using the Fisher's exact test, as well as qualitative comparisons.

It can be concluded that not many studies provide validations for their results, namely 4 out of 19 studies that provide empirical evidence. Thus, it is important for future research that findings on the use of PM in organizations are validated.

### 2.2.4 Demographics of the Studies

Most research has been conducted in several European countries, where most research was done in Germany, as can be seen in *Figure 6 Demographic Trend of Publication (country)*. This research from Germany was done at three different institutes, and from different authors, so there does not seem to be any relation. The two studies from the Netherlands are also done at different universities, and the same holds for the studies from Italy and South Korea. Both studies from Australia are by the same authors, at the same university. Moreover, both studies investigate a case at the Dutch APG (Algemene Pensioen Groep), a large provider of services to pension funds. However, the papers are independent from each other, there is no reference from the latter paper to the former.

From the studies that were executed through a collaboration by different countries, 14 studies were done in European countries, as can be seen in *Figure 7 Demographic Trend of Publication (continent)*. Thus, the amount of research done in Europe is much more as compared to the other continents. From the studies that were done in collaboration across countries, one study was a collaboration across different continents, namely Europe & Australia.

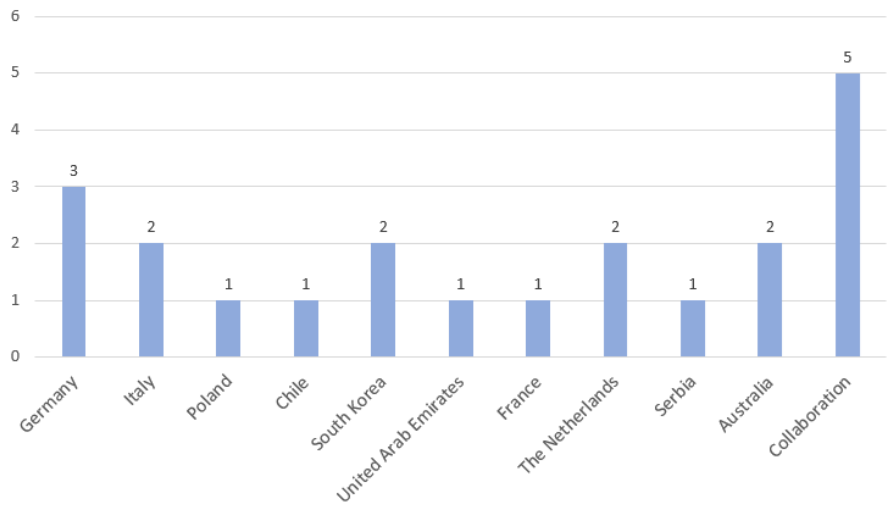



*Figure 6 Demographic Trend of Publication (country)*

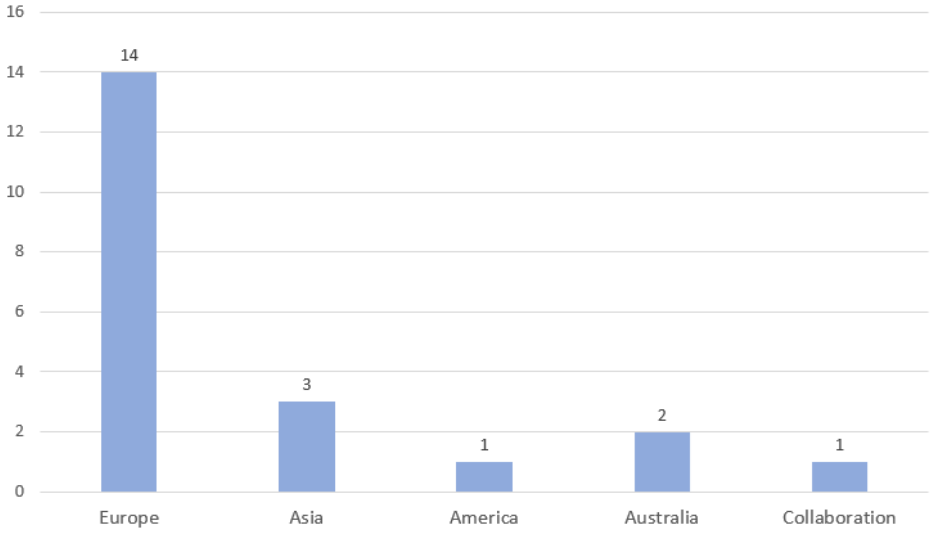



*Figure 7 Demographic Trend of Publication (continent)*

## 2.3 Discussion

Apparent from the literature search is that mainly analyses about PM have been executed. These analyses may include observations, (dis)advantages, challenges, guidelines, etc. From these 11 analysis studies, 4 studies were focussed on SMEs, and the remaining 7 studies on large organizations or other organizational contexts. All analysis studies address different topics, there are no continuations of other papers.

After analysis studies, the most frequent studies are studies in which an implementation is achieved, and studies where a method is developed. In both cases, there were 2 papers, both focussing on SMEs. The papers address different implementations and methods, there is no dependency across papers. This holds for all papers; no paper is a continuation of another paper. The other types of results are method, reflection, framework, system, procedure, methodology and model, which all comprise of one paper. An overview is given in *Figure 4 Selected Studies with respect to Type of Results*. Although guidelines are given regarding the use of PM in organizations, none of the studies explicitly provides steps on how to start with PM in SMEs.

With respect to the type of company and/or dataset, two studies investigate PM at the same company. The company that is investigated in both studies is the Algemene Pension Groep (APG), and the studies are executed by the same authors. However, the studies are independent, the latter paper is not a continuation of the former, and there is no reference from the latter paper to the former. Four studies do not study a particular company or dataset. Two of these studies retrieve information from experts as part of their empirical study, these are a Delphi study and a focus group study. The other two studies are a reflection and a literature review and do not provide empirical evidence.

The type of process that is investigated differs per study. Some processes that reoccurred a couple of times were purchasing processes, manufacturing processes, and customer journey processes. However, there was no relation between those papers. Three studies do not study a particular process. One of these is a reflection, which gives a more general guideline on PM in organizations. Moreover, [41] provides a holistic view of opportunities and challenges, where organizations as a whole are taken into account. Lastly, [25] conducts a literature review on PM, where the value of PM for the complete business is addressed.

Six studies mentioned PM methodologies, where only one paper investigated projects in which the mentioned methodology was used. This is the only paper of the six that is focussed on SMEs. Four studies mention PM methodologies but do not continue to use them, or at least do not specify this. One study follows a framework, which can be categorized as a PM methodology according to the paper [52].

Regarding the type of PM tool that is used in the studies, most studies make use of the ProM tool. After ProM, the most frequently used tool is Celonis, followed by Disco. Some studies make use of several tools, and one study employs a focus group study where the participants used different types of PM tools. The type of discovery algorithm differed per study as well. One study [11] mentioned that the Disco tool makes use of the Fuzzy algorithm. Studies which made use of the ProM tool used several algorithms, where the Alpha Miner, Inductive Miner and Heuristic Miner were the most frequently used. The studies in which Celonis was used did not address the type of algorithm. One study made use of an algorithm specific to their study, and one study specified mechanisms instead of particular discovery algorithms.

With respect to the empirical evidence, two papers do not provide empirical evidence. From the studies that provide empirical evidence, most conducted a case study, namely 15 out of 19. These case studies were either performed at a specific company, multiple companies, or made use of a publicly available dataset. Two studies conducted a focus group study, where one investigated two SMEs, and the other obtained results from experts that operate in different industries. The other types of empirical evidence were a simulation, a Delphi study, and an illustrative example. One study conducted a case study as well as a simulation and was therefore categorized in both types of empirical evidence. An overview of the types of empirical evidence can be found in *Figure 5 Selected Studies with respect to Type of Empirical Evidence*. Note that one study contained a case study as well as a simulation.

Four studies explicitly mention their validation technique. One of these is focussed on SMEs and uses a simulation to validate the results from their case study. Two studies validated their findings with a cross-case analysis, and the Fisher's exact test, respectively. The other study proposed a specified procedure which was validated in their case study.

With reference to the demographics of the studies per country, most studies were conducted in Europe, namely 14 out of the 21. Three studies were from Asia, where two studies were conducted in South Korea, both from different universities. Two studies were done in

Australia, by the same authors, from the same university. These studies were not related, however. Five studies were collaborations between countries, where four of them consisted of solely European countries. The other study was a collaboration between European and Australian authors.

## 2.4 Limitations and Threats to Validity

The main threat to the validity of this SLR is that there may be more relevant literature that has not been included. Reasons for this could be that not all existing relevant articles were included in the searched digital libraries, and that the search term used might not have covered all relevant material. This threat was mitigated by searching through 5 large digital libraries, and using synonyms, such as organization, company, and enterprises, as well as Boolean conjunctions. Moreover, an elaborate search about PM in organizations was conducted by [26]. The references from this paper were checked, and some additional papers about PM in large organizations were included.

Another threat to the validity of this research is that the found studies might not address the research questions. To mitigate this threat, quality criteria were established, and the studies were checked according to these criteria. Only the papers that scored at least 2 points on these criteria were included.

## 2.5 Summary

This SLR provides an understanding of the methods, techniques, approaches, and findings regarding the application of PM in SMEs that have been published in the last decade (2012 to 2022). It gives an overview of the empirical evidence found in these studies, as well as the evaluation and validation approaches used. This research was performed following the techniques undertaken in the SLR from [15]. Since solely 11 studies about the application of PM in SMEs were found, some papers about the application of PM in large organizations, or organizations from which no size information was available, were included. Below, the most important findings for each of the research questions are summarized:

**RQ1.1**: *What methods/techniques/approaches/findings regarding the application of PM in SMEs have been published in the last decade (2012 to 2022)?* It was found that 4 analyses, 2 implementations, 2 methods, a reflection, a framework, and a system were produced about the use of PM in SMEs. Some of the analyses provide challenges and guidelines, others describe more general findings. The implementations and methods were based on PM techniques, mostly in combination with another technique. No method was presented solely on the use of PM in SMEs.

With respect to larger organizations and other organizational aspects, 7 analyses were performed, one model was produced, one methodology, and one procedure. The analyses provide many challenges and guidelines, but not all results may be generalized to SMEs. The model, methodology and procedure all addressed some more specific steps and guidelines in comparison to the analyses.

Most research about PM in SMEs was published in recent years; 2020, 2021, and 2022.

**RQ1.2**: *What empirical evidence has been produced in the scientific literature about methods/techniques/approaches/findings regarding the application of PM in SMEs that were published after 2012?* From the 21 selected studies, 19 studies provided empirical evidence. From these studies, by far most empirical evidence was achieved through case studies, namely 15. Other empirical evidence was produced by focus group studies, a Delphi study and an illustrative example. Some papers evaluated PM challenges that were found in

several other studies, but no findings were explicitly evaluated. Therefore, it might be useful to evaluate the findings from some studies further.

**RQ1.3**: *What evaluation approaches have been used in empirical studies to validate the proposed methods/techniques/approaches/findings regarding the application of PM in SMEs?* Regarding validation techniques, there were 4 studies that clearly stated their validation method. These studies all used different validation techniques. One study validated the findings from their case study through a simulation. Another study validated an established procedure through a single case study. Moreover, one study performed a cross-case analysis to validate their model, and lastly, one study validated their findings using the Fisher's exact test. Thus, it can be suggested to further validate results on the use of PM in organizations in future research.

Based on findings from this SLR, it is clear that more research on PM in SMEs is needed. Moreover, from the SLR, no studies were found that develop a methodology on the use of PM in SMEs. Furthermore, a need to develop a PM methodology with clear guidance was emphasized. This study resolves this gap by developing a methodology on the use of PM in SMEs.

# 3 Research Methodology

In this chapter, the research methodology that will be followed in this research is discussed. Moreover, the manner in which the methodology will be applied is presented.

## 3.1 Design Science Methodology

This research follows the Design Science Methodology (DSM) from [57]. The DSM is a proven methodology for doing design science in information systems research. It provides guidelines for doing research on an artefact in a context. Such an artefact may e.g., be a method, technique, or algorithm used in information systems. The context is the design, development, maintenance and use of software and information systems. This aligns with our research, since a methodology will be developed to apply PM in SMEs. The artefact (the methodology) will be empirically investigated in two case studies. The first case study will be done to refine the developed methodology, and the second case study will be done to validate the refined methodology. Moreover, expert and practitioner evaluations will be used for the validation of the methodology.

The case studies for this research concern the application of the PM methodology, that will be developed in this research, at eMagiz. eMagiz is a Dutch Enterprise Integration Platform as a Service (iPaaS) that enables quick and easy connections between applications and systems so that data streams can be automated and managed optimally. With around 30 employees, eMagiz can be classified as an SME [45]. At the start of this research, the employees of the SME had no familiarity with PM.

This case study research can be classified as experimental research. More specifically, single-case mechanism experiments will be conducted. In single-case mechanism experiments, the researcher studies individual cases, investigates phenomena that are produced by the architecture of a case, and intervenes with the case [57]. Since a PM methodology, PROMISE, will be developed in this research, and the methodology will be applied at an SME, where no application of PM exists yet, the research can be classified as such. The purpose of a single-case mechanism experiment is to validate a new technology, investigate problems in the field, and evaluate implementations. In this research, the purpose is to refine and validate PROMISE.

The DSM is followed to properly structure the research and to maximise the validity and value of the research outcomes. In design science, two activities are iterated: (i) the design of an artefact that improves something for stakeholders, and (ii) the empirical investigation of the performance of the artefact in a context. To properly conduct design science, the DSM design cycle as depicted in *Figure 8 Design Cycle from [57]* is followed.

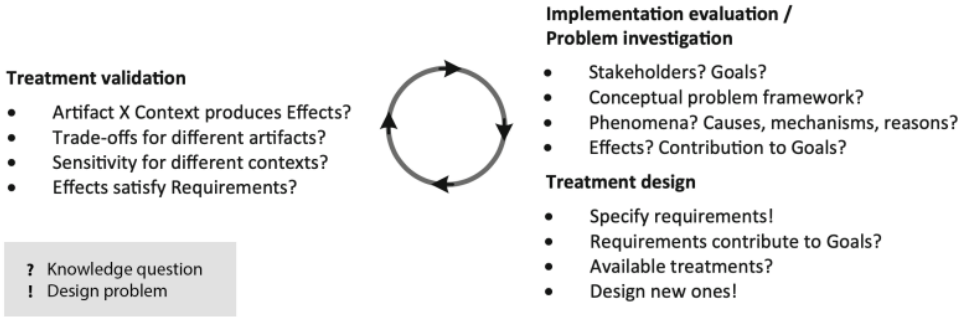



*Figure 8 Design Cycle from* [57]

## 3.2 Application of the DSM

The design cycle consists of three phases:

1. **Problem investigation**: The problem investigation is the first phase, and its goal is to understand the problem before formulation of the requirements and before starting with the design. In this research, the problem investigation is achieved through analysing findings from literature on PM in SMEs and PM methodologies. At the end of this phase, stakeholders and stakeholder goals are addressed.
2. **Treatment design**: During this second phase, the requirements are formulated, and the artefact is designed. In this research, the requirements are formulated through investigating requirements for effective PM methodologies available from literature, as well as findings from the problem investigation. The methodology proposed in this research is based on a combination of existing PM methodologies, as well as findings on the use of PM in SMEs. After the development of the methodology, the methodology is refined through a case study.
3. **Treatment validation**: In this last phase, the artefact is validated by demonstrating that it can contribute to stakeholder goals in the problem context. This is achieved by applying the refined methodology in an additional case study and obtaining expert and practitioner evaluations.

## 3.3 Summary

*Figure 9 Application of the Design Cycle* shows which chapters, research questions, and key activities correspond to each phase of the design cycle. For clarity purposes, a phase has been added, namely **Treatment Refinement**. Refinements may be applied during treatment design, but since two case studies will be executed, one to refine the methodology, and one to validate the methodology, this phase has been added separately.

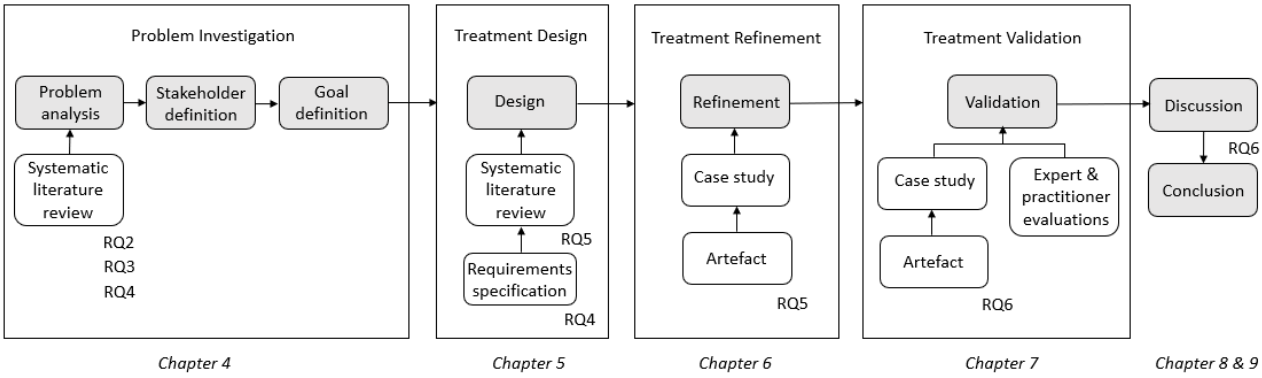



*Figure 9 Application of the Design Cycle*

# 4 Problem Investigation

In this chapter, the difference between PM in SMEs and large organizations will be investigated, using findings from the papers identified in the SLR. Moreover, existing PM methodologies will be evaluated, and a mapping of the phases of existing PM methodologies will be presented. Lastly, stakeholders and goals for this research will be defined.

## 4.1 PM in SMEs Versus Large Organizations

### 4.1.1 Results from the SLR

From the literature search, described in Chapter *2 Systematic Literature Review*, 11 studies were found that focus on PM in SMEs. These studies will be used to define differences on the use of PM in SMEs in comparison to large organizations. Moreover, challenges or other findings that could be relevant for defining our methodology will be discussed. Lastly, findings from the studies that focus on PM in large organizations or other organizational contexts, identified in Section *2.1.6 Review of Additional Papers*, will be included, where an overview of PM guidelines found from the literature will be given. While findings from studies on large organizations cannot directly be generalized to SMEs [46], some findings may hold for any PM project, independent of an organization's size.

SMEs mainly differ from large organizations with respect to their size and annual turnover. In the European Commission recommendation 2003/361/EC [62], it is stated that SMEs are enterprises that have less than 250 employees and an annual turnover of up to EUR 50 million [45]. Moreover, SMEs typically have less process maturity [27], and no internal skills to lead necessary process evolutions [18]. Furthermore, SMEs are characterized by low managerial skills and a low formalization level [32][51], which means that employees tend to have multiple roles to fulfil. SMEs do appear to house deeper IS/IT knowledge [17], and have short and immediate communication channels with decisionmakers [32]. Due to these differences, results from research on large organizations cannot necessarily be generalized to SMEs [26], as found by [46].

As a result of the lack of internal skills to lead necessary process evolutions, SMEs sometimes choose to have an out-of-date process rather than evolve it [18]. However, organizations should aim for the continuous usage of PM for it to be most effective [6]. Therefore, it is important that change management and automation efforts are implemented, and that PM is seen as a continuous company-wide activity [6], especially for SMEs.

Another challenge that was found from the studies focussing on PM in SMEs is the activity of choosing an appropriate case ID [11][53]. Choosing an appropriate case ID is an important step since it influences the outcome of the PM analysis largely, and thus the manner in which the data is interpreted [11]. This challenge seems likely to arise in organizations other than SMEs as well.

More PM challenges specific to SMEs are identified by [26], where challenges on the use of PM in organizations found from literature are evaluated in an SME context. Four PM challenges appeared to be SME specific: (i) preparation of event log data; pre-processing and cleaning of data such that it is suitable for applying PM, (ii) poor documentation quality; the documentation that describes the desired, or target, process is unreadable or too high level, (iii) awareness; creation of awareness of the benefits and costs for PM within the organization, and (iv) shifting manpower; the shift of manpower to fulfil PM tasks in the organization.

Some guidelines to combat the above-mentioned challenges are given by [26] as well: (i) find the right balance for the number of events in the dataset that is to be investigated, and

(ii) ensure that top management is involved in supporting PM in the organization. No guidelines are given on how to deal with the challenge of poor documentation quality. Moreover, the guidelines are very general, no elaboration is given on what could be a right balance for the number of events in a dataset, or on how to best involve top management. Important to note as well, is that [26] follows a different definition for SMEs in comparison to the definition used in our research. According to [26], SMEs are “companies that have less than 500 employees or have a revenue less than 50 million Euro per year”, p.127. This definition was obtained from the Institute for SME research in Bonn, Germany [61]. The company that was investigated in the paper employs round about 480 full time employees with small subsidiaries in other countries. According to the definition used in our research, this company would not be categorized as an SME, since the number of employees is > 250. Thus, the results from [26] might not immediately be generalizable to our study either. Nevertheless, since the findings are quite generic, they are expected to apply to organizations of a range of sizes.

### 4.1.2 Challenges, Guidelines, and other Findings

A literature search on PM challenges in organizations was executed by [26], and a table including all identified challenges from the literature review was established in the study. This table can be found through the following link: Applying Process Mining in Small and Medium sized IT Enterprises - Challenges and Guidelines | Zenodo [26]. Note that these challenges are not necessarily SME specific. One additional challenge was found in [19], also not necessarily SME specific. This challenge is the computational complexity, which concerns the time needed to compute results and the ability to produce relevant results.

Since an elaborate overview of PM challenges was already created by [26], the studies from our literature review were not further investigated for challenges. However, an overview of PM guidelines based on findings from literature is not available yet. Thus, the studies from our literature review, both studies on the use of PM in SMEs as well as studies focussing on PM in large organizations or other organizational contexts, were reviewed for guidelines. An overview of all the found PM guidelines can be found in *Table 1 PM Guidelines from Literature*. One additional study was added, namely the Process Mining Manifesto [3], since this is a widely cited study that provides several PM guidelines. Only guidelines provided in [3] that focus on organizational contexts have been included. Guidelines on e.g., algorithms or other technical details have been excluded.

*Table 1 PM Guidelines from Literature*

| Reference | Code | Guideline |
|---|---|---|
| [26] | LG1 | Begin with simple processes. |
| [26] | LG2 | Focus on core functionalities of a process mining software. |
| [26] | LG3 | Create a comprehensive knowledge base. |
| [26] | LG4 | Involve data protection stakeholders from the beginning. |
| [26] | LG5 | Consider process versions when evaluating event data. |
| [26] | LG6 | Find the right balance between precision and abstraction when creating a data set. |
| [26][52] | LG7 | Ensure top management support for process mining. |
| [6][54][30][3] | LG8 | PM should be a continuous company-wide activity. |
| [6] | LG9 | Organizations should implement change management and automation efforts. |
| [30] | LG10 | Process properties are more important to consider as compared to the type of processes. |
| [30] | LG11 | A sufficient amount of data must exist. |

| | | |
|---|---|---|
| [30] | LG12 | PM needs to align with the strategy and other operations of the company. |
| [30][3] | LG13 | Data needs to be high quality and complete. |
| [30][52] | LG14 | Employees need to agree on a transparent data policy before implementation. |
| [30][52] | LG15 | Governance has to be considered to determine who should be involved in PM activities and who is responsible for managing them. |
| [52] | LG16 | When new systems are introduced, event logging must be integrated. |
| [52] | LG17 | Event logs should be distributed through a centralized portal. |
| [3] | LG18 | Log extraction should be driven by questions. |
| [3] | LG19 | Models should emphasize the aspects that are relevant for a particular type of user. |

The guidelines provided in the literature are quite general and allow for multiple interpretations. For example, LG1: Begin with simple processes, is likely to be independent of the size of an organization. Thus, while the guidelines may not be specific for SMEs, it can be assumed that they are relevant due to their widely interpretable nature, and therefore can give some guidance for the development of our methodology.

*Table 2 Challenges, Guidelines, and other Findings on PM in SMEs from Literature* gives an overview of the challenges, guidelines and other findings on PM in SMEs specifically. Note that some challenges, guidelines, or other findings may apply to large organizations as well.

*Table 2 Challenges, Guidelines, and other Findings on PM in SMEs from Literature*

| Reference | Code | Challenges (C), guidelines (G), other findings (F) |
|---|---|---|
| [45] | LF1 | SMEs mainly differ from large organizations with respect to their size and annual turnover. |
| [27] | LF2 | SMEs typically have less process maturity. |
| [32][51] | LF3 | SMEs are characterized by low managerial skills. |
| [32][51] | LF4 | SMEs are characterized by a low formalization level. |
| [17] | LF5 | SMEs appear to house deeper IS/IT knowledge. |
| [32] | LF6 | SMEs have short and immediate communication channels with decision makers. |
| [18] | LC1 | SMEs typically have no internal skills to lead necessary process evolutions. |
| [18] | LC2 | SMEs sometimes choose to have an out-of-date process rather than evolve it. |
| [11][53] | LC3 | Choosing an appropriate case ID appears to be difficult. |
| [26] | LC4 | Preparation of event log data. |
| [26] | LC5 | Poor documentation quality. |
| [26] | LC6 | Creation of awareness. |
| [26] | LC7 | Shifting manpower. |
| [6] | LG20 | PM should be seen as a continuous company-wide activity. |
| [6] | LG21 | Organizations should implement change management and automation efforts. |

## 4.2 Existing PM Methodologies

For the development of a methodology on the use of PM in SMEs, it seems useful to first evaluate existing PM methodologies. From the literature search in our study, no methodologies on the use of PM in SMEs were found. However, two studies presented methods based on PM techniques [45][18]. From these studies, [45] develops a method for collaboration of SMEs based on PM techniques. The study is not focussed on the use of PM in organizations, it merely uses PM techniques to ensure collaboration between SMEs. Thus, the results from [45] are not very useful for the development of our methodology. The other study [18] develops a method on the Continual Evolution For Organisational Processes (CEFOP). They address the challenge of continuous usage of PM. The two main intentions of this method are (i) characterize the as-is process, and (ii) imagine the as-if process. The CEFOP process model is presented in *Figure 10 The map of the CEFOP Method, obtained from [18]*.

As is clear from *Figure 10 The map of the CEFOP Method, obtained from [18]*, the method does not provide clear PM steps. Nevertheless, the importance of establishing the existing process and modelling the desired process is emphasized and will thus be taken into account for the development of our methodology.

With respect to the studies on organizations without size restrictions, one study presented a methodology [20]. This methodology defines PM activities including tools to be used to support the activities, as well as results and reasons for using the specified tool for each activity. The methodology is given in *Figure 11 PM Methodology from [20]*.

This methodology defines much clearer steps as compared to the CEFOP method. However, the focus lies more on which tool to use during which step as compared to how to apply PM. For example, the process of selecting a dataset or business process is not described. Moreover, the methodology is not specialized for SMEs. Nevertheless, the methodology may be useful to consider when developing our methodology.

After having searched Google Scholar using the term “process mining methodologies”, a recent study [60] was found that compares existing PM methodologies with PM practices. The four PM methodologies that are compared in the study are the L*Life cycle methodology [4], the PMPM [33], PM$^2$ [23], and the PM project proposal [8]. Below, each of the methodologies is described in short.

### 4.2.1 L*Life cycle Methodology

First of all, the L*Life cycle methodology [4] describes five phases in a PM project and is based on the practical application of PM on more than 100 organizations. The visualization, including the five phases, can be found in *Figure 12 L*Life cycle Methodology from [4]*.

In stage 0, three types of PM projects are distinguished: (i) data-driven, (ii) question-driven, and (iii) goal-driven. Data-driven projects are mainly explorative, question-driven projects aim to answer specific (business) questions, and goal-driven projects strive to improve a process with the use of KPIs. Based on the type of project, activities may differ in consequent stages. Next to identifying the type of project, in this stage, the PM project needs to be planned with e.g., milestones and resource allocations.

In stage 1, data extraction has to be done, such that an event log can be obtained. In case of question-driven and goal-driven projects, the questions and KPIs, respectively, have to be defined here. No further guidelines for data extraction are given here.

In stage 2, the process model has to be created, where the activities in the process model have to refer to events in the event log. The process model is enhanced in stage 3 by adding

perspectives. These perspectives are an organizational, case, and time perspective. In stage 4, three activities are to be executed: (i) detect, (ii) predict, and (iii) recommend. This stage can only be executed when the event logs are of high quality, and the processes are structured.

### 4.2.2 PMPM Methodology

The second PM methodology that will be discussed is the PMPM [33]. This methodology consists of 6 phases, as can be seen in *Figure 13 PMPM Methodology from [33]*.

The Scoping phase of the PMPM concerns the identification of processes, and the determination of the objectives, where the same types of PM process projects as described in stage 0 of the L*Lifecycle methodology are considered. Moreover, this phase includes the determination of tools and techniques. The steps that are described are quite detailed, however, the tools and techniques are not specified.

In the Data Understanding phase, data is located, explored, and verified. It does not yet concern the extraction of data; this is done in the Event Log Creation phase. In the Event Log Creation phase, three dimensions are considered: (i) historic data vs live data, (ii) timeframe, and (iii) perspective. After the consideration of these dimensions, the data has to be extracted and prepared.

The Process Mining phase concerns the identification and application of PM techniques to answer business questions. Three main types of PM are mentioned here, namely (i) discovery, (ii) conformance, and (iii) enhancement. Moreover, a PM framework [1] is described here.

In the Evaluation phase, the modelled results have to be verified, validated, and accredited. Moreover, it has to be decided whether the PM project should be elaborated. In the Deployment phase, it has to be identified if and how the modelled process can be improved, and the results have to be presented to the organization.

### 4.2.3 PM^2 Methodology

A third methodology is $PM^2$, which consists of 6 phases as shown in *Figure 14 PM2 Methodology from [23]*.

The objective of the Planning stage is setting up the project and determining the research questions. More specifically, the business processes have to be selected, research questions have to be identified, and a project team has to be composed. After that, the Extraction stage has to be conducted, which concerns the determination of the scope, extraction of event data, and the transferring of process knowledge. The event logs are created in the third stage, namely Data Processing. It seems that the authors interpret extraction of event data as the collection of all event data, which is then transformed into event logs through e.g., aggregating events and enriching logs, so that it can be better used for PM. In the Data Processing stage, the event logs are actually taken from the system.

The fourth stage is about Mining and Analysis. Four types of activity are distinguished for this stage, namely (i) process discovery, (ii) conformance checking, (iii) enhancement, and (iv) process analytics. The first three activities are PM techniques, which have also been described in other methodologies, the fourth activity is a complementary analysis technique.

In the Evaluation stage, the results have to be diagnosed, verified, and validated. Lastly, in the Process Improvement and Support stage, improvements have to be implemented and operational support should be provided.

### 4.2.4 PM Project Proposal Methodology

The last PM methodology that will be discussed is the PM project proposal [8]. The methodology consists of 4 stages, as presented in *Figure 15 PM Project Proposal Methodology from [8]*.

Since the main activities are already clear from *Figure 15 PM Project Proposal Methodology from [8]*, the phases will only be described in short. First of all, the goal of the Project Definition stage is to understand the business process and its main problems. The second stage, Data Preparation, is meant to extract the data and assess its quality. In the third stage, the Process Analysis, PM techniques are applied to discover a process model and to analyse the performance. In the last phase, the Process Redesign, improvements are suggested, assessed, and implemented.

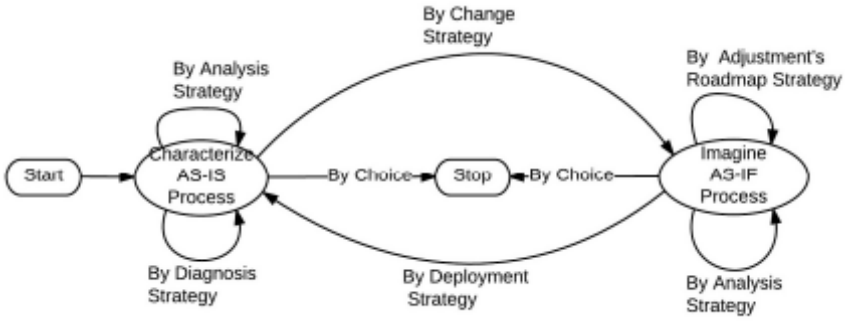



*Figure 10 The map of the CEFOP Method, obtained from* [18]

| Activity | Chosen Tool | Result | Reason |
|---|---|---|---|
| Import event log as CSV file | Disco | Event log prepared for process mining | Disco supports csv input type |
| Pre-processing | Disco | Removed incomplete cases, limited amount of start and end activities, included only activities that refer to a particular product | Instead of using three different filters in ProM to reach the same result, Disco offers Endpoint and Attributes filter |
| Automatically generate process model | Disco | Fuzzy process model | Disco (fuzzy) miner is used when log data are unstructured and complex, to simplify the model |
| Present process model in performance and frequency perspectives | Disco | Visually discovered loops, waiting times and bottlenecks | Performance analysis can only be accomplished in ProM if Petri net model notation is used to present the process model |
| Export event log as XES | Disco | Event log as XES file | Filtered event log is exported as XES in order to be imported into ProM |
| Import XES file | ProM | Event log prepared for process mining | Filtered event log is imported into ProM for further case/time and social network analysis |
| Apply pattern abstraction visualizer on event log | ProM | Discovered tandem array pattern type loops | More detailed pattern information than initially discovered with Disco |
| Apply handover of work plug-in on event log with originators | ProM | Handover of work social network | Disco does not support social network (organizational) mining |
| Present ranking view of social network | ProM | Discover employees crucial for the process execution | Disco does not support social network (organizational) mining |

*Figure 11 PM Methodology from* [20]

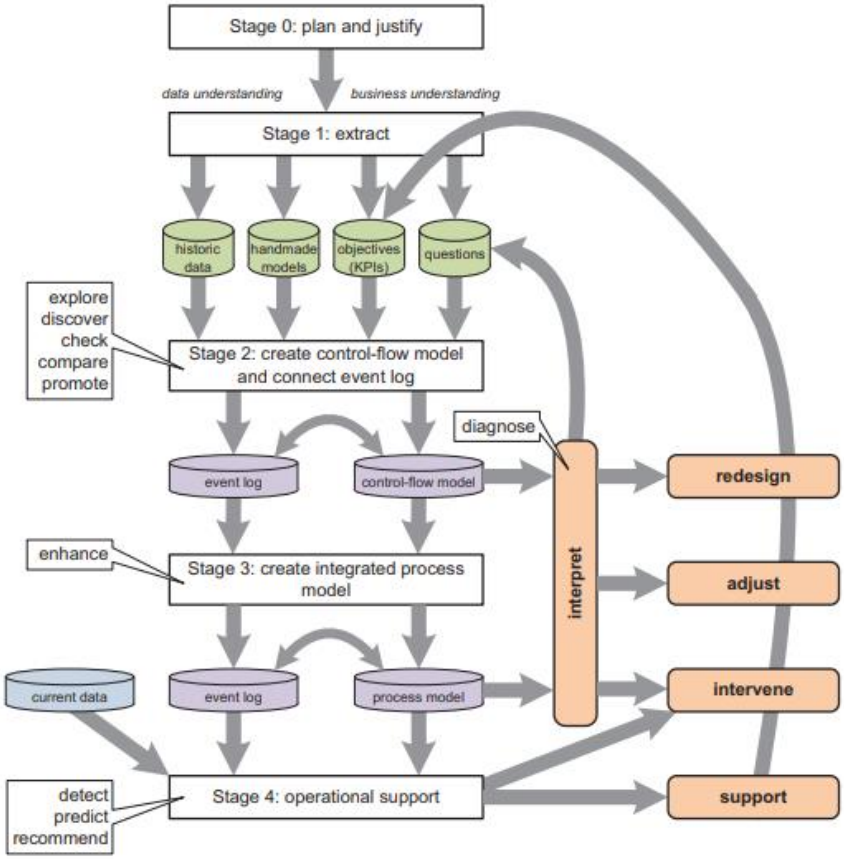



*Figure 12 L*Life cycle Methodology from* [4]


*Figure 13 PMPM Methodology from* [33]

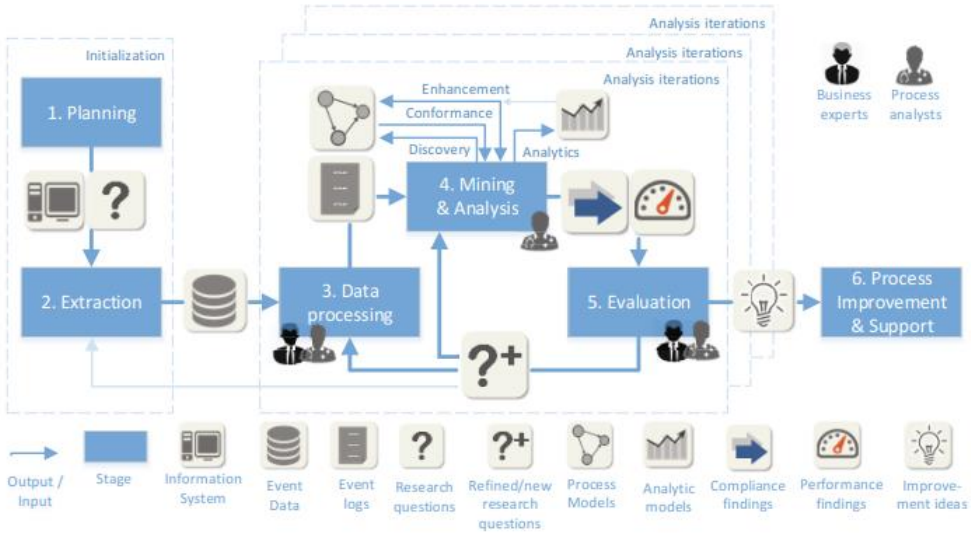



*Figure 14 PM² Methodology from* [23]

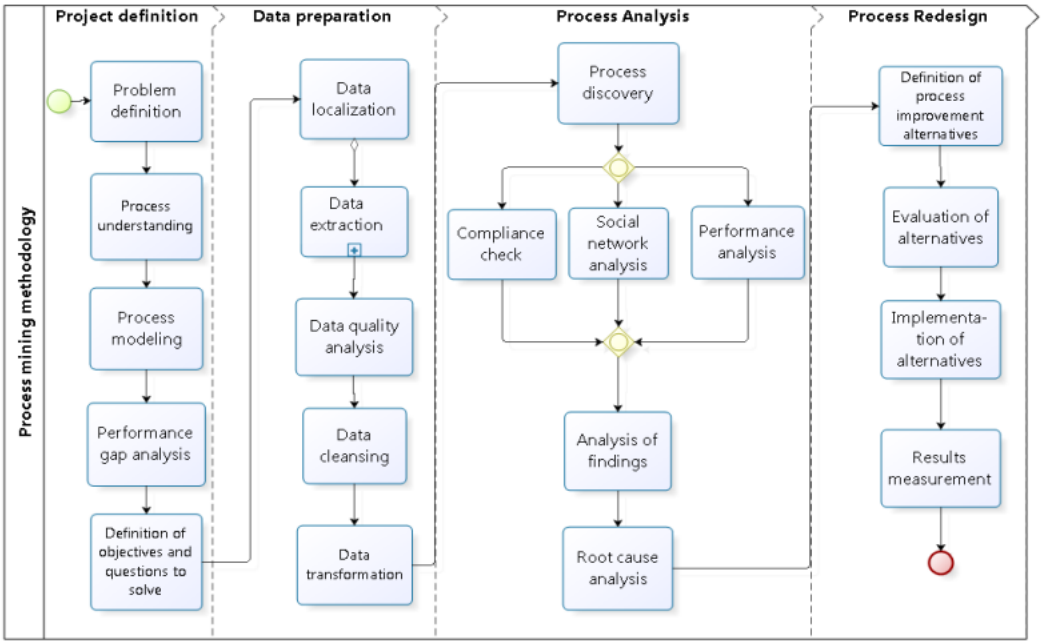



*Figure 15 PM Project Proposal Methodology from* [8]

### 4.2.5 Mapping of the Phases of PM Methodologies

In [60], the four above-described methodologies are presented in a comparison framework. In this comparison framework, elements are presented that should be addressed in PM methodologies according to the authors, as investigated in their study. The PM methodologies are compared based on these elements. However, it is clear from the phases from the methodologies that there is quite some overlap among them. Therefore, it seems useful to also have a mapping of the phases of the four PM methodologies to check which activities are addressed by the methodologies. Therefore, a mapping has been made, where phases, or stages, of all methodologies have been combined or split, in such a way that a general overview of all phases addressed in all methodologies could be made. The mapping is given in *Table 3 Mapping of the Phases of PM Methodologies*. Note that the categorization of the phases is generic. Not all phases fit exactly into a defined phase, but it has been done as precise as possible.

*Table 3 Mapping of the Phases of PM Methodologies*

| **Phases** | **L*Lifecycle** [4] | **PMPM** [33] | **PM²** [23] | **PM project proposal** [8] |
|---|---|---|---|---|
| **Phase 1: Plan, Scope & Define** *Identify the type of PM project, identify business processes, determine tools & techniques.* | Stage 0: Plan and Justify | 1. Scoping | Stage 1: Planning | 1. Project definition |
| **Phase 2: Data Exploration & Understanding** *Locate, explore, and verify event data.* | - | 2. Data Understanding | Stage 2: Extraction | - |
| **Phase 3: Event Log Creation** *Create event logs and extract this event log data.* | Stage 1: Extract | 3. Event Log Creation | Stage 3: Data Processing | 2. Data preparation |
| **Phase 4: Process Models Creation** *Make process models based on the event log data.* | Stage 2: Create Control-Flow Model and Connect Event Log | 4. Process Mining | Stage 4: Mining and Analysis | 3. Process Analysis |
| **Phase 5: Analysis & Enhancement** *Analyse process models, enhance models, address PM techniques,* | Stage 3: Create Integrated Process Model | 4. Process Mining | Stage 4: Mining and Analysis | 3. Process Analysis |

| | | | | |
|---|---|---|---|---|
| *answer business questions.* | | | | |
| **Phase 6: Evaluation** *Verify & validate results.* | - | 5. Evaluation | Stage 5: Evaluation | - |
| **Phase 7: Process Improvement & Presenting** *Identify potential improvements, implement improvements, make recommendations & present results.* | Stage 4: Operational Support | 6. Deployment | Stage 6: Process Improvement and Support | 4. Process Redesign |

In the study from [60], 19 PM project elements have been formulated. These elements should be addressed in PM methodologies, to give PM practitioners guidance and support when applying PM and to stimulate the adoption of PM in organizations [60]. The elements have been derived from practitioner experiences and available literature on PM challenges and enablers. The practitioner experiences were obtained from interviewees having a variety of roles related to PM, and working in organizations of different sizes, ranging from small organizations to large organizations.

From the 19 PM project elements, 2 elements were not addressed at all in any of the methodologies. These are organizational willingness and the creation of process dashboards. The remaining elements were either partially or fully addressed in some or all of the methodologies. Thus, for our methodology it is important to address the 2 elements that have not been addressed in the evaluated methodologies, and to add to the completeness of the remaining elements.

Regarding the methodology from [20], organizational willingness and the creation of process dashboards were not addressed either. The activities given in the methodology are generally less elaborate as compared to the four methodologies evaluated by [60]. However, the methodology from [20] does give some more information about PM tools.

## 4.3 Stakeholders

According to [57], part of the problem investigation is the identification of stakeholders. A stakeholder is a person, group of persons, or institution that is affected by the treatment of the problem [57]. To identify the stakeholders in this study, the taxonomy from [10] will be used. This taxonomy is a conceptual framework for classifying development stakeholders based on an onion model. Depending on the artefact that is developed, only a subset of entities is relevant [10]. First, possible stakeholders will be identified, after which they are classified according to the taxonomy from [10], as is suggested by [57]. This classification concerns the stakeholder role and the level of involvement.

The PM (adapted) methodology should be usable by any practitioner in an SME, where the goal is to implement PM within the SME. To achieve this, mainly resources from within the SME are needed. First of all, management is needed to shift manpower or hire experts to conduct PM [26][52]. Moreover, architects that have knowledge about existing processes within the SME are needed, so that it can be determined which process could be analysed.

Furthermore, once a process is selected, it is important that a process expert is in place who has extensive knowledge about the process. In contrast to the architect, the process expert should have more thorough knowledge about a singular process. Next to that, the PM activities will have to be executed. This could be done by a designated PM team. Depending on the knowledge on PM that resides within the SME, a PM expert may need to be hired to guide the PM activities. Since PM should be a continuous activity [6], a maintenance team will also be needed, to ensure that the PM results remain as desired. To obtain the data that is needed to perform the PM activities, some technical insight may be needed, and some more technical activities may have to be executed, which could be done by a development team. Lastly, data protection stakeholders need to be involved [26]. Data is the main source of the PM activities, and this data is vulnerable to hackers. To mitigate the threat of hackers, these data protection stakeholders are in place. Several more stakeholders may be needed, but it was tried to include at least the necessary stakeholders above.

All stakeholders fulfil a stakeholder role and have a certain level of involvement. An overview of the stakeholders including a classification into roles, descriptions of the roles, and their level of involvement according to [10] is given in *Table 4 Stakeholders*. The architect has two roles, namely the role of normal operator, which is performed by providing an overview of existing processes, and the role of functional beneficiary, since the architect should benefit from the output of the PM activities. The reason for this, is that PM should clarify existing processes and identify process inefficiencies, which should aid the architect in the daily practices. Next to the architect, the CEO should benefit from the methodology, because the results from PM have value for the business, in terms of monetary values as well as non-monetary values [12].

*Table 4 Stakeholders*

| **STAKEHOLDER** | **STAKEHOLDER ROLE** | **STAKEHOLDER ROLE DESCRIPTION ADAPTED FROM** [10] | **INVOLVEMENT** |
|---|---|---|---|
| MANAGEMENT | Sponsor | Initiate development and obtain funding for it. | Low |
| ARCHITECT | Normal Operator | Have routine interaction with the system. | High |
| | Functional Beneficiary | Benefit from the results or outputs of the system. | Medium |
| PROCESS EXPERT | Consultant | Support some aspect of the system development. | Low |
| PM TEAM | Developer | Engineers, analysts, designers, programmers, etc. that are directly involved in the system development. | Low |
| PM EXPERT | Operational Support | Give advice on how to operate the system. | High |

| | | | |
|---|---|---|---|
| MAINTENANCE TEAM | Maintenance Operator | Interact with the system to keep it running. | High |
| DEVELOPMENT TEAM | Developer | Engineers, analysts, designers, programmers, etc. that are directly involved in the system development. | Low |
| DATA PROTECTION STAKEHOLDER | Developer | Engineers, analysts, designers, programmers, etc. that are directly involved in the system development. | Low |
| CEO | Functional Beneficiary | Benefit from the results or outputs of the system. | Medium |
| HACKER | Threat Agent | Responsible for, or attempt to, bring harm to the organization. | Low |

## 4.4 Goals

The goal of the PM methodology is to guide practitioners in the application of PM, such that PM can be implemented optimally in SMEs. In Section *5.1 Methodology Requirements*, the interpretation of optimally is elaborated on. Apart from this main goal, several stakeholder goals can be formulated. In *Table 5 Stakeholder Goals*, these goals are listed for each stakeholder, apart from the hacker stakeholder.

*Table 5 Stakeholder Goals*

| STAKEHOLDER | GOAL |
|---|---|
| MANAGEMENT | To ensure that the right resources are in place to conduct the PM activities. |
| ARCHITECT | To have a clear overview of the current processes within the organization.<br>To reduce the number of process inefficiencies within the organization. |
| PROCESS EXPERT | To have detailed knowledge about the process selected for PM. |
| PM TEAM | To execute the PM activities efficiently and effectively. |
| PM EXPERT | To support in performing the PM activities if needed such that they can be executed successfully. |
| MAINTENANCE TEAM | To monitor the outcome of PM and to suggest improvements. |
| DEVELOPMENT TEAM | To deliver the data needed to perform PM. |
| DATA PROTECTION STAKEHOLDER | To protect the data that is used for PM against privacy and security threats. |

| | |
|---|---|
| CEO | To ensure that the results from PM have value for the business. |

## 4.5 Summary

In this chapter, the difference between PM in SMEs versus large organizations was investigated using the literature identified in the SLR. To summarize the findings, a table containing guidelines on PM in organizations, and a table containing challenges, guidelines, and other findings on PM in SMEs was created.

Moreover, several PM methodologies were discussed in this chapter and an initial mapping of phases was achieved. The CEFOP Method from [18] and the PM Methodology from [20] may be used when formulating guidelines for our methodology. The phases from the L*Lifecycle [4], PMPM [33], $PM^2$ [23], and PM project proposal [8] methodologies were used for an initial establishment of phases for our methodology. This was achieved through mapping the phases of these methodologies. The phases will be defined, evaluated and refined during the remainder of this research.

The stakeholders and the stakeholder goals were also formulated in this chapter. In total, 10 stakeholders and 10 goals were identified. Moreover, each stakeholder was assigned a stakeholder role according to the taxonomy from [10], as suggested by [57].

# 5 Treatment Design

In this chapter, requirements for PROMISE will be formulated. Then, the methodology phases as defined in the previous chapter, and PM project elements from [60] will be mapped. After that, guidelines for each of the PM project elements will be defined. Lastly, a first version of PROMISE will be presented.

## 5.1 Methodology Requirements

For the development of a methodology, it is important to first completely understand the definition of a methodology, as well as requirements for methodologies. As mentioned before, a methodology is about revealing practices of researchers, as well as ideas and presuppositions behind those practices, in a systematic way [14]. The words method and methodology are often interchanged [40], so it seems important to give a clear definition of these words, which will be followed in this thesis. According to [42], methodologies deal with general principles to generate knowledge, while methods are techniques and procedures to follow to conduct research. Moreover, methods are determined by methodologies [42]. In this thesis, this is interpreted as the following: A methodology will be developed, since the result is expected to be applicable in any organization, and, thus, yields general principles. The methodology can be specified to work for a single organization, in which case it would become a method.

In the literature, no general requirements for methodologies could be found. However, based on our interpretation as described above, the following requirement can be formulated:

**R1**: The PM (adapted) methodology must be usable in any SME.

Since PM can be applied in any organization [5], this seems to be a reasonable requirement.

Our research question for this thesis, as defined in Section *1.4 Research Questions*, is: *How to design a methodology on the use of PM that gives practical guidelines so that PM can be implemented optimally in SMEs?*

This research question yields the need for two more requirements, namely regarding (i) practical guidelines, and (ii) optimal implementation.

With respect to the practical guidelines, the purpose of these is that any practitioner who has a basic level of knowledge, experience, and skills in the field of data and business processes, should be able to understand and use the methodology. Thus, any such practitioner should be able to set up PM in their organization by following the methodology, where the practical guidelines should aid in this. This yields the following requirement:

**R2**: The PM (adapted) methodology must be understandable and usable by any practitioner in an SME, who has a basic level of knowledge, experience, and skills in the field of data and business processes.

Regarding the optimal implementation, this means that the methodology should ensure that PM is implemented in the best or most favourable way. This can be interpreted in the following, namely that by applying this methodology, the highest business values that can be obtained by applying PM can be achieved. According to [12], the business values that PM can bring to organizations are (i) process efficiency, (ii) monetary values, and (iii) non-monetary values. However, no clear PM requirements are given that would yield business values. It also seems quite reasonable to assume that many factors and relations play a role in achieving these business values. Nevertheless, 3 key features of PM have been identified in a study [12] that afford the creation of business values. These are (i) data & connectivity, (ii) process visualization, and (iii) process analytics. Thus, it seems logical that, at least,

these PM features will have to be addressed in our methodology to strive for an optimal implementation of PM. Thus, the following requirements, based on the description given in [12], can be formulated:

**R3**: The PM methodology must address features that extract, integrate, and combine process-related data.

**R4**: The PM methodology must address features that visualize process execution.

**R5**: The PM methodology must address features that generate various process related KPIs.

As described at the end of Section *4.2 Existing PM Methodologies*, PM project elements that should be addressed in PM methodologies have been defined in [60]. Therefore, the last requirement will be the following:

**R6**: The PM methodology must address all PM project elements defined in [60].

In *Table 6 Requirements*, all requirements are given.

*Table 6 Requirements*

| **Requirements** |
| --- |
| R1. The PM (adapted) methodology must be usable in any SME. |
| R2. The PM (adapted) methodology must be understandable and usable by any practitioner in an SME, who has a basic level of knowledge, experience, and skills in the field of data and business processes. |
| R3. The PM methodology must address features that extract, integrate, and combine process-related data. |
| R4. The PM methodology must address features that visualize process execution. |
| R5. The PM methodology must address features that generate various process related KPIs. |
| R6. The PM methodology must address all PM project elements defined in [60]. |

## 5.2 Phases & Elements Mapping

For the development of PROMISE, the PM project elements from [60] will be used as steppingstones. Moreover, the comparison framework from [60] will be used to check which methodologies (partially) address these elements, and what appears to be missing. Elements that have been addressed in existing PM methodologies will be evaluated in our case studies, and these case studies will also be used to add to the completeness of (parts of) elements that lack in existing PM methodologies.

First, a categorization of the PM project elements into phases is made. The phases have been derived from *Table 3 Mapping of the Phases of PM Methodologies*, since this will give the most complete overview, where all phases from existing PM methodologies are addressed. Later, after categorizing the project elements and evaluating them in our first case study, the phases may be changed according to the findings. The categorization can be found in *Table 7 Phases & PM Project Elements*. Note that the PM project elements have been categorized according to what seems reasonable for each phase. It may be the case that the project elements are (partially) addressed in different phases from the evaluated PM methodologies.

The requirements R3, R4, and R5, that concern PM features, have been categorized into the phases as well. This is mainly done to understand how and where the requirements may be

addressed. The requirements will not specifically be evaluated here, this will be done in Section *7.3 Validation*.

*Table 7 Phases & PM Project Elements*

| **Phase** | **PM Project Elements from** [60] |
|---|---|
| **Across Phases** | - Iterative nature |
| **Phase 1: Plan, Scope & Define**<br>*Identify the type of PM project, identify business processes, determine tools & techniques.* | - Organizational willingness<br>- Stakeholder involvement<br>- Linking business goals to PM projects<br>- Vendor selection<br>- Process selection<br>- Project goal<br>- Desired insights and KPI selection<br>- Familiarity with process mining |
| **Phase 2: Data Exploration & Understanding**<br>*Locate, explore, and verify event data.* | - Data availability<br>- R3 |
| **Phase 3: Event Log Creation**<br>*Create event logs and extract this event log data.* | - Data extraction and preparation<br>- R3 |
| **Phase 4: Process Models Creation**<br>*Make process models based on the event log data.* | - Creation of process dashboards<br>- R4 |
| **Phase 5: Analysis & Enhancement**<br>*Analyse process models, enhance models, answer business questions.* | - Analysis of dashboard<br>- R5 |
| **Phase 6: Evaluation**<br>*Verify & validate results.* | - Interpretation and conclusion<br>- Validation<br>- R5 |
| **Phase 7: Process Improvement & Presenting**<br>*Identify potential improvements, implement improvements, make recommendations & present results.* | - Defining improvement actions<br>- Quantify, select, monitor improvements<br>- Communicating quick wins<br>- Continuous effort<br>- R5 |

A difficulty that arose during the categorization, is that [60] does not provide definitions for the PM project elements. The interpretations of these elements were based on the gap analysis from [60], as well as findings from the PM methodologies and definitions from literature. Afterwards, the categorization of the project elements, as well as our interpretation of the elements could be verified with one of the authors of [60]. The interpretations will be elaborated on in Section *5.3 Phases & Elements Descriptions*.

As mentioned before, the PM project elements were found in organizations ranging in size from small to large. However, since the PM project elements allow for many interpretations, it seems that all elements can be assumed to be relevant. The manner in which they are applicable for SMEs will depend on the findings from literature on PM in SMEs, as well as our case study.

Below, each of the phases is described, where all PM project elements are addressed, and guidelines are formulated. This is done based on (i) findings from the literature on PM in

SMEs, (ii) findings from the literature on PM in large organizations, and (iii) existing PM methodologies as described in this thesis.

## 5.3 Phases & Elements Descriptions

### 5.3.1 Across Phases

The iterative nature addresses the need to refine and revalidate steps that have been taken during the application of PM. For example, after having defined a process, it may be found that the data quality is insufficient, meaning that a process reselection may have to be done. So, across phases, it is important to check whether it is needed to go back to a previous phase. Mentioned by one of the authors of [60] was that it is important to verify and validate findings after each phase, especially during the formulation of research questions. Next to this, it was emphasized that, mainly for SMEs, it is important to move across phases quickly to present an MVP (minimum viable product) as soon as possible. The reason for this, is that SMEs generally have a lower budget, and PM is not the highest priority, as found during interviews from [60]. The presentation of results and an MVP should help to gain trust.

**G1**: Verify findings after each phase and move to previous phases if required.

**G2**: Move across phases quickly to present an MVP as soon as possible.

### 5.3.2 Phase 1: Plan, Scope & Define

*Identify the type of PM project, identify business processes, determine tools & techniques.*

#### *5.3.2.1 Organizational Willingness*

It is crucial that the organization is willing to put effort into PM [60]. Therefore, it is important to ensure top management support for PM [26][52] because the management needs to be willing to shift manpower or hire experts to conduct PM, which is a challenge especially for SMEs [26]. To convince top management, success stories of PM on example processes could be used [26].

However, since SMEs are characterized by low managerial skills and a low formalization level [32][49], it seems important to involve and convince not only top management but also employees that have knowledge about existing data and processes. PM should be a company-wide activity [6][3][30][54], and involving such employees from the beginning might help to ensure this. Moreover, SMEs have short and immediate communication channels with decision makers [32], which may constrain an open culture and result in resistance to IT security investments [32]. Therefore, it seems important to maintain an open culture and involve not only top management but employees which knowledge may be beneficial to PM as well. This will also help in creating awareness for the benefits and costs of PM, which is a challenge specifically for SMEs [26].

It is also important to consider who should be involved in PM activities and who is responsible for managing them [30][52]. Especially for SMEs, which have limited resources [18], it should be clear who is able to invest time into PM activities. Here, it should also be taken into account that PM should be a continuous activity, so people will have to be able to not only set up a PM project, but also to maintain it. The most important stakeholders are the PM expert and process expert [23], so these will definitely have to be appointed.

**G3**: Involve top management as well as employees that have knowledge about existing data and processes.

**G4**: Convince the company of the importance of PM, e.g., by providing success stories of PM on example processes.

**G5**: It needs to be established who should be involved in PM activities and who is responsible for managing them, taking into account the continuity of PM.

#### *5.3.2.2 Stakeholder Involvement*

Data protection stakeholders should be involved from the beginning [26]. The following needs to be achieved: (i) inform workers council when a PM software is acquired, (ii) establish rules for PM, (iii) determine where event data can be stored safely and (iv) ask process participants for approval [26].

Moreover, employees will need to agree on a transparent data policy [30][52]. This should ensure a centralized vision and should prevent the emergence of differing policies within the company [52].

**G6**: Involve data protection stakeholders.

**G7**: Ensure that employees agree on a transparent data policy.

#### *5.3.2.3 Linking Business Goals to PM Projects*

It is important that PM aligns with the strategy and other operations of the company [30]. Therefore, the purpose of the PM project needs to be linked to the strategy of the company. This can be achieved through identifying research questions [23], and linking them to business goals [60]. Important here is that a readiness check is done to evaluate whether enough data and commitment is present, as found by one of the authors of [60]. The reason for this, is that many organizations are not familiar with PM yet, causing a lack of understandability. To prevent the company from stopping PM activities, it needs to be established whether enough commitment can be given.

**G8**: Link the strategy of the company to the PM goals.

**G9**: Check whether the company is able to give enough data and commitment.

#### *5.3.2.4 Vendor Selection*

For choosing a PM software, the core functionalities of existing tools have to be considered [26]. If it is important to understand the working of the algorithm that is producing PM models, it is suggested to use an academic tool such as ProM (available at www.processmining.org). The reason for this, is that ProM provides several PM algorithms that can be mathematically explained, such as the Alpha Miner, Heuristic Miner, and the Inductive Miner [43]. Commercial tools, such as Disco (available at https://fluxicon.com/disco/) and Celonis (available at https://www.celonis.com/), make use of a Fuzzy algorithm which is not academic. However, commercial tools do generally provide a more intuitive, understandable and easy to use interface. Moreover, they provide the possibility to create dashboards with e.g., KPIs, while ProM does not have this functionality.

It is also possible to use a combination of tools, such as suggested in [20]. To conclude, it is important to find a PM tool that provides required functionalities.

**G10**: Choose a suitable PM tool by considering the core functionalities of existing tools and the desired functionalities.

#### *5.3.2.5 Process Selection*

For the selection of a process, it is more important to consider process properties as compared to the type of process [30]. Such process properties are e.g., the amount of data produced and the variations of a process, where it is desired that the number of repetitions is high and that variations can be identified [30]. This does not mean that the process should be complex, on the contrary, it is advised to begin with simple processes [26]. Moreover, the

people involved in the process should be considered, and the strategic goals should be taken into account. The reason for this, is that SMEs typically have no internal skills to lead necessary process evolutions, and therefore sometimes choose to have an out-of-date process [18]. To ensure that process evolutions can be made, the stakeholders who will be involved in the to-be analysed process should agree with the process selection, and enough resources should be available to mine the process.

To give a more concrete guideline, the number of process steps should be higher than 2. The reason for this, is that processes of two steps appeared to not yield useful findings, as found by one of the authors of [60].

**G11**: Begin with a simple process, with a minimum of 3 process steps.

**G12**: Select a process by considering process properties, stakeholders and strategic goals.

#### *5.3.2.6 Project Goal*

Three types of PM projects may be set up: (i) data-driven, (ii) goal-driven, and (iii) question-driven projects [8][33][4]. Data-driven projects are explorative and are powered by the availability of event data. Goal-driven projects aim to improve a process with respect to particular KPIs. Question-driven projects aspire to answer specific questions.

Since SMEs typically have less process maturity [27] it can be advisable to start with a question driven project [4]. Moreover, question-driven projects help to scope the project and guide data extraction efforts [4]. The most difficult project to apply is a data-driven project [33], because of its explorative character [1]. Moreover, a data-driven project is often not possible in terms of time and budget [33]. Because SMEs typically have limited resources [18], this type of project is inadvisable. Goal-driven projects may be set up as well, but a disadvantage of goal-driven projects is that it may be difficult to determine how PM can be used [33].

**G13**: Start with a question-driven project, or, if it is clear how to use PM to achieve a certain goal, a goal-driven project may be set up.

#### *5.3.2.7 Desired Insights and KPI Selection*

Three types of PM techniques exist: (i) discovery, (ii) conformance, and (iii) enhancement [1]. Process discovery techniques produce a model using event logs without any a-priori information. Conformance techniques compare an existing process with an event log of that process. Enhancement techniques extend or improve an existing process model based on event logs from the actual process.

Depending on the type of PM tool, many discovery, conformance, and enhancement techniques exist [33]. Discovery techniques need to balance four criteria, namely fitness, precision, generalization, and simplicity [4]. The Fuzzy Miner aims to balance these four criteria [4]. However, it is not clear how this algorithm exactly works. So, if insight into the working of the algorithm is needed, another algorithm such as the Heuristic Miner or the Inductive Miner can be chosen. Note that most commercial tools only provide a Fuzzy Miner.

Several types of conformance or enhancement techniques can be applied as well, depending on the type of PM tool. Therefore, it is important to consider the types of PM techniques that need to be used before selecting a PM tool. The type of techniques that are needed depends on the project goals and project scope. For example, if the goal of applying PM is only to discover what causes extra costs, discovery techniques may be sufficient. However, if the intention is to also reduce those costs, and improve the process, all types of PM techniques may be needed.

Next to choosing the types of PM, it is important that KPIs are formulated in case the project is goal driven. For question-driven projects it is important that questions are formulated. The formulation, or adjustment of questions and KPIs is an iterative process [4][23].

**G14**: Consider the types of PM techniques that are needed with respect to the PM project goals and scope.

**G15**: Formulate KPIs for a goal-driven project and formulate questions for a question-driven project.

#### *5.3.2.8 Familiarity with Process Mining*

A basic understanding of PM is beneficial for all involved in the evaluation of the results, because the interpretation of findings can be difficult and time-consuming otherwise [23]. Moreover, it is advised to start with question-driven projects when organizations are not familiar with PM [4]. This is already included in G13.

In the literature, no more information regarding the familiarity with PM was found. However, it seems important that all who are involved in the PM project should have some basic understanding of PM. The reason for this, is that they could have some insight that may benefit the execution of several PM activities. For example, an employee might remember some event data that is stored that may be relevant.

**G16**: Ensure that all who are involved with steps of the PM project have a basic understanding of PM.

### 5.3.3 Phase 2: Data Exploration and Understanding

*Locate, explore, and verify event data.*

#### *5.3.3.1 Data Availability*

For a PM project, data needs to be available [33]. It is necessary that process events are automatically recorded and can be extracted, and that there is some level of guarantee that the recorded events match reality [3]. The data should contain at least the following three attributes: case, activity, and timestamp [5]. The case refers to a process instance, activity to a task or operation, and timestamp to the time of the event [5]. In the previous phase, the activity to select a process based on process properties, such as the amount of data, was described. The activity in this phase is about understanding and exploring this data more thoroughly, such as by checking for a case ID, activity, and timestamp. New or adjusted KPIs of objectives may emerge [4], and it may be needed to alter the scope based on the data understanding [33].

**G17**: Make sure that data is available, and that the data can be extracted.

**G18**: Ensure that the data contains a case ID, activity, and timestamp.

Next to the availability of data, it is important for PM that the data is of high quality and complete [30][3]. The quality of data can be judged according to four criteria: (i) trustworthiness; it needs to be safe to assume that recorded events actually happened and that the attributes of the event data are correct, (ii) completeness; no events may be missing, (iii) semantics; the events should be well-defined, and (iv) safeness; privacy and security concerns are addressed when recording the events [3][33]. To benefit from PM, organizations should strive for event logs of the highest possible quality level [3]. In case the data does not meet the criteria, the data should be improved in order to apply PM, or other data should be sought.

**G19**: Check whether the data quality is sufficient and strive for the highest possible quality level of event logs.

It should also be considered what the right balance is for the number of events in the data set, in other words the precision and abstraction level [26]. The reason for this, is that too many events may clutter the results, but too few events might cause the results to be meaningless [26]. If it appears that there are too many events, the scope may be reduced, e.g., only focussing on purchasing events instead of sales and purchasing events. Similarly, if there appear to be too few events, the scope could be augmented. Note that it is also possible to filter events, this can be done in the next phase. Thus, if it appears to be difficult to find the right balance, this could be solved to some extent in the next phase. Moreover, the choice for the type of case ID can influence the number of events, which will be discussed in the next phase as well.

**G20**: Find the right balance for the number of events in the data set.

The log extraction should be driven by questions, because this will help in extracting meaningful event data [3]. Since the event data is already selected in this phase, it is important that it is evident that the selected data can help in answering the questions. The questions should have already been formulated in the previous phase (G15).

**G21**: Ensure that the selected data can help in answering the research questions.

### 5.3.4 Phase 3: Event Log Creation

*Create event logs and extract this event log data.*

#### *5.3.4.1 Data Extraction and Preparation*

For the data extraction, a case ID, activity, and timestamp need to be selected. In the previous phase, it should already have been checked whether it would be possible to select these aspects from the dataset. For SMEs, but likely also for larger organization, choosing an appropriate case ID appears to be difficult [11][52]. The choice of an appropriate case ID is an important step and can completely change the outcome of the analysis, and as a result the manner in which the data is interpreted [11]. Unfortunately, no guidelines on the selection of a case ID could be found in the literature. Nevertheless, we will give some suggestions here that can be validated during the case study.

First of all, for the selection of a case ID it is important to thoroughly understand the meaning of the data, which should already have been evaluated during the quality assessment (G19). However, if some parts of the data remained unclear, this should be solved here, because it is necessary to understand the meaning of a potential case ID. Next, it is important to consider the research questions, and visualize a potential process model. This could help to better understand the type of case ID that is needed to answer the research questions. Lastly, it needs to be taken into consideration how many activities belong to a case ID. Too many activities may cause the process model to become cluttered, while too few activities may cause the process model to become meaningless. All in all, choosing an appropriate case ID is likely to be an iterative process, where many potential case IDs are evaluated.

**G22**: Take time to choose an appropriate case ID, ensuring to evaluate all possible case IDs.

For the selection of the activities, the research questions need to be taken into account as well, same for the timestamps [8]. Moreover, it is important to understand that the selected timestamp should be the timestamp that belongs to the selected activity. Once the required

dataset has been constructed, the data can be extracted. The dataset can often be exported to different kinds of data files [33], and PM tools generally support several input types [20].

**G23**: Select a case ID, activity and timestamp by taking into account the questions that should be answered by the PM analysis and ensuring that the selected timestamp belongs to the selected activity.

**G24**: Extract the data once the required dataset has been constructed.

After the data has been extracted, the data has to be prepared. This includes several tasks.

First of all, event logs may be enriched with various additional attributes [38][23]. This can be done by deriving or computing additional events and attributes based on the event log itself, or by adding external data [23]. The additional attributes can give more insight into the process or may be used to apply filtering. Moreover, unnecessary data aspect and records can be removed, and outliers or missing values can be resolved [33]. If event logs come from more than one source, it is needed to merge different datasets into one dataset [33].

**G25**: Prepare the extracted dataset so that it is suitable for further processing.

Once the desired dataset has been achieved, the event logs may be filtered. This should be done in case the number of events of the extract event log seems too large [33][23][8][4]. Filtering is a data processing step that can help reduce complexity and focus on a specific part of the dataset [23]. Three types of filtering techniques can be applied: (i) slice and dice; remove events based on the values recorded for a specific attribute, (ii) variance-based filtering; group similar traces, and (iii) compliance-based filtering; remove events that do not comply with a given rule or fit a given process model [23]. Whether to apply filtering, and which technique to apply depends on the dataset.

**G26**: Apply filtering if it is needed to reduce complexity or to focus on a specific part of the dataset.

### 5.3.5 Phase 4: Process Models Creation

*Make process models based on the event log data.*

#### *5.3.5.1 Creation of Process Dashboards*

PM dashboards are interfaces that include graphical indicators based on PM techniques [29]. Several commercial PM tools, such as Disco and Celonis, provide the possibility to make dashboards, which may visualize KPIs, and other elements that are relevant to track. These dashboards can also show created process models such as in [29]. Below, the creation of process models is discussed first. Afterwards, the creation of PM dashboards is addressed.

Process models can be created with several PM tools, such as ProM, Celonis, and Disco. ProM provides several different algorithms to mine process models, whereas most commercial tools only provide the Fuzzy Miner. All algorithms will require a case ID, activity, and timestamp, as selected in the previous phase. Next to that, several PM techniques can be applied, namely discovery, conformance and enhancement. This has already been discussed *Section 5.3.2.7 Desired Insights and KPI Selection*, and a suitable PM tool should have been chosen in Phase 1, as discussed in Section *5.3.2.4 Vendor Selection*.

The process models should highlight the aspects that are relevant for a particular type of user [3]. This means that the models may show different perspectives, such as dataflow, time, and resources, as well as different levels of granularity and precision. Moreover, different levels may be created, such as a strategic level, tactical level, and operational level

[3]. Lastly, differing process versions have to be taken into account [26]. The desired visualization of the process models should be discussed with the involved stakeholders, mentioned in Section *5.3.2.2 Stakeholder Involvement*.

**G27**: Create a process model with the chosen PM tool using its desired algorithms and techniques.

**G28**: Ensure that the process models highlight the aspects that are relevant for a particular type of user and take into account differing process versions.

If conformance checking is done, reference models are needed to compare the discovered process with the documented process. However, for SMEs, a good quality of documentation or reference models is considered challenging [26]. This can be resolved by creating a comprehensive knowledge base [26], to address all available knowledge. Such a comprehensive knowledge base can be created with the help of the involved stakeholders.

**G29**: Create a comprehensive knowledge base with involved stakeholders for conformance checking.

The event log data can be either pre mortem, meaning that it is current data, or post mortem, which is historic data [1]. For the creation of dashboards, both types of data can be used. However, real-time data can be better used for monitoring. Such monitoring can help to achieve actionable insights, which is a key factor driving PM adoption [55]. Nevertheless, historic data can be useful for the creation of dashboards as well, since it can give great insight into the process and the results from the PM activities. In the literature, no specific guidelines on the design of a PM dashboard could be found. However, several commercial PM tools give guidance and show examples of dashboards.

**G30**: Create PM dashboards to gain insight into the results from the PM activities.

### 5.3.6 Phase 5: Analysis & Enhancement

*Analyse process models, enhance models, answer business questions.*

#### *5.3.6.1 Analysis of Dashboard*

The purpose of analysing the dashboard, including the process models, is to answer the business questions [8]. Thus, it is important to check the business questions, and to see whether these can be answered by the process models or other dashboard features. The following core activities have to be executed: (i) compare the discovered process to the desired process, (ii) analyse performance indicators and bottlenecks, and (iii) analyse the relationship between resources and activities [8]. The comparison of the discovered process to the desired process can be easily achieved if conformance checking with the PM tool can be done. Otherwise, the process will have to be thoroughly analysed with the help of the involved stakeholders. It is important to involve the right stakeholders to analyse the results, which should be the stakeholders with extensive knowledge about the process, as well as about the PM activities. The reason for this, is that the results should be interpreted correctly, which may be difficult if stakeholders have no knowledge about PM or about the steps that have been taken. Thus, ideally, process experts are involved to guide the analysis, such that useful results can be obtained [23].

**G31**: Conduct analyses in close collaboration with process analysts and business experts in a highly iterative and interactive manner.

**G32**: Make sure to compare the discovered process to the desired process, check for performance indicators and bottlenecks, and analyse the relationship between resources and activities.

Next to analysing the models, the models may be enhanced. This can be achieved by adding additional visual analytics, such as histograms of events per case [23]. Moreover, several additional analyses may be added, such as a social network analysis [8]. Next to that, the process models may be digitally animated to enlighten the observation of activities that produce congestion [8]. Lastly, additional performance information could be added to the models [23].

**G33**: Enhance the process model by adding e.g., additional visual analytics and digital animations.

### 5.3.7 Phase 6: Evaluation

*Verify & validate results.*

#### *5.3.7.1 Interpretation & Conclusion*

The purpose of the evaluation is to relate the findings from the analysis to improvement ideas that achieve the project's goals [23]. For this, it is important to check whether the results are interpreted correctly, to distinguish outstanding results from expected results, and to identify or refine research questions for potential further iterations [23]. To keep improving the process, new projects should be suggested to ensure that PM is applied in a long-term basis [33]. This will ensure that PM is seen as a continuous activity so that it is effective [6]. Lastly, conclusions have to be drawn with respect to the research questions.

**G34**: Check whether the results are interpreted correctly and draw conclusions with respect to the research questions.

**G35**: Suggest actions for improvements and think about other possible elaborations of PM in the organization.

#### *5.3.7.2 Validation*

To evaluate the results further, it is important to verify and validate the results [23][33]. It is essential that process experts are involved for the verification and validation of the results, because it can be difficult to determine the causes of unexpected results from the analysis [23].

For the verification of the results, the correctness of the findings has to be investigated [23]. This can be done by checking the soundness of the models [33]. The soundness of a model guarantees the absence of live locks, deadlocks, and other anomalies [7].

For the validation of the results, the findings have to be compared to the claims of process stakeholders [23]. These process stakeholders are experts of the process that was investigated. In this way, the degree in which the model represents the real process has to be determined [33]. To evaluate the quality of a process model, the following metrics can be used: (i) fitness; the degree of observed behaviour being captured in the process model, (ii) precision; the degree in which the model allows for too much behaviour, (iii) generalization; the degree in which the system is described and not only the data, and (iv) simplicity; the degree in which the model is understandable [33]. The process mining evaluation framework from [50] can be used to guide the process.

It is also important to evaluate the degree in which the results meet the objectives that were set for the PM project [33]. This should be done by the initiator of the PM project, often the process owner [33].

**G36**: Verify the PM results by investigating the correctness of the findings, e.g., by checking the soundness of the models.

**G37**: Validate the results by checking the degree of fitness, precision, generalization, and simplicity of the model.

**G38**: Evaluate the degree in which the results meet the PM project objectives.

### 5.3.8 Phase 7: Process Improvement & Presenting

*Identify potential improvements, implement improvements, make recommendations & present results.*

#### *5.3.8.1 Defining Improvement Actions*

Ideas for improvement, which should have been defined in the previous phase, should be turned into actions, and several more improvement actions may be defined [23]. These improvement actions may be the following: (i) redesign; changes to the process may be made due to the insights obtained, (ii) adjust; adjustments to the process can be made, such as changes in resources, (iii) intervene; problems may be revealed regarding particular cases or resources, and (iv) support; PM can be used for operational support [33][1][4]. Redesign and adjust actions have to do with strategic or tactical decisions, while intervene and support actions improve the operational process and can be implemented more quickly [33]. Regarding technicality, redesign and support activities have technical implications, while adjust and intervene actions are about changing processes, in nontechnical terms [33].

One or more of the above-mentioned improvement actions should be identified to be helpful to achieve the business goals and improve the process [33]. Operational support is the most ambitious form of PM, and only possible for structured processes [4]. It may be provided by detecting running cases that are problematic, predicting their future, or suggesting recommended actions [23]. Two improvement actions for any organization that applies PM are to integrate event logging, especially when new systems are introduced, and to distribute event logs through a centralized portal, such that they are more easily accessible [52]. Since SMEs appear to house deeper IS/IT knowledge [17], these improvement actions may be even better executable by SMEs.

**G39**: Define improvement actions concerning one or more of the following activities: redesign, adjust, intervene, and support.

**G40**: Consider the integration of event logs with systems and the distribution of event logs through a centralized portal.

#### *5.3.8.2 Quantify, Select, Monitor Improvements*

After defining improvement actions, they have to be prioritised [8]. Moreover, it is important to decide on how the improvements can be monitored. This means that the state of the process has to be analysed several moments in time [33].

**G41**: Prioritise improvement actions and decide on the manner in which improvements will be monitored.

#### *5.3.8.3 Communicating Quick Wins*

Recommendations and predictions need to be presented to the people working on the corresponding cases [4]. This information, as well as a summary of the project addressing the insights gained can be used to improve the process [33].

**G42**: Communicate the recommendations, predictions, and other results to the involved stakeholders.

#### *5.3.8.4 Continuous Effort*

Evident from the literature, and mentioned several times before, is that PM should be a continuous activity, as indicated by [6][3][30][54]. Therefore, it is important that change management and automation efforts are applied by the organization [6]. Ensuring that a continuous effort is being made by companies can partly be achieved by deciding on elaborations for the PM project [33], and following up on the improvement actions. Moreover, an analysis project may be set up to measure the improvements [23].

To ensure that improvements are realized, change management plays a key role [6]. Several theories and approaches on change management exist [16]. Successful change management focusses on both strategic and operational issues, and some important practices are (i) disciplined project management, (ii) clear accountability and goals, (iii) communication, (iv) staff involvement, and (v) management commitment [44]. The successful management of change is a highly required skill [16], and it therefore seems important to invest a good amount of time in this.

**G43**: Ensure that PM is a continuous activity in the organization, e.g., by elaborating the PM project, following up on improvement actions, and measuring improvements.

**G44**: Implement change management to ensure that improvements are realized.

## 5.4 Overview of PROMISE

### 5.4.1 The PROMISE Visualization

The PROMISE Visualization is given in *Figure 16 PROMISE Visualization*. It shows the phases that need to be followed to end up with the desired PM results. For each phase, guidelines are given that should aid the practitioner in accomplishing the phases. The guidelines are given in the order in which they should be followed. In comparison to the phases defined in *Table 7 Phases & PM Project Elements*, two additional phases appear in the PROMISE Visualization, namely Phase 1: Business Understanding and Phase 9: Change Management. Reasons for this are given below. An overview of all phases and the specific guidelines belonging to each phase is given in Appendix *D PROMISE Phases & Guidelines*. The two guidelines that hold for all phases (G1&G2) have been added to the first phase.

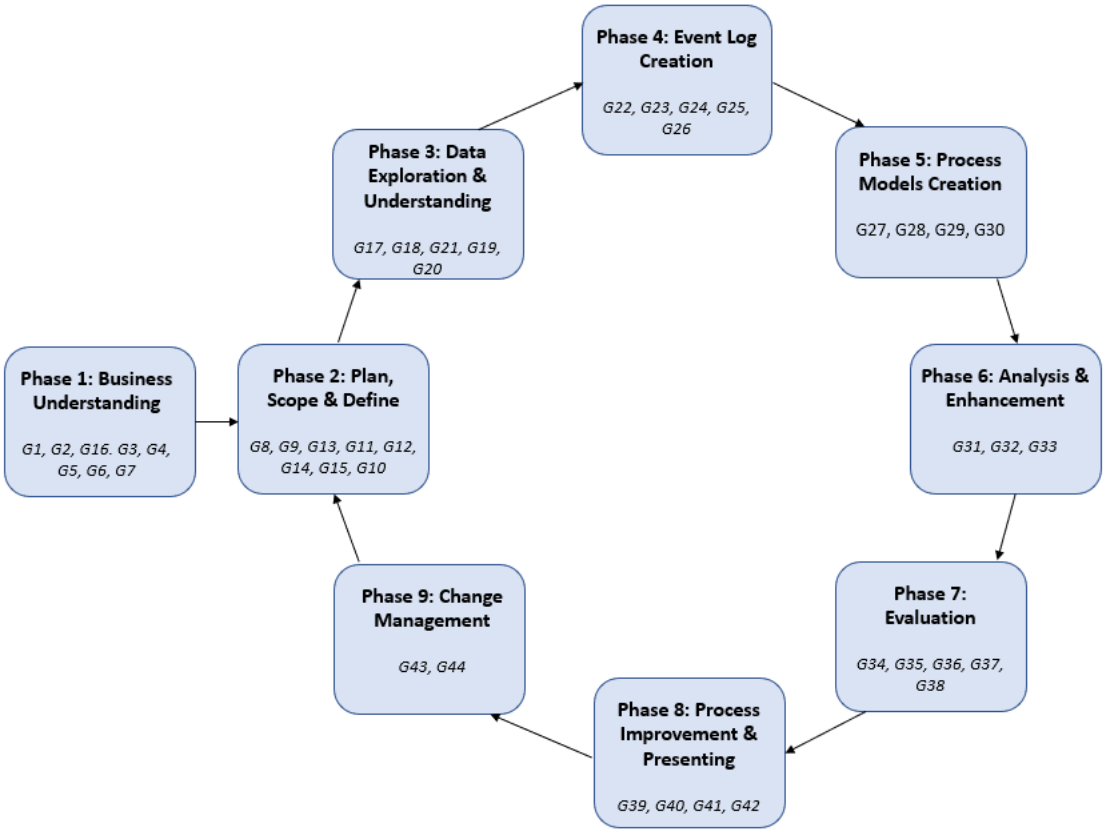



*Figure 16 PROMISE Visualization*

Apparent from *Table 7 Phases & PM Project Elements*, is that many PM project elements belong to Phase 1: Plan, Scope & Define. Most of these project elements are either not addressed or only partially addressed in the PM methodologies evaluated in [60]. Moreover, after formulating guidelines that address these elements, it was found that many guidelines belonged to this phase, and that a clear split could be made. Part of the guidelines addressed a need for business understanding, concerning a familiarization with the company and the involvement of stakeholders. The other part of the guidelines concerned the planning and scoping, focussed on defining the PM project. Therefore, Phase 1: Plan, Scope & Define was split into Phase 1: Business Understanding, and Phase 2: Plan, Scope & Define.

Moreover, Phase 9: Change Management was added, which represents the need for the continuity of PM. It has been addressed several times in this study that this is an important aspect to consider when applying PM. However, existing PM methodologies do not address this (clearly) [60]. To represent this need for PM being a continuous activity, this phase was added. Phase 9 should be entered after PM results have been obtained and presented, and process improvements have been suggested and prioritised. The purpose of the phase is to monitor the current PM results, and it functions as a bridge between the presentation of the PM results and the start of working out potential improvements or setting up a new PM project by going through the phases again. The phase that needs to be entered when starting with an improvement is Phase 2. The reason for this, is that a business understanding needs to be obtained in Phase 1, which should already have been established during the first iteration of the phases. Working out the improvements, however, should be seen as a new (small) PM project, which is planned, scoped and defined in Phase 2. This cycle within the visualization should ensure that continuous efforts are made to

achieve the desired business values that can be delivered by PM, and to effectively implement change management.

Below, a short description of each of the phases is given.

### 5.4.2 Phase 1: Business Understanding

*Determine organizational willingness, involve stakeholders, and familiarise stakeholders with PM.*

In this phase, a clear understanding of the business needs to be obtained. This concerns involving top management, employees that have knowledge about data and processes in the organization, and data protection employees. Moreover, other employees that are involved in the PM project need to be informed about PM, such that they have basic knowledge about it. The company will need to be convinced of the importance of PM, and employees have to agree on a transparent data policy.

### 5.4.3 Phase 2: Plan, Scope & Define

*Identify the type of PM project, identify business processes, determine tools & techniques.*

The purpose of this phase is to clarify and shape the PM project. First, the strategy of the company needs to be established. Then, research questions have to be formulated and linked to the business goals. Moreover, the type of PM project has to be chosen and a process needs to be selected. Lastly, it has to be determined which PM techniques and tools are needed, which can be guided by formulating KPIs or questions. Several activities in this phase may be iterated over several times, such as the process selection.

### 5.4.4 Phase 3: Data Exploration & Understanding

*Locate, explore, and verify event data.*

In this phase, the availability of the data needs to be checked, making sure that it contains a case ID, activity, and timestamp, and can be extracted. Moreover, the quality of the data has to be evaluated, and it needs to be checked whether the selected data can help in answering the research questions. Furthermore, the right balance for the number of events in the data set has to be found. Several rounds to check whether the dataset contains the required properties should be executed, and the objectives and scope should be revised to see whether the dataset is relevant.

### 5.4.5 Phase 4: Event Log Creation

*Create event logs, extract and prepare this event log data.*

After a dataset has been selected, event logs have to be created and extracted, which is done in this phase. Furthermore, the dataset needs to be prepared to make it suitable for subsequent PM activities. Lastly, the relevancy and appropriateness of the dataset need to be checked through several rounds, which also holds for the preparation of the dataset.

### 5.4.6 Phase 5: Process Model Creation

*Make process models based on the event log data.*

The goal of this phase is to create the process models. This is to be achieved by using the chosen PM tool and PM techniques. It needs to be taken into account that the process models should highlight the desired aspects and process version(s). Moreover, a comprehensive knowledge base has to be created such that conformance checking can be done. Next to process models, process dashboards can be created, where both activities should be revised several times to end up with the best results.

### 5.4.7 Phase 6: Analysis & Enhancement

*Analyse process models, enhance models, answer business questions.*

In this phase, the findings from the PM activities have to be analysed, preferably with process experts. Bottlenecks and performance indicators need to be checked for, and relationships have to be investigated. This should be done in a highly iterative and interactive manner. Lastly, the model may be enhanced using e.g., additional visual analytics and digital animations.

### 5.4.8 Phase 7: Evaluation

*Verify & validate results.*

The PM outputs should be evaluated by verifying and validating the results. It needs to be checked whether the results have been interpreted correctly and conclusions with respect to the research questions have to be formulated. Moreover, actions for improvements and other potential elaborations of the project may be suggested. To verify the results, the correctness of the findings needs to be investigated, and the results have to be validated by checking the degree of fitness, precision, generalization, and simplicity of the model. This should be done iteratively to ensure the completeness and accuracy of the findings.

### 5.4.9 Phase 8: Process Improvement & Presenting

*Identify potential improvements, implement improvements, make recommendations & present results.*

The suggestions for improvement made in the previous phase can be identified further in this phase. Two improvement actions that should be considered are the distribution of event logs through a centralized portal and the integration of event logs. The improvement actions need to be prioritised and it needs to be determined how they can be monitored. Lastly, all results have to be communicated to the involved stakeholders.

### 5.4.10 Phase 9: Change Management

*Ensure PM continuity, realize improvements.*

In this phase, it has to be determined how PM can be a continuous activity within the organization. This can be done by elaborating the PM project, making improvements, and measuring those. To ensure that the improvements are realized, change management needs to be implemented. Once this phase is finished, phase 2 may be entered to start with the implementation of (small) PM projects to improve PM results.

## 5.5 Summary

In this chapter, requirements for the methodology were formulated. These concerned the usability and understandability of the methodology, as well as aspects that the methodology should address. Next to this, a mapping of PM elements and phases was achieved, and for each phase guidelines were formulated. It appears to be important to verify findings after each phase and to move across phases quickly to present an MVP as soon as possible. Moreover, it is important to obtain a good business understanding, and to involve and convince stakeholders. Several more guidelines were formulated, which can be found in Appendix *D PROMISE Phases & Guidelines*. The phases and guidelines were also depicted in a visualization, and short descriptions for each phase have been presented.

# 6 Treatment Refinement

To refine the methodology designed in this study, a case study was performed at eMagiz, an SME located in Enschede, the Netherlands. In this case study, PROMISE was applied to implement PM. The SME had no familiarity with PM, so the methodology was used to start up their first PM project and report on the findings. Afterwards, the methodology was refined. The results of applying PROMISE are described below.

## 6.1 Results per Phase

### 6.1.1 Phase 1: Business Understanding

eMagiz is a Dutch Enterprise integration Platform as a Service (iPaaS) that enables quick and easy connections between applications and systems. This ensures that data streams are automated and managed optimally. The company has around 30 employees working in one building and manages 3 main teams: (i) Sales & Marketing, (ii) Expert Services, and (iii) Development, where the Development team is split into an ILM team and a Cloud team.

To ensure that all who may be involved with the PM project have a basic understanding of PM, the concept was shortly presented during a knowledge share. Thereafter, several employees were asked about potential processes to be evaluated and asked about the availability of data. The company supervisor of this study gave several suggestions for processes that could be analysed, and this was investigated further. Because it was already established that a project would be undertaken to start with PM, it was not needed to convince the company of the importance of PM. Moreover, it was not necessary to persuade top management, because supervisors for the project had already been appointed.

With respect to data protection, rules and regulations about this were already in place within the company. These rules had to be followed for the PM project. Moreover, it was decided that no customer data would be used for the PM project, such that no permission from customers would have to be obtained to execute the PM project. Only data from the company itself would be used for the project. Furthermore, it was decided that company names would have to be anonymized in this report. So, the companies were given pseudonyms, which are used in this report.

In *Table 8 Stakeholders Case Study 1*, the stakeholders needed for the case study have been identified, according to the stakeholders and goals defined in *Table 5 Stakeholder Goals*.

*Table 8 Stakeholders Case Study 1*

| STAKEHOLDER | CASE STUDY STAKEHOLDER |
|---|---|
| MANAGEMENT | Manager Expert Services |
| ARCHITECT | CTO |
| PROCESS EXPERT | ILM Process Experts |
| PM TEAM | Student & Supervisors |
| PM EXPERT | First UT Supervisor |
| MAINTENANCE TEAM | Expert Services Team |
| DEVELOPMENT TEAM | ILM Development Team |
| DATA PROTECTION STAKEHOLDER | Information Security Officer |
| CEO | Manager Expert Services |

### 6.1.2 Phase 2: Plan, Scope & Define

The main goal of the PM project for the company was to find out whether PM could be useful for the company. Thus, no plan with respect to the type of process or dataset was in place.

Moreover, not much data was stored in general, and if data was stored, it could not be obtained easily. Therefore, no clear dataset could be selected. To mitigate this challenge, several potential processes and datasets were identified. Eventually, it was decided to further investigate integration lifecycle management (ILM). ILM is the management of integrations from their first creation to final disposition.

In eMagiz, five phases are followed to implement ILM, namely:

1. Capture; to draw systems and integrations.
2. Design; to add message definitions and mappings.
3. Create; to generate integrations.
4. Deploy; to deploy integrations.
5. Manage; to monitor integrations.

ILM data models were accessible in Mendix, and the data could be viewed using pgAdmin. It was chosen to investigate ILM because this could give more insight into how integrations are currently managed within the organization. eMagiz is a low code development company, meaning that it should be possible to set up integrations quickly. Therefore, it could be interesting to e.g., see how long it takes to set up the first integration until the first release. However, since no clear dataset was in place, it had to be investigated first whether it was possible to identify such processes. Thus, the data was analysed to select a process that could be used for mining.

Selecting an ILM process posed a challenge, because the number of ILM data models was very large, and the meaning of all entities, relations, and attributes was not immediately clear. Therefore, several processes were identified first, after which three main potential processes were selected initially: (i) release process, (ii) flows & flow versions process, and (iii) testing process. In general, the ILM data was quite incomplete, e.g., most entities did not contain timestamp attributes. Moreover, some data was missing, e.g., flows may be removed, after which they are completely erased from the data. Due to these factors, the potential processes had to be investigated further before choosing one of the processes. This mainly concerned of checking for timestamps, because it was clear that several case IDs or events could be chosen. The company supervisor also helped in the choice of a process by considering the importance and relevance of all the processes. After taking into account these factors, it was chosen to first investigate the release process further.

The release process starts with a user activating created integrations on an environment. The release process ends once such a release is replaced by a new release. A release is a combination of different versions of integrations and is created in the deploy phase from eMagiz. Apart from phases, eMagiz holds three environments, namely test, acceptance and production. Releases may be promoted from one environment to the other. Moreover, several release versions may be created before a release is activated, and releases can be created by different users if they have access rights for it.

Process properties were considered by creating several datasets using different case IDs, activities, and timestamps. This was the only way in which process properties such as the amount of data produced and the variations of a process could be considered, because no predefined dataset was available. These process properties were considered before formulating questions, because some more familiarization with the data was needed. The reason for this, is that no clear idea of what the data would exactly represent was present yet.

It was decided to first model the release process with respect to the environment on which releases are made or to which they are promoted, which should be a simple process. It should show the type of environment, either test, acceptance, or production, and the order in which releases are made on these environments (e.g., test → acceptance → production).

After having investigated the data more thoroughly, the following questions were formulated:

1. In which order do customers activate releases with respect to the test, acceptance, and production environment?
2. How does the order in which customers activate releases differ from the expected order?
3. On which environment do customers mainly activate releases?
4. What are the release patterns for each customer?
5. Which steps, patterns, and customers yield a high throughput time?

The types of PM techniques that will be used are discovery, conformance checking, and enhancement. Discovery techniques will be used to answer the questions defined above, and conformance checking will be done to see how the answers differ from what is expected. Enhancement will be done by defining improvement actions.

The type of PM tool that will be used is Celonis. The reason for this, is that the Celonis tool is very intuitive and can help to quickly gain insights due to several features, e.g., selecting different process versions. Moreover, Celonis provides the possibility to create PM dashboards. In *Table 9 Tools & Techniques Case Study 1*, an overview of the tools and techniques that will be used is given.

*Table 9 Tools & Techniques Case Study 1*

| Tools | Techniques | Purpose |
|---|---|---|
| Data viewing tool *(Mendix)* | No specific technique needed | View data models |
| Data query tool *(PgAdmin)* | Query language *(SQL)* | View data, extract & create datasets *(CSV)* |
| PM tool *(Celonis)* | Discovery algorithms, dashboarding *(Fuzzy Miner)* | Create process models & dashboards |

### 6.1.3 Phase 3: Data Exploration & Understanding

In the previous phase, it was already found that (i) the data can help in answering the research questions, (ii) the data contains a case ID, activity, and timestamp, and (iii) the data is available and can be extracted. The reason for this already having been investigated in the previous phase, is that these aspects were considered when selecting a process based on the process properties. More specifically, the research questions were formulated after having selected a process, because it was not yet clear what could be investigated. That is because it was not yet known what (types of) data could be provided and what insights could be useful.

To check whether the data quality was sufficient, the data had to be evaluated with respect to the trustworthiness, completeness, semantics, and safeness of the data. This was achieved by creating datasets with different case IDs, activities, and timestamps. The reason for several datasets having been created is that, while the process had already been defined, it was not yet exactly clear what the data would represent. For example, several

relationships between entities exist, and several entities could be used to seemingly represent the same information. To evaluate the quality, several quality requirements were formulated, which are given in *Table 10 Quality Aspects & Requirements Case Study 1*.

*Table 10 Quality Aspects & Requirements Case Study 1*

| Quality Aspect | Requirement |
|---|---|
| Trustworthiness | The releases should represent real releases, meaning that they should actually have happened and are correct, e.g., with respect to when a release was made. |
| Completeness | The release information should be complete, meaning that no releases should be removed. |
| Semantics | It should be clear what the meaning of the data is. I.e., the meaning of the case IDs, events, and timestamps should be evident. |
| Safeness | Privacy and security concerns should be addressed when recording the events. |

Taking into account the quality requirements, case IDs, activities and timestamps were chosen. Several datasets were created, and evaluated with respect to the quality requirements, after which one dataset was selected. Thus, while according to the methodology a case ID, activity, and timestamp should be chosen in the next phase, this was already done in this phase.

The trustworthiness of the chosen dataset seemed good, it was checked with several employees that the correct entities and relationships were chosen. With respect to the completeness, the release information was expected to be good. No releases were removed from the entity being used, so the data appeared to be complete. The semantics of the data was discussed with some employees as well and seemed to be right. With respect to the safeness, only technical eMagiz data was evaluated, no sensitive employee data, so the safeness was good.

The right balance for the number of events in the data set is explained in the next phase, because it was difficult to evaluate it without filtering.

### 6.1.4 Phase 4: Event Log Creation

As mentioned before, the case ID, activity, and timestamp were chosen in the previous phase. In *Table 11 Attributes Case Study 1*, the chosen case ID, activity, and timestamp are described.

*Table 11 Attributes Case Study 1*

| Type of Attribute | Notation | Description |
|---|---|---|
| Case ID | MessageBus id + ChangeEvent summary version | Releases belong to a message bus, which has an id. When a release is set to active, this is stored in the 'summary' attribute of the entity 'ChangeEvent', including the version number. Together with the message bus id, this version is unique. Thus, the case ID is the id from the message bus to which the release |

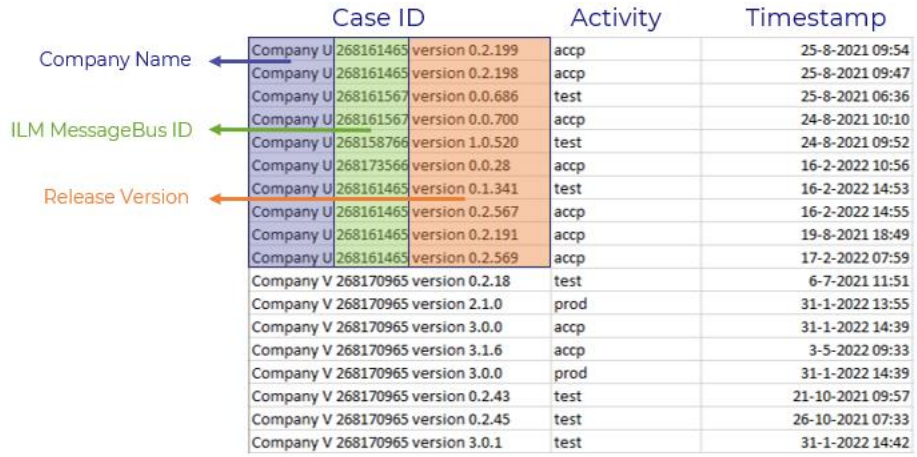
| | | |
|---|---|---|
| | | belongs, together with the version of the release. |
| Activity | Bus environment | The message bus is related to a deploy bus. This deploy bus contains the attribute ‘environment’, which can be either prod, accp, or test. These represent the production, acceptance, and test environment, respectively, which are the activities in the PM model. |
| Timestamp | ChangeEvent createdate | The ‘ChangeEvent’ entity has an attribute ‘createdate’, which contains the timestamp of when a release is set to active. This attribute will therefore be used as timestamp. |

In short, releases are a unique combination of the version number of a release and the ILM message bus in which the release is activated. A release may run on the three environments (test, accp, prod) at the same time, or different releases can run on different environments. Releases can be promoted from one environment to the other, and releases can be stopped on an environment, and re-activated. When a release is activated on an environment, the timestamp of activation is stored. Thus, a release within a message bus (case ID) can be activated on an environment (activity) at a specific time (timestamp).

The data was extracted by downloading it to a CSV file, a functionality provided by PgAdmin. Two filters were applied to make the dataset more suitable. First of all, only customers were included in the dataset. Sometimes environments are used to e.g., do some personal testing. Therefore, it was necessary to filter on this. Secondly, it was needed to filter on events that contained the string ‘was set as the active release’, to ensure that only these events were included, and not e.g., deleted releases.

To find the right balance for the number of events in the data set, the data was evaluated after filtering. After filtering, the dataset contained 19 412 events. No specific guidelines with respect to the number of events could be found in the literature. Based on our own interpretation, this number of events seemed sufficient, taking into account that the dataset contains 3 types of activities (test, acceptance, production).

Part of the extracted dataset is shown in *Figure 17 Data Extract Case Study 1*, where the case ID, activity and timestamp are indicated.




| Case ID | | | Activity | Timestamp |
|---|---|---|---|---|
| Company U | 268161465 | version 0.2.199 | accp | 25-8-2021 09:54 |
| Company U | 268161465 | version 0.2.198 | accp | 25-8-2021 09:47 |
| Company U | 268161567 | version 0.0.686 | test | 25-8-2021 06:36 |
| Company U | 268161567 | version 0.0.700 | accp | 24-8-2021 10:10 |
| Company U | 268158766 | version 1.0.520 | test | 24-8-2021 09:52 |
| Company U | 268173566 | version 0.0.28 | accp | 16-2-2022 10:56 |
| Company U | 268161465 | version 0.1.341 | test | 16-2-2022 14:53 |
| Company U | 268161465 | version 0.2.567 | accp | 16-2-2022 14:55 |
| Company U | 268161465 | version 0.2.191 | accp | 19-8-2021 18:49 |
| Company U | 268161465 | version 0.2.569 | accp | 17-2-2022 07:59 |
| Company V | 268170965 | version 0.2.18 | test | 6-7-2021 11:51 |
| Company V | 268170965 | version 2.1.0 | prod | 31-1-2022 13:55 |
| Company V | 268170965 | version 3.0.0 | accp | 31-1-2022 14:39 |
| Company V | 268170965 | version 3.1.6 | accp | 3-5-2022 09:33 |
| Company V | 268170965 | version 3.0.0 | prod | 31-1-2022 14:39 |
| Company V | 268170965 | version 0.2.43 | test | 21-10-2021 09:57 |
| Company V | 268170965 | version 0.2.45 | test | 26-10-2021 07:33 |
| Company V | 268170965 | version 3.0.1 | test | 31-1-2022 14:42 |

*Figure 17 Data Extract Case Study 1*

### 6.1.5 Phase 5: Process Model Creation

For the creation of process models, the Celonis tool was used, which makes use of the Fuzzy Miner. This means that the algorithm with which process models are built is not based on a specific mathematical model. However, since the goal is to gain quick insights, and not to thoroughly understand the mathematical explanation behind the models, this seemed to be alright for this case study. Moreover, Celonis provides several useful functionalities, such as the possibility to include all process variants in one model, to filter on process variants, to analyse throughput times and to create dashboards. Therefore, the models from Celonis will be presented here and will be used for evaluation in next phases.

The Celonis model based on case frequency, with all variants, is given in *Figure 18 Process Model Case Study 1 (case frequency)*. This is a relevant model, because it shows how many times a release is activated on an environment, and to which environment a release is promoted. It shows many different release versions, which is the goal, because the pattern in which releases are done should be shown. Celonis models based on throughput time are given in *Figure 19 Process Model Case Study 1 (throughput time - median)*, *Figure 20 Process Model Case Study 1 (throughput time - average)*, *and Figure 21 Process Model Case Study 1 (throughput time - trimmed mean)*. The trimmed mean is the mean of the rows that remain after trimming off 5% of the top rows, and 5% of the bottom rows (https://docs.celonis.com/en/trimmed_mean.html). The figures with respect to the throughput time are relevant to see how long it takes before a release is activated on a different environment.

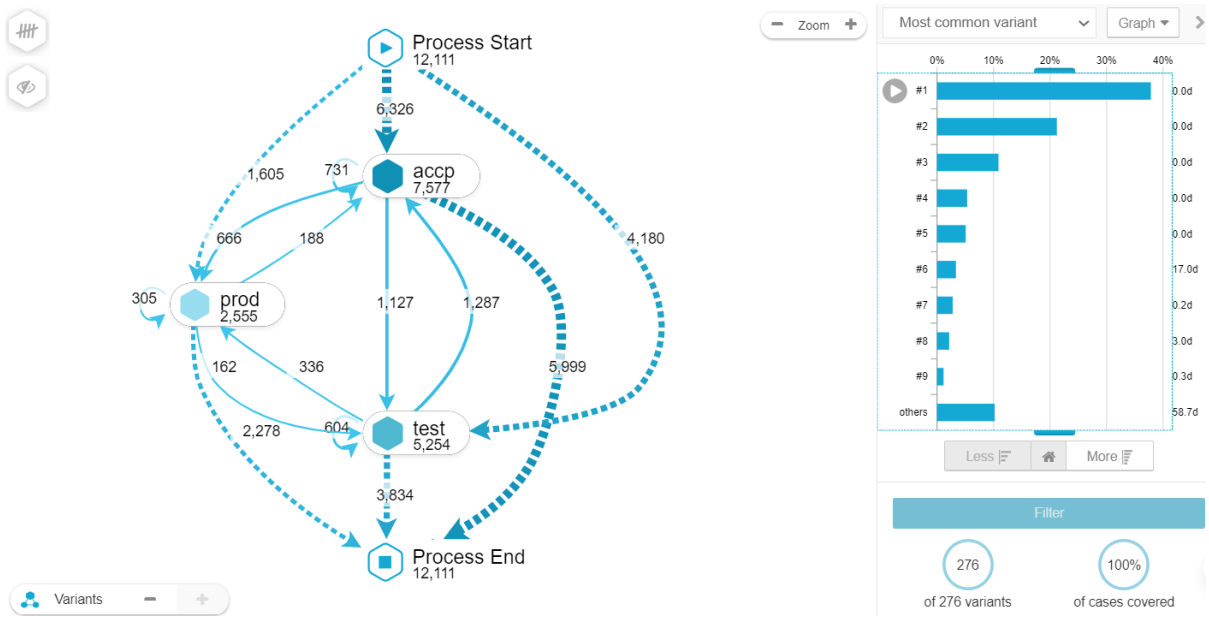



*Figure 18 Process Model Case Study 1 (case frequency)*

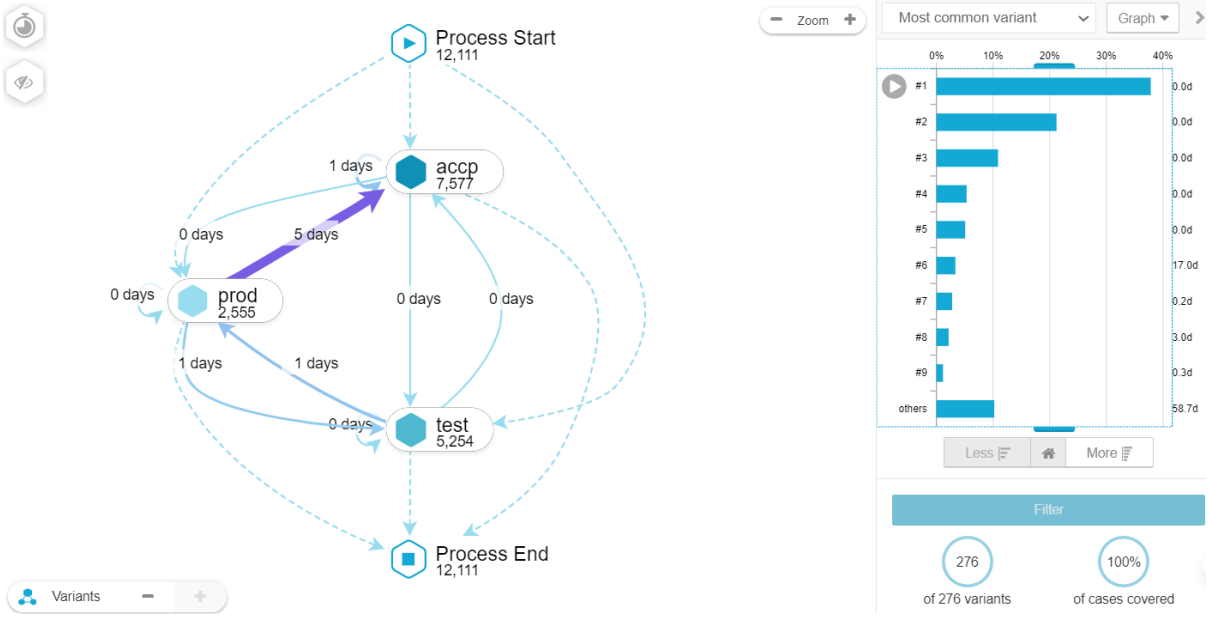



*Figure 19 Process Model Case Study 1 (throughput time - median)*

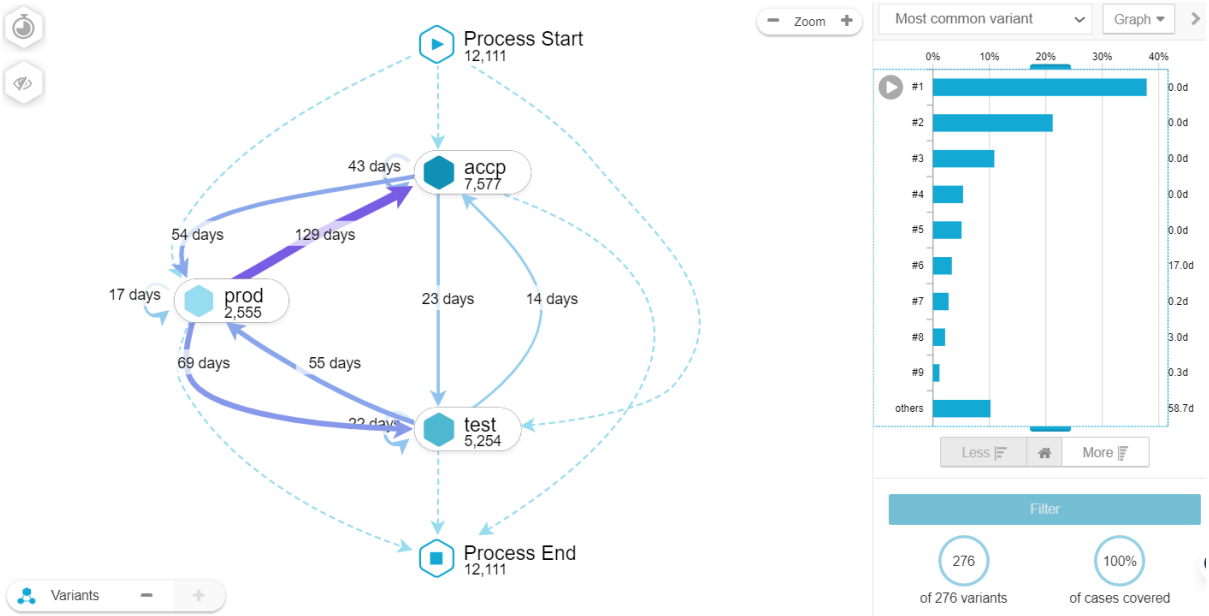



*Figure 20 Process Model Case Study 1 (throughput time - average)*

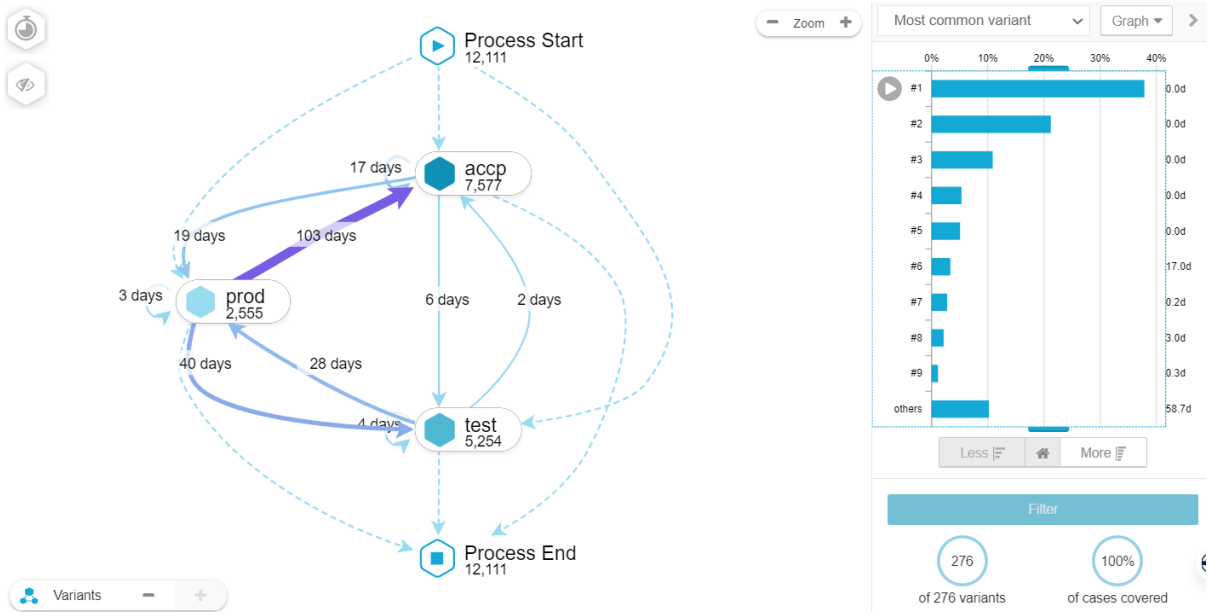



*Figure 21 Process Model Case Study 1 (throughput time - trimmed mean)*

The level that is shown in the figures is the operational level, since the figures depict how releases are made. Behind the release process is a strategy however, which concerns the order in which releases are promoted to another environment. The desired order is that a release is made in the test environment, then promoted to the acceptance environment, and finally promoted to the production environment. The reference model is given in *Figure 22 Reference Model Case Study 1*.


*Figure 22 Reference Model Case Study 1*

This reference model was made with the conformance checking functionality in Celonis. The results from conformance checking can be found in *Figure 23 Conformance Checking Case Study 1*.

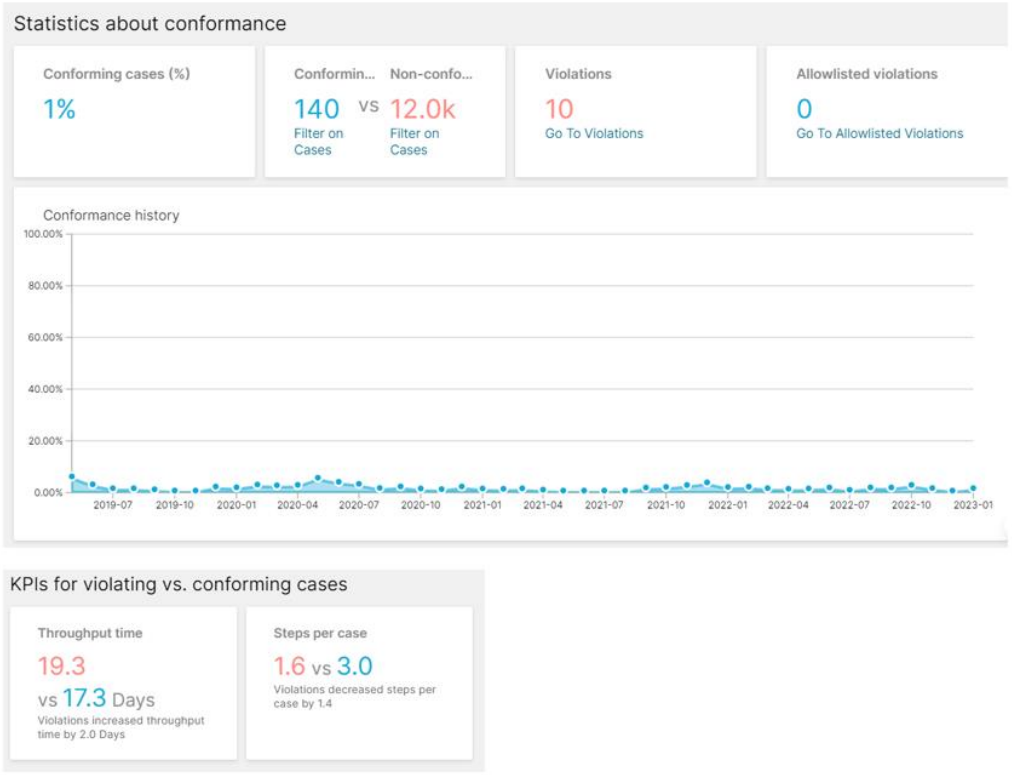



*Figure 23 Conformance Checking Case Study 1*

To give a clear overview of all the findings, a dashboard was created in Celonis. *Figure 24 Dashboard Impression Case Study 1* gives an impression of the dashboard. While dashboards are very useful for monitoring with the use of live data, dashboards can also be created to give additional insights if live data is unavailable. This is the case here, because the event logs are historic, their dates running from January 2019 till January 2023.

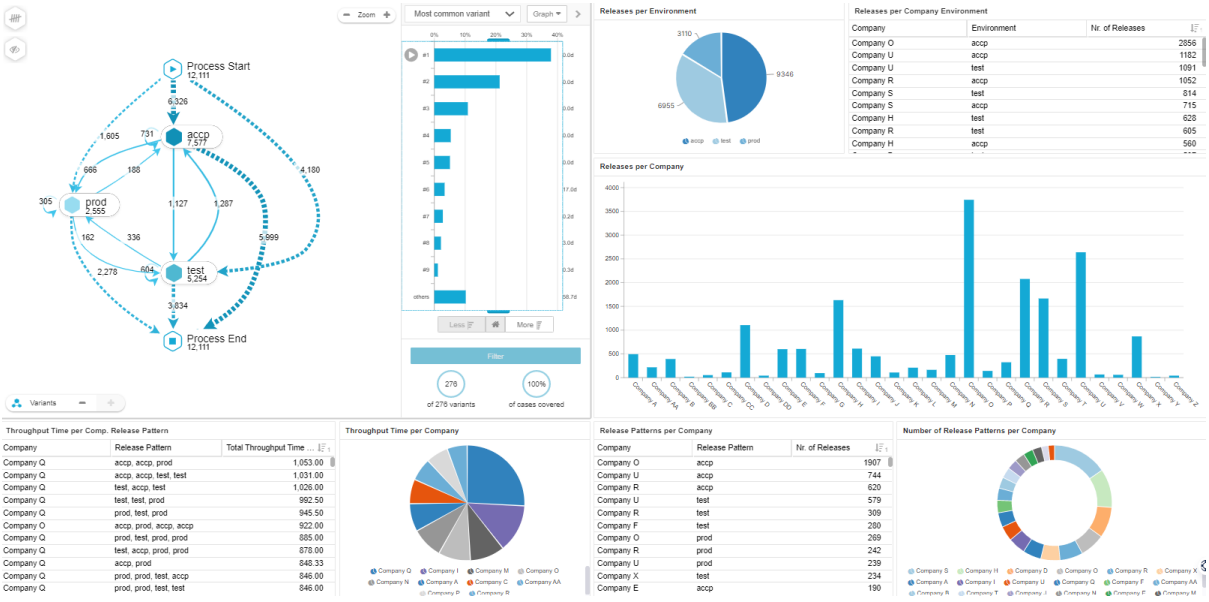



*Figure 24 Dashboard Impression Case Study 1 (text is not readable for privacy reasons)*

### 6.1.6 Phase 6: Analysis & Enhancement

The findings of the PM activities were evaluated with several process experts. Below, the findings are reported.

As can be seen in *Figure 22 Reference Model Case Study 1*, the desired process is that a release is created on the test environment, then promoted to the acceptance environment, and finally promoted to the production environment. However, this is only the 10th most common variant, and covers only 1% of all cases. The most common variant is a release only being activated on acceptance. This variant covers 38% of the cases. After that, the most common variant is a release only being activated on test, followed by the variant of a release only being activated on production. Thus, it is most common to activate a release on only one environment, instead of promoting releases from one environment to another.

The most common variant where a release is promoted from one environment to another is from test to acceptance, which is part of the desired process model. However, after this, the most common variant is a promotion from acceptance to test. With respect to the variants in which three activities are presented, the most common variant is the desired order of activities, namely test, then acceptance, then production, which covers 140 out of 12.0k cases. The second most common variant containing three activities is acceptance, then test, then production, which covers 89 cases. Lastly, the order of activities acceptance, then production, then test is followed in 35 cases. It is clear from the results that the desired order of activities is not executed often.

With respect to conformance checking, several findings can be reported. First of all, as mentioned before, the percentage of conforming cases is only 1%, which corresponds with 140 cases out of the total of 12.0k cases. Secondly, in 52% of the cases, a release is first activated on acceptance, which should be test. Thirdly, in 13% of the cases, a release is first activated on production. Fourthly, the number of steps per case for violating cases is 1.6 versus 3.0 for conforming cases. This is to be expected, because the most common variants were releases activated on a single environment. The throughput time of violating cases is 19.3 days as compared to 17.3 days for conforming cases. This is surprising, because the average number of steps per case for violating cases is lower.

With respect to the throughput time, it appears that it takes longest to promote a release from production to acceptance. This was the case for the median, average, and trimmed average throughput time. This order of promotion is not part of the desired order, which may explain the increased throughput time for violating cases as mentioned above. After that, reactivating a release on acceptance and promoting a release from production to test, as well as promoting a release from test to production appear to take much time. These are not part of the desired order of activities either. With respect to the desired order of activities (test → accp → prod), the median throughput time shows that this should take 0 days. Looking at the figures for mean throughput time and trimmed mean throughput time, it should take 68 and 21 days, respectively, where promoting a release from acceptance to production takes longest in both cases.

A surprising finding is that sometimes releases are reactivated on an environment. For example, in 3% of the cases, a release is activated on acceptance, and then again activated on acceptance. Based on the data, no clear explanation could be found, and no immediate explanation could be given by the process experts.

In *Figure 24 Dashboard Impression Case Study 1*, an impression of the dashboard with the most important PM results can be viewed. The dashboard shows the process model, where it is possible to show all types of models as given in *Phase 5: Process Model Creation*. The

dashboard also shows the releases per company, environment, and company environment. Moreover, the release patterns per company, as well as the number of release patterns per company is shown. Lastly, the throughput times per company and the throughput times per company release pattern are given. The dashboard in Celonis is interactive, so when hovering over it or when clicking on certain components, more or different information can be viewed.

### 6.1.7 Phase 7: Evaluation

In this phase, the research questions, as formulated in *Phase 2: Plan, Scope & Define* will be answered.

*TRQ1: In which order do customers activate releases with respect to the test, acceptance, and production environment?*

Customers mainly activate releases on a single environment, without promoting releases to other environments. The main environment on which customers activate releases is acceptance (38%), followed by test (21%), and then production (11%). After that, the most common variant is a release activated on test and then promoted to acceptance (5%). The most common variant consisting of three activities is a release activated on test, promoted to acceptance, and then promoted to production (1%).

*TRQ2: How does the order in which customers activate releases differ from the desired order?*

The desired order for customers to activate releases is to activate a release on test, promote it to acceptance, and then promote it to production. This order of activities is only performed in 1% of the cases, namely 140 out of 12.0k cases. It is the 10$^{th}$ most common variant. While it is desired to start with a release on the test environment, in 52% of the cases, a release is first activated on the acceptance environment. Moreover, releases are mainly activated on only one environment, while the desired release process consists of three activities.

*TRQ3: On which environment do customers mainly activate releases?*

Customers mainly activate releases on acceptance, 9 346 releases have been activated on acceptance. This is followed by test, on which 6 955 releases have been activated. The number of releases activated on production is 3 110. This adds up to a total number of releases of 19 411. Note that this total number is greater than the 12.0k that was mentioned before, because releases can be reactivated on an environment.

*TRQ4: What are the release patterns for each customer?*

All customers have several different release patterns. The customer with most releases is Company O, with 3 743 releases. The number of release patterns that belongs to this customer is 51. After that, Company U produces most releases, followed by Company R, with 2 639 and 2 077 releases respectively. The number of release patterns for these customers are 30 and 47, respectively. So, while Company U produces more releases, Company R produces more release patterns. The company with the highest number of release patterns is Company S, with 113 patterns. The number of releases from Company S is 1 665. The highest number of release patterns is followed by Company H with 79 patterns, and Company D with 63 patterns, with 1 631 and 1 105 releases, respectively.

The customer with the lowest number of releases is Company Y, with 15 releases. The company also has the lowest number of release patterns, namely 3. After that, Company BB has the lowest number of releases, namely 19, with 4 release patterns. This is followed by Company Z and Company DD, both with 42 releases, and both 10 release patterns.

The most common release pattern differs per company, but overall, the most common release pattern is a single release on acceptance, as mentioned before.

*TRQ5: Which steps, patterns, and customers yield a high throughput time?*

The step that takes most time is the promotion of a release from production to acceptance. This was the case for the median, average, and trimmed average throughput time. This is followed by reactivating a release on acceptance, promoting a release from production to test, and promoting a release from test to acceptance. The release pattern that has the highest total throughput time, taking into consideration all companies, is prod, test, prod, prod. Per company, the release pattern with the highest total throughput time is accp, accp, prod, from Company Q. This company holds 11 release patterns out of the 12 release patterns that have the highest total throughput time. Logically, this company holds first place in having the highest total throughput time. The company with the lowest total throughput time is Company BB.

To verify the PM results, three checks may be executed, as formulated by [33], p, 42:

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

All three checks could be verified for the Celonis models. Note that the frequencies here are the number of cases, and not the number of flows. E.g., accp has 6 326 + 1 287 + 188 + 731 = 8 532 incoming cases, and 666 + 1 127 + 5 999 + 731 = 8 523 outcoming cases, and thus the same frequency.

To validate the PM results, the degree of fitness, precision, generalization, and simplicity of the model should be checked. The Celonis models were constructed using the Fuzzy Miner, which aims to balance these four criteria [4]. No cases were excluded from the event log that should have been included. The only filters that were applied were to filter out non-customers and to only include the activity of activating a release. Moreover, looking at the process models, they allow for all possible patterns. Thus, the degree of fitness is very good. While the variant explorer process model shows all possible behaviour, the model is still precise. The reason for this, is that the number of cases is shown for each step or activity. With respect to the degree of generalization, the different release patterns show how the eMagiz environment is used to activate releases. So, the data does not only describe the data, but also shows the usage patterns of customers. Lastly, the model is quite simplistic, because only three types of activities are identified. However, this does mean that the model is very understandable, and also very complete, because all events could be used to make a model. Moreover, the model does not appear to be too simplistic, because useful results could be obtained. In conclusion, the models could be validated against all four criteria.

The PM project objective was to find out whether PM could be useful for the company. This project objective has been met, because several important findings were done. Moreover, all research questions for this project have been answered. The most important finding is that releases are not activated according to the eMagiz best practice. First of all, it appears that the acceptance environment is used as a test environment. The reason for this, is that many releases are activated on acceptance, while much less releases are activated on test, even less than on production. Secondly, all possible patterns between the three activities are executed by almost all companies. Thus, it appears that no particular pattern to activate releases is followed by any company.

With respect to the throughput time, it is clear that some companies have a much higher throughput time than other companies. This high amount of throughput time does not seem to have a clear correlation with the number of releases or release patterns per company. Next to this, it was found that promoting a release from production to acceptance, which is not best practice, yields the highest throughput time in general.

Improvement actions will be discussed in the next phase.

### 6.1.8 Phase 8: Process Improvement & Presenting

Several improvement actions can be defined based on the findings. First of all, it would be interesting to see why the test environment is generally not used as a test environment at the moment. Secondly, it should be investigated why the best practice of making releases is not followed at the moment. Thirdly, it should be evaluated whether certain companies do make use of a particular release pattern, and why. The answers to these findings could help to improve the eMagiz platform. For example, if certain functionalities are missing from the test environment, which cause the user to not use it, these functionalities may be added. Moreover, if it appears that the test environment is not useful, the test environment could potentially be removed. Another aspect that should be investigated is releases being activated on one environment several times. For example, a common pattern is that releases are activated multiple times on the acceptance environment. With respect to the findings regarding throughput time, it would be interesting to check which companies have the highest throughput time, and why. It is likely that there is some correlation between the throughput time and the efficiency of teams. This could be investigated.

In *Table 12 Findings, Actions Points, Stakeholders & Benefits Case Study 1*, the findings, corresponding action points, responsible stakeholder(s), and potential benefits are summarized.

*Table 12 Findings, Actions Points, Stakeholders & Benefits Case Study 1*

| Finding | Action Point | Responsible Stakeholder(s) | Benefit |
|---|---|---|---|
| The best practice of making releases is not being followed. | Investigate why the best practice is not followed, e.g., through user interviews. | Product Owner & UI/UX Expert | Adaptations to the eMagiz platform could be made, such that it yields a better user experience for customers.<br><br>Customers could be guided in following the best practice, such that integrations can be created more effectively and efficiently. |
| The acceptance environment seems to be used as a test environment, instead of the designated test environment. | Investigate why users do not use the test environment much, e.g., through user interviews. | Product Owner & UI/UX Expert | Adaptations to the eMagiz platform could be made, such that the test environment is better usable, or |

| | | | |
|---|---|---|---|
| | | | potentially could be removed. |
| Differences with respect to release patterns among different customers exist. | Investigate whether customers try to follow a specific release pattern and why, e.g., through user interviews & further studying the data. | Expert Services Team | A best practice that is better achievable may be found.<br><br>Some customers may be guided to better understand their release patterns, such that integrations can be created more effectively and efficiently. |
| In some cases, releases are reactivated on a single environment. | Investigate why releases are reactivated on a single environment, e.g., through further studying the (raw) data & user interviews. | Product Owner & UI/UX Expert | Remove reactivated releases from the data, because they might disturb the data.<br><br>Resolve a potential bug that leads to releases being activated multiple times. |
| Companies appear to have large differences in throughput time. | Investigate where the large throughput time comes from, e.g., through studying the data further & user interviews. | Product Owner & UI/UX Expert | Some companies may be guided to create integrations more effectively and efficiently. |

Next to activities specific for the release process, several other improvement actions can be formulated. First of all, several ILM processes were analysed at the start of the project. These processes, or other ILM processes, could be investigated further. For this, it may be needed to store additional data, because e.g., timestamps were missing. Moreover, event logs may be integrated in such a way that ILM data can be extracted. This means that a system could be set up to automatically generate event logs that are suitable for PM. This way, it could also be possible to mine current data, which is great for monitoring. Next to ILM processes, several other processes may be identified, and investigated further. For those processes, it might also be needed to store additional information. To ease the extraction of event logs of different processes, a centralized portal could be setup, through which event logs can be distributed.

For this research, with the purpose of validating the designed methodology further, a different process will be studied. This will be explained in the next phase. The prioritization and refinement of improvement actions will be done by the eMagiz team and will thus not further be described in this study.

The final results were presented to the eMagiz employees during a knowledge share. Employees that collaborated on the project were updated during the execution of the phases.

### 6.1.9 Phase 9: Change Management

To ensure that PM is a continuous activity in the organization, the improvement actions as defined in the previous phase should be executed. This will be done by the eMagiz team, as mentioned in the previous phase. To ensure continuity, other processes relevant for eMagiz may be investigated as well. This will be part of this study, since it will also ensure that the methodology can be validated. The process that will be studied is the ticket process from support. This will be explained in Chapter *7 Treatment Validation*.

With respect to change management, existing practices can be followed for this. eMagiz is a company that implements many changes and follows certain guidelines for this. For example, complete product version migrations are done. Due to this experience, no new change management practices are formulated at the moment.

## 6.2 Refinement

Based on the findings from applying PROMISE in the case study, several conclusions can be drawn. First of all, the guidelines gave some guidance, but the amount of guidance was not as desired, especially in the first five phases. It would be clearer if specific steps were given for all required activities. Secondly, the order in which the guidelines were given differed in practice. Several activities were executed in different phases. To resolve this, the order of the activities, and in which phases the activities are categorized should be reconsidered. Thirdly, the descriptions of the phases and the explanations of the guidelines were not given in the visualization. It would be good if an overview of the main activities for each phase could be given in a (separate) visualization. Lastly, it should be possible to end the PM activities, e.g., if resources become unavailable, and it should be possible to enter the Business Understanding phase again if a change in the business has occurred which would cause the Business Understanding to change, e.g., a change in the strategy of the company.

Based on these findings, first of all, the guidelines were transformed into steps, and the order of the steps was changed. *Figure 25 Refined PROMISE Visualization* shows the refined visualization. For clarity purposes, this visualization has been enhanced. The enhanced visualization can be found in *Figure 26 Refined PROMISE Visualization Enhanced*. This enhanced visualization includes two additional arrows: End of PM Activities, representing the ending of the PM activities, and Business Change, representing the re-entering of the Business Understanding phase.

Apart from these adaptations, an additional cycle has been added to the visualization, *Figure 26 Refined PROMISE Visualization Enhanced*. This cycle represents the three types of PM techniques: discovery, conformance, and enhancement. The reason for the PM techniques having been added to the visualization, is that this could help practitioners to decide on which PM technique(s) to use, and it clarifies in which phase a particular PM technique is emphasized. Instead of following all phases, it is possible to follow only part of the phase, e.g., by only using discovery techniques.

The specific steps that have to be executed in each phase of the methodology are not displayed in the visualization anymore but are given in *Figure 27 PROMISE Pillars*. These pillars explain the main activities and deliverables per phase, as well as the needed PM steps for each phase. In Appendix *E PROMISE Steps*, all steps are given.

Important to keep in mind is that (i) findings should be verified after each phase, (ii) previous phases should be executed again if required, and (iii) phases should be completed quickly to present an MVP as soon as possible. These findings also held for the initial methodology. It was chosen not to display these findings in the PROMISE visualization for clarity purposes.

## 6.3 Summary

In this chapter, PROMISE was applied in a case study performed at eMagiz. For each phase, the findings were described. It was found that the order of the guidelines formulated in the previous chapter differs in practice. Moreover, the methodology did not yet give the desired amount of guidance. Therefore, PROMISE was refined, where the guidelines were transformed into steps, and the order of the steps was changed. Moreover, the PROMISE visualization was enhanced, and PM techniques were added to the model. Lastly, methodology pillars were created that contain a description of each phase, including the steps that have to be executed, as well as the deliverables for each phase.

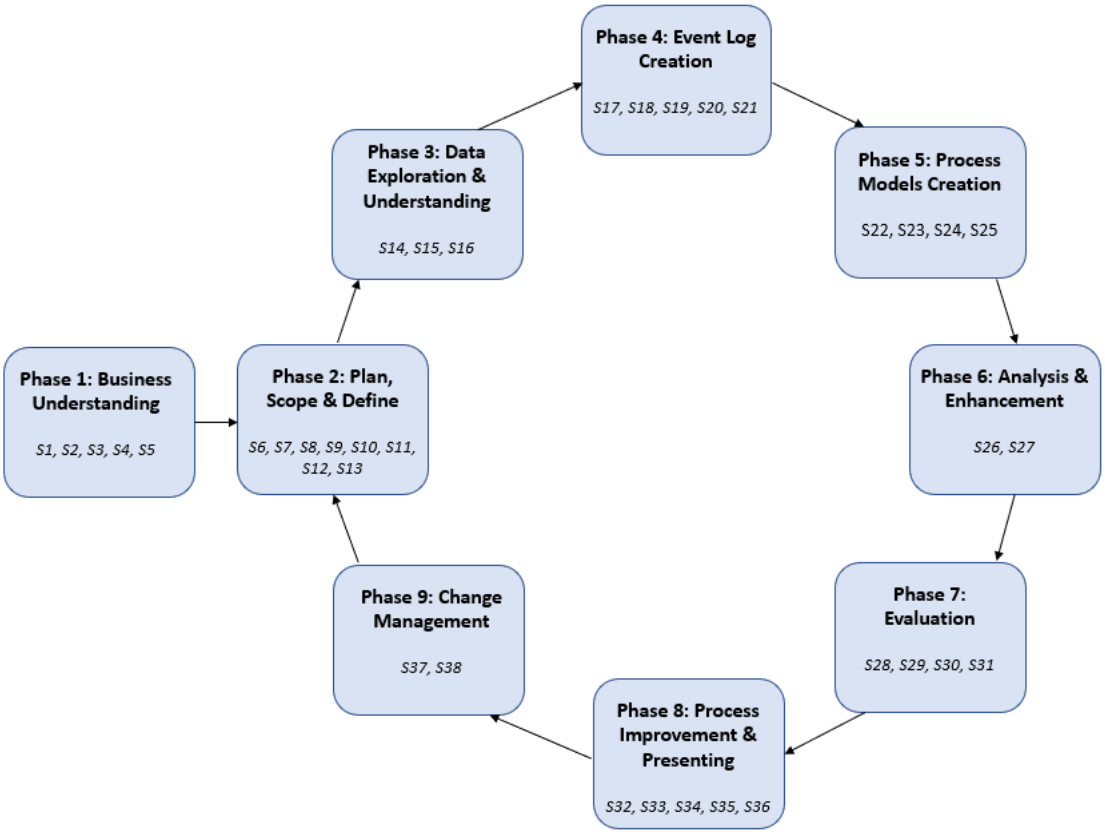



*Figure 25 Refined PROMISE Visualization*

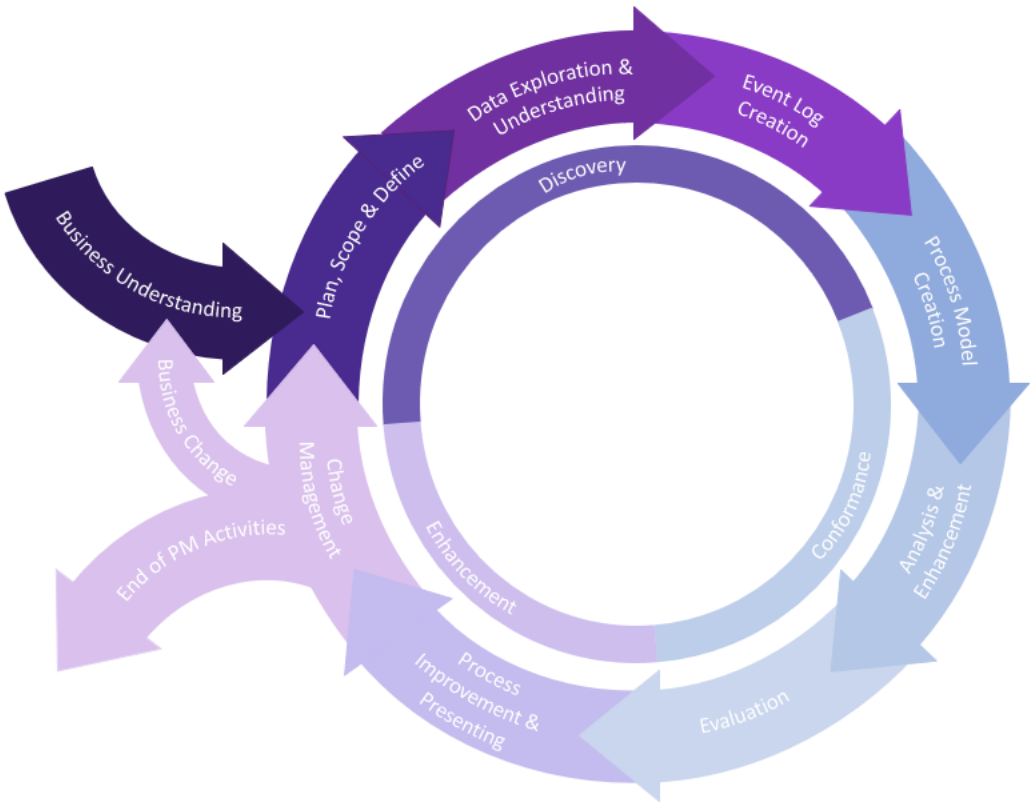



*Figure 26 Refined PROMISE Visualization Enhanced*

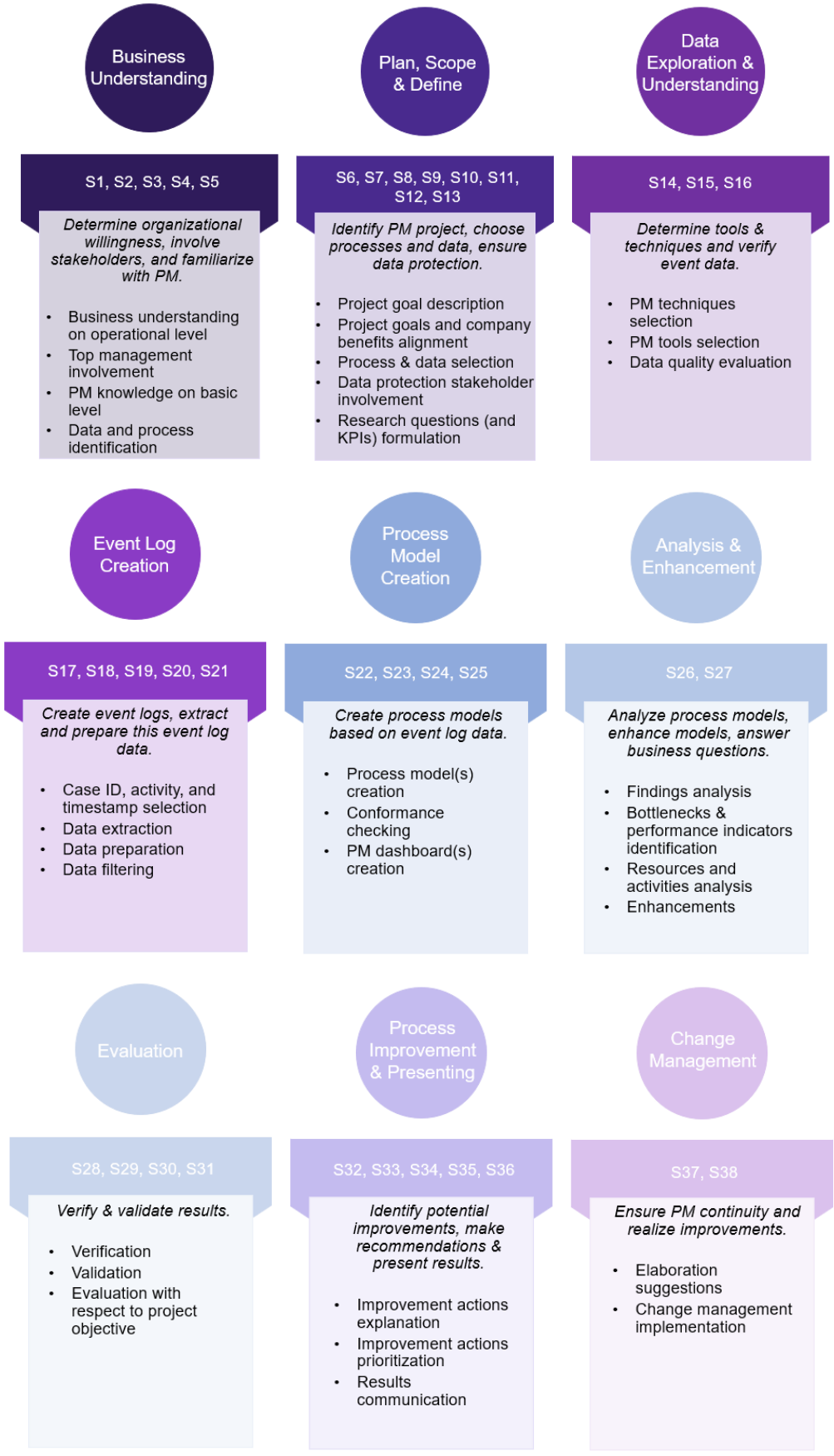



*Figure 27 PROMISE Pillars*

# 7 Treatment Validation

To validate the (refined) methodology, it was applied in another case study at eMagiz. The results of applying PROMISE are described below. After applying the methodology, expert and practitioner evaluations were obtained. Moreover, it was checked whether the requirements defined in Section *5.1 Methodology Requirements* have been met.

## 7.1 Results per Phase

### 7.1.1 Business Understanding

Customer Support is a shared service for eMagiz and their partner CAPE, which share a building. Support provides, as is clear from the name, support to eMagiz and CAPE. Such support ranges from helping customers with particular issues to helping employees setting up their work environment.

About 5 to 10 people work for Support, which meant that it was not difficult to ensure them to have a basic understanding of PM. The Support Process Expert could provide a process that was desired to mine, including the relevant data for it. Moreover, the Support Process Expert was willing to help analyse this process during the required PM activities. The process that could be mined was the ticket process, which is about the registration of support actions. When an action has to be undertaken for a customer, e.g., when an incident happens in the eMagiz environment of a customer, this is registered using tickets. It was also evaluated whether other processes could be analysed, but no data for other processes was available, and the ticket process seemed significant and useful to mine.

In *Table 13 Stakeholders Case Study 2*, the stakeholders needed for the case study have been identified, according to the stakeholders and goals defined in *Table 5 Stakeholder Goals*.

*Table 13 Stakeholders Case Study 2*

| STAKEHOLDER | CASE STUDY STAKEHOLDER |
|---|---|
| MANAGEMENT | Team Lead Support |
| ARCHITECT | Support Process Expert |
| PROCESS EXPERT | Support Process Expert |
| PM TEAM | Student & Supervisors |
| PM EXPERT | First UT Supervisor |
| MAINTENANCE TEAM | Support Team |
| DEVELOPMENT TEAM | Support Team |
| DATA PROTECTION STAKEHOLDER | Information Security Officer |
| CEO | Team Lead Support |

### 7.1.2 Plan, Scope & Define

The goal of this PM project is to gain more insight into the support ticket process. The ticket process starts with a new ticket being created, meaning that a new activity is registered, and it ends once the ticket is completed, meaning that the activity has been completed. Investigating this process can have several benefits. First of all, it could be helpful in understanding how tickets are registered at the moment. Currently, several approaches to registering tickets can be taken, no best practice for this is in place. Therefore, it would be interesting to see whether a certain pattern to register tickets is being followed. Secondly, based on the discovered process model, a new process model could be determined, which could be seen as a new best practice standard for the registration of tickets. This could guide employees in their ticket registration process and could ensure that tickets are registered in a more effective or logical way. Lastly, several important findings with respect to customers

may be done. For example, it would be interesting to see which customer has most ticket registrations, in other words, which customer required most supporting activities.

A case ID, activity, and timestamp could easily be selected, no ambiguity arose. Moreover, the data could quickly be extracted from CAPE Service Point, the system that is used to register tickets, which was done by Support. Apart from the case ID, activity, and timestamp, several additional attributes were retrieved. The reason for this, is that additional properties should be investigated, such as the company to which a ticket belongs. With respect to the number of activities in the dataset, 18 types of activities are present. At the moment, it is not known how many types of activities the desired process should contain.

Based on the extracted data and the interest of Support in the ticket process, it is likely that the company is able to give enough data and commitment. A categorization of the people involved in, and responsible for, the PM activities has been given in *Table 13 Stakeholders Case Study 2*, which also includes data protection stakeholders. The same data protection rules as for the first case study were in place. With respect to the transparent data policy, it was agreed that data could be mined using anonymized names for the employees as well as for the companies.

The following questions were formulated:

1. What is the most common variant of activities in the ticket process?
2. Which steps, projects, and companies yield a high throughput time?
3. Are tickets solved and closed by different employees?
4. Which project, service, and company holds the highest number of tickets?
5. What could be a desired ticket process?
6. How does the current ticket process differ from the desired ticket process?

### 7.1.3 Data Exploration & Understanding

The types of PM techniques that will be used are discovery, conformance, and enhancement. Discovery techniques will be used to answer the first four questions. Conformance checking, in combination with enhancement techniques, will be done for the fifth and sixth research question. The reason for this combination of conformance and enhancement is that no ticket process model exists yet. Therefore, a desired process model is made first, which is an enhancement technique, and is then checked with the current ticket process, which is conformance checking. Enhancement will also be done through defining improvement actions.

In *Table 14 Tools & Techniques Case Study 2*, an overview of the tools and techniques that will be used is given.

*Table 14 Tools & Techniques Case Study 2*

| **Tools** | **Techniques** | **Purpose** |
|---|---|---|
| Data query tool<br>*(CAPE Service Point)* | Query language (*system specific*) | Extract data & create datasets<br>*(CSV)* |
| Data visualization & analysis tool<br>*(Excel)* | No specific technique needed | View data & transform datasets |
| PM tool<br>*(Celonis)* | Discovery algorithms, dashboarding<br>*(Fuzzy Miner)* | Create process models & dashboards |

To check whether the data quality was sufficient, the data had to be evaluated with respect to the trustworthiness, completeness, semantics, and safeness of the data. To evaluate the quality, several quality requirements were formulated, which are given in *Table 15 Quality Aspects & Requirements Case Study 2*.

*Table 15 Quality Aspects & Requirements Case Study 2*

| Quality Aspect | Requirement |
|---|---|
| Trustworthiness | The ticket activities should represent real activities, meaning that they should actually have happened and are correct, e.g., with respect to when a ticket was completed. |
| Completeness | The ticket information should be complete, meaning that no ticket activities should be removed. |
| Semantics | It should be clear what the meaning of the data is. I.e., the meaning of the case IDs, events, and timestamps should be evident. |
| Safeness | Privacy and security concerns should be addressed when recording the events. |

The data quality requirements were evaluated with the Support Process Expert, and it was found that the data quality was good on all aspects. With respect to the safeness, privacy concerns will be addressed by anonymizing names of employees.

### 7.1.4 Event Log Creation

In *Table 16 Attributes Case Study 2*, the chosen case ID, activity, and timestamp are described. Several other attributes that are needed to answer the research questions or that could turn out to be useful were also stored. These are given in *Table 16 Attributes Case Study 2* as well.

*Table 16 Attributes Case Study 2*

| Type of Attribute | Notation | Description |
|---|---|---|
| Case ID | ticketnumber | When a supporting action for a company has to be undertaken, this is registered using tickets, where each ticket has a unique ticket number and represents a unique supporting activity. |
| Activity | status | The status of a supporting action is recorded as the status of a ticket. Examples of statuses are new, accepted, and closed. In total, 18 statuses were present in the dataset. |
| Timestamp | statusdate | The timestamp of an activity is the timestamp of a change in the recorded status, which represents a change in the status of a supporting activity. For example, a status may be changed from new to |

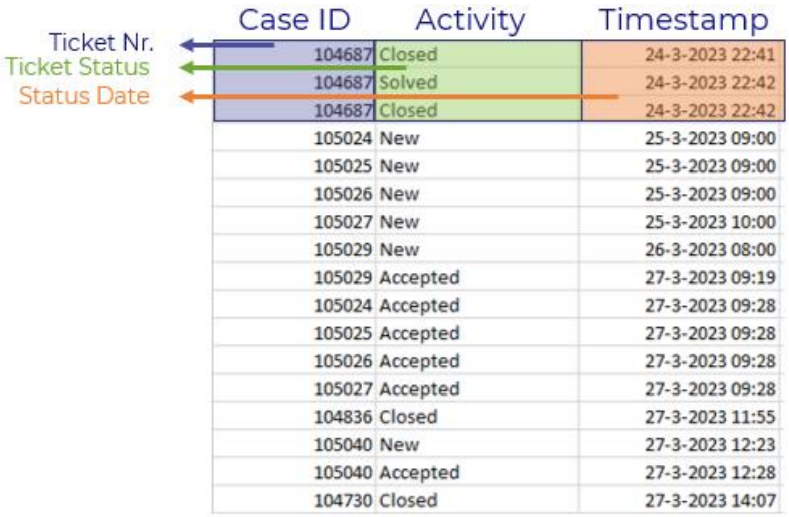
| | | |
|---|---|---|
| | | accepted, where this new timestamp would belong to the ‘accepted’ status. |
| Additional | emailaddress | Name of the employee who has changed the ticket status. |
| Additional | prioritylevel | Priority level of the activity, (1-5, from high to low). |
| Additional | servicename | Name of the service into which a ticket can be categorized, e.g., ‘incident management’, or ‘service alert’. |
| Additional | projectname | Name of the project to which a ticket belongs. |
| Additional | companyname | Name of the company to which a ticket belongs. |

In short, supporting actions for customers are registered using tickets, where each ticket, or action, has a unique ticket number. Each ticket can have multiple statuses, which represent the status of a supporting action, where each recorded status has a timestamp. Next to this, several properties are stored that give more information about the action, such as the type of project or company for which the action was performed.

The data was extracted by the Support Team. To ensure that the dataset was suitable, the names of the employees had to be anonymized. This was done by giving all of the employees a pseudonym. Moreover, several of the additional columns had missing values. The missing values of the (integer) column ‘priority level’ was given a 0. This is seemed to be a good value, because no priority level of 0 exists, meaning that the missing value are distinguishable. For all other additional (string) attributes, the missing values were replaced by the string ‘unknown’. The missing values were not filtered out of the dataset, because it was expected that this would give a more skewed image. Moreover, the attributes appeared to still be useful, because the missing values were distinguishable.

The number of events in the dataset was > 100 000, which is the limit for the Celonis tool. Therefore, several events had to be removed from the dataset. The dates from the dataset ranged from 2011 till 2023, so it was chosen to remove some of the oldest data. All data that belonged to a deleted ticket was removed as well. This meant that data from 2011, 2012, 2013, and partly 2014 was removed. Thus, the filtered dataset contained activities from the years 2014 till 2023. The number of events was 99 994. This number of events seemed sufficient; it should be possible to show many ticket registration processes. If the models appear to be too cluttered, it is possible to add filters in Celonis, e.g., to include only recent years.

Part of the extracted dataset is shown in *Figure 28 Data Extract Case Study 2*, where the case ID, activity and timestamp are indicated.




| Case ID | Activity | Timestamp |
|---|---|---|
| 104687 | Closed | 24-3-2023 22:41 |
| 104687 | Solved | 24-3-2023 22:42 |
| 104687 | Closed | 24-3-2023 22:42 |
| 105024 | New | 25-3-2023 09:00 |
| 105025 | New | 25-3-2023 09:00 |
| 105026 | New | 25-3-2023 09:00 |
| 105027 | New | 25-3-2023 10:00 |
| 105029 | New | 26-3-2023 08:00 |
| 105029 | Accepted | 27-3-2023 09:19 |
| 105024 | Accepted | 27-3-2023 09:28 |
| 105025 | Accepted | 27-3-2023 09:28 |
| 105026 | Accepted | 27-3-2023 09:28 |
| 105027 | Accepted | 27-3-2023 09:28 |
| 104836 | Closed | 27-3-2023 11:55 |
| 105040 | New | 27-3-2023 12:23 |
| 105040 | Accepted | 27-3-2023 12:28 |
| 104730 | Closed | 27-3-2023 14:07 |

*Figure 28 Data Extract Case Study 2*

### 7.1.5 Process Model Creation

Several process models were created in Celonis. *Figure 29 Process Model Case Study 2 (case frequency)* shows a process model based on case frequency, with the 12 most common variants. The reason for 12 variants being displayed, is that displaying more variants makes the model more cluttered, while the percentage of cases covered from 12 variants on only increases by 1% by adding a variant. Moreover, the current percentage of cases covered is 70%, so it should give a fair picture.

*Figure 30 Process Model Case Study 2 (throughput time - median)*, *Figure 31 Process Model Case Study 2 (throughput time - average)*, and *Figure 32 Process Model Case Study 2 (throughput time - trimmed mean)* show the process models based on throughput times, the median, average, and trimmed mean throughput times respectively. All contain the 12 most common variants.

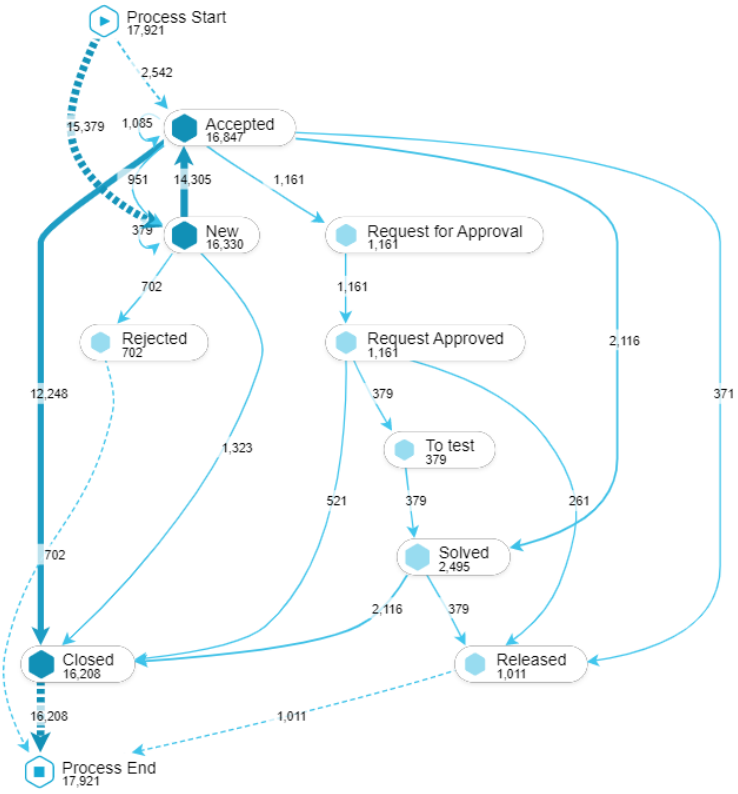



*Figure 29 Process Model Case Study 2 (case frequency)*

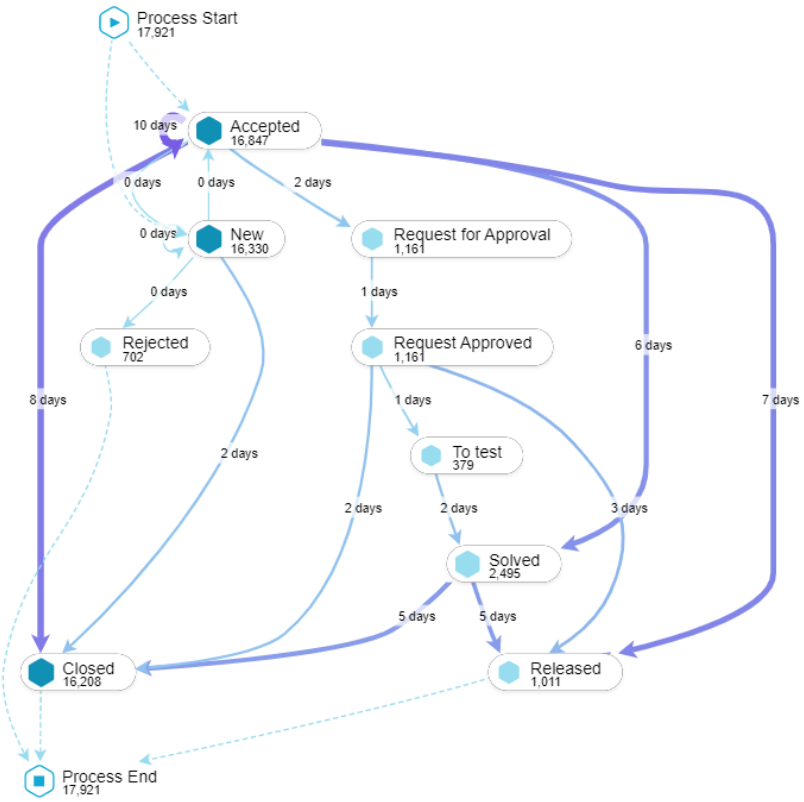



*Figure 30 Process Model Case Study 2 (throughput time - median)*

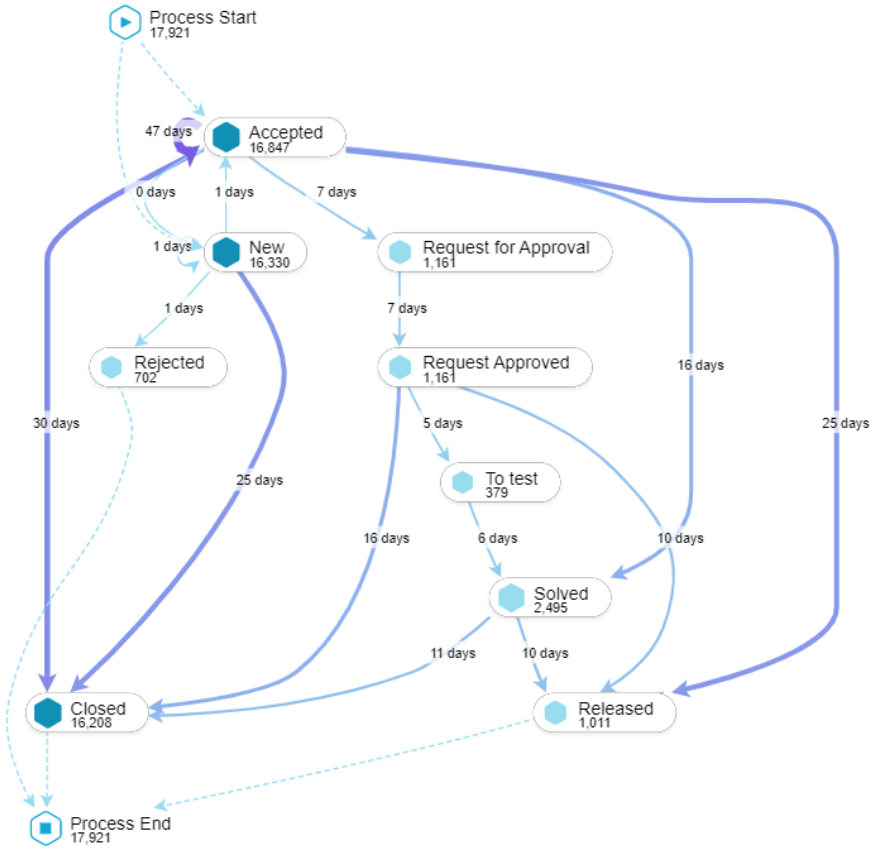



*Figure 31 Process Model Case Study 2 (throughput time - average)*

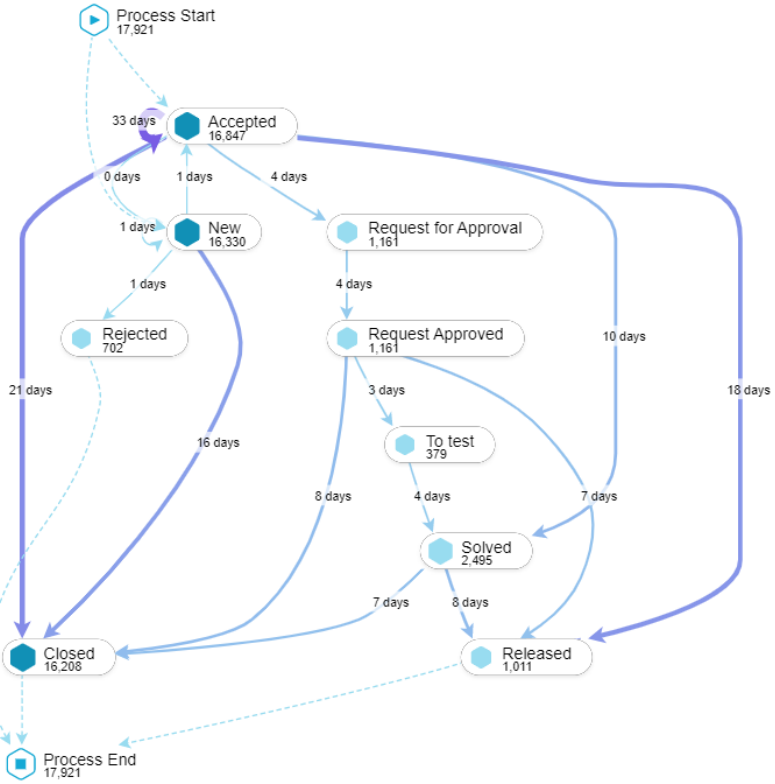



*Figure 32 Process Model Case Study 2 (throughput time - trimmed mean)*

Together with the Support Team, a desired process model was created in Celonis, given in *Figure 33 Reference Model Case Study 2*. Note that this process model only shows the desired activities, so e.g., the activity 'not solved' is not present in the model. In Figure, the desired process model is depicted.

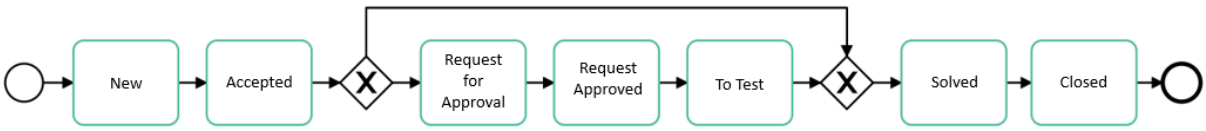



*Figure 33 Reference Model Case Study 2*

After a ticket has been accepted, either a request for approval to the company has to be made, or the ticket is worked on by Support, after which the next status is 'Solved'.

The results from conformance checking with this model can be found in *Figure 34 Conformance Checking Case Study 2*.

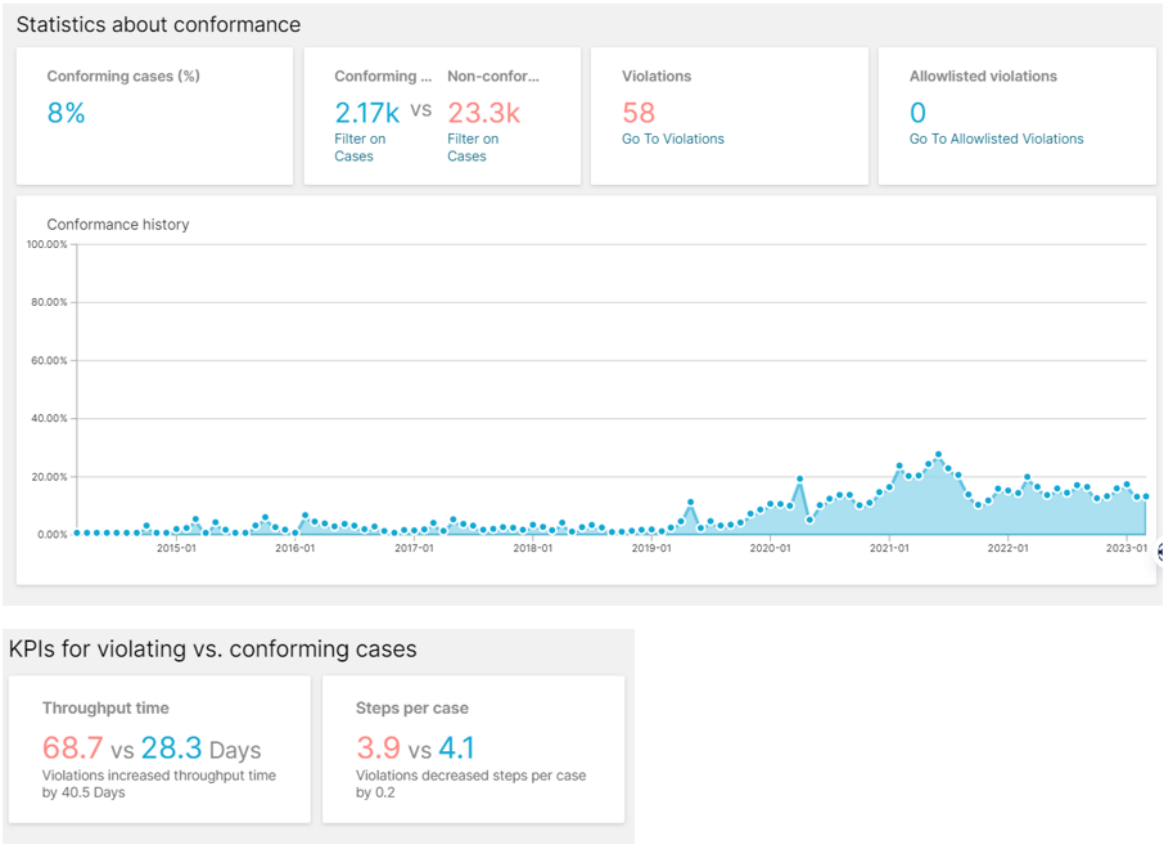



*Figure 34 Conformance Checking Case Study 2*

Next to this desired process model, a best practice process model was created with the Support Team, which also shows undesired, but realistic activities, such as 'Not Solved'. This model can be used by Support in the future. The model is given in *Figure 35 Best Practice Process Model Case Study 2*.

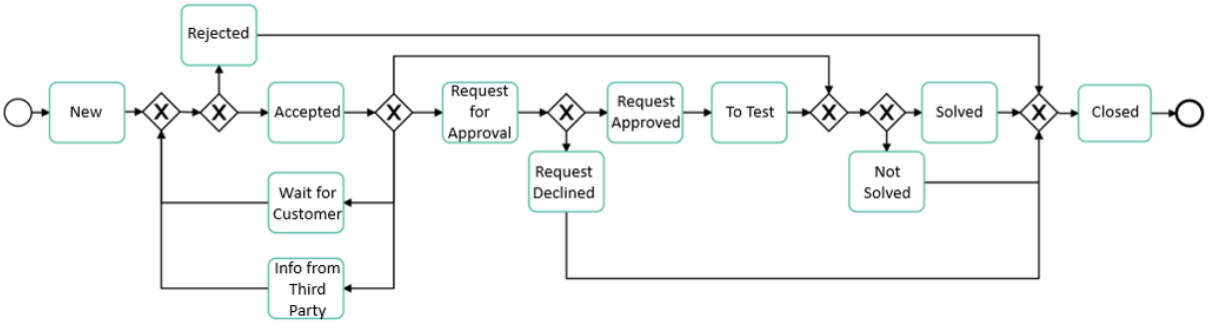



*Figure 35 Best Practice Process Model Case Study 2*

Two dashboards were created in Celonis to give a clear overview of all the findings. One dashboard, given in *Figure 36 Dashboard Impression Tickets Case Study 2*, gives an impression of the dashboard containing findings with respect to the number of tickets. An impression of the other dashboard, given in *Figure 37 Dashboard Impression Throughput Times Case Study 2*, contains findings with respect to throughput times.

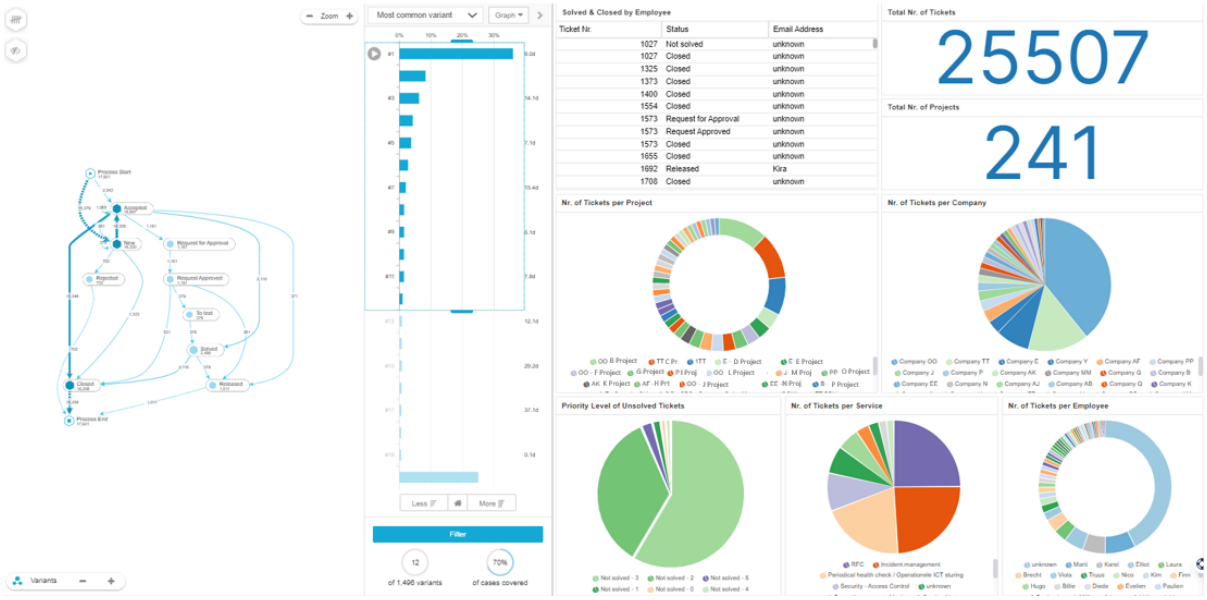



*Figure 36 Dashboard Impression Tickets Case Study 2 (text is not readable for privacy reasons)*

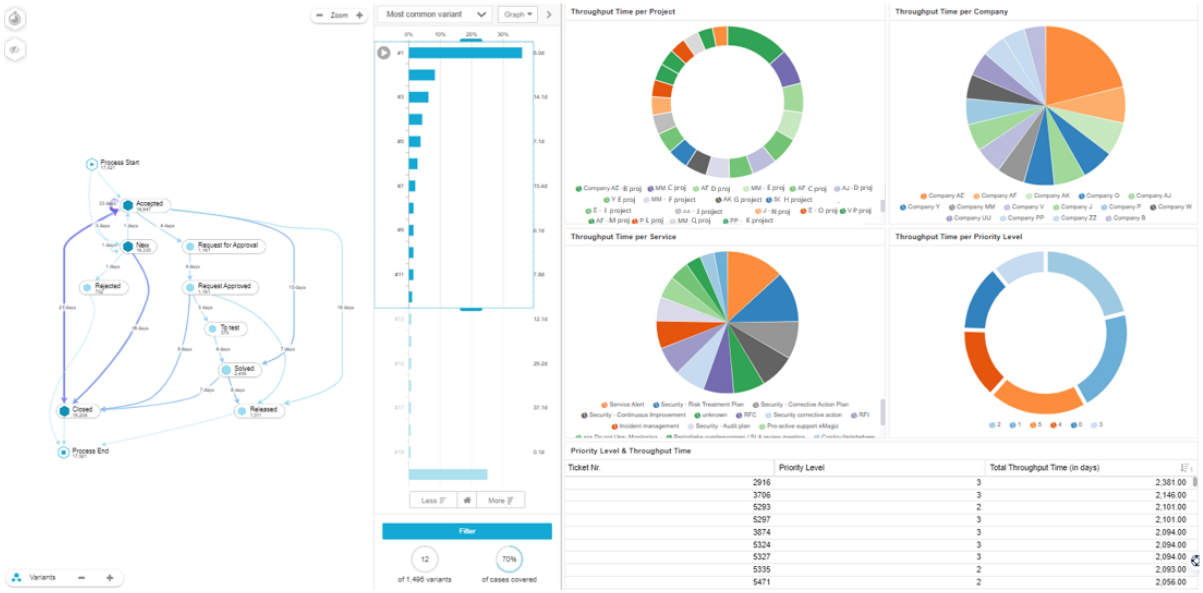



*Figure 37 Dashboard Impression Throughput Times Case Study 2 (text is not readable for privacy reasons)*

### 7.1.6 Analysis & Enhancement

The findings of the PM activities were evaluated with the Support Process Expert and another person from the Support Team. Below, the findings are reported.

The most common variant is a ticket being registered as New, then Accepted, and then Closed. This variant covers 36% of all cases. As can be seen in *Figure 33 Reference Model Case Study 2*, the variant is not part of the desired process, because the status Solved should be added. The second most common variant does follow the desired pattern, namely New, Accepted, Solved, Closed. This variant covers 8% of all cases. With conformance checking, it was also found that 8% of all cases is conforming to the desired process model, which corresponds with 2.17k cases versus 23.3k non-conforming cases. Based on this, it can be concluded that the desired process in which a request has to be made to a client is never followed, because the above-mentioned variant already covers all conforming cases. The most common variant in which a request is made to a client is the 7$^{th}$ most common variant, covers 2% of all cases, and does not conform to the desired process, because the activities 'to Test' and 'Solved' are not included. The most common variant thereafter does not include the activity 'Closed'.

The following findings are done considering the 12 most common variants. The most common step in the process is a process starting by setting the ticket status to New. This is done in 15 379 cases out of the 17 921 cases. If a process was not started by setting the ticket status to New, it was done by setting the ticket status immediately to Accepted, which was done in 2 542 cases. The second most common step in the process is a status being set from New to Accepted, which was done in 14 305 cases. This is followed by a status being set from Accepted to Closed, done in 12 248 cases. After this, the most common step is from Accepted to Solved. However, the frequency for this is much lower, namely 2 116.

In 1 085 cases a status is immediately set to Accepted after having been set to Accepted. In 379 cases, a status from New is reset. These are the only two activities which are reset. All other activities move to a different status.

With respect to the throughput time, it appears to take longest before a status is reset to Accepted. This was the case for the median, average, and trimmed mean throughput time,

with 10, 47, and 33 days respectively. This process step is not part of the desired process. After that, the three models with respect to throughput times show that a status being set from Accepted to Closed takes longest, namely 8, 30, and 21 days respectively. This is part of the desired process model, and it seems logical that this takes longest, because it represents the duration of executing a supporting action.

Looking at the results from conformance checking, in *Figure 34 Conformance Checking Case Study 2*, it can be seen that the throughput time of conforming cases is 28.3 days versus 68.7 days for violating cases. This is likely due to some extreme outliers. For example, setting a release from Released to Known Issue took 531 days in one case. While this only represents one case, it is an extreme outlier, and it may be expected that more such outliers exist for violating cases.

In *Figure 36 Dashboard Impression Tickets Case Study* 2, important results with respect to the number of tickets can be viewed. For example, the number of tickets per company, service, and project. This dashboard can be used to answer research question 4. The dashboard also shows the solved & closed status for each employee, which can be used to answer research question 3. Apart from the answers to these research questions, several other interesting findings can be presented here on request, e.g., through filtering since the dashboard is interactive.

In *Figure 37 Dashboard Impression Throughput Times Case Study* 2, important results with respect to the total throughput times can be viewed. This dashboard can be used to answer research question 2. Apart from the answer to research question 2, this dashboard provides several other interesting insights, such as the throughput time per project, company, and service. The dashboard is also interactive, and with filtering more specific findings may be presented.

The questions 1,5, and 6 were answered during conformance checking. The answers to all questions will be discussed in the next phase.

### 7.1.7 Evaluation

In this phase, the questions, as formulated in Phase 2: *Plan, Scope & Define*, will be answered.

*TVQ1: What is the most common variant of activities in the ticket process?*

The most common variant of activities in the ticket process is a ticket status being set from New to Accepted to Closed. This variant of activities is followed in 36% of all cases. After that, the most common variant is a ticket status being set from New to Accepted, to Solved, to Closed. This variant covers 8% of all cases.

*TVQ2: Which steps, projects, and companies yield a high throughput time?*

The step that yields the highest throughput time is the step from Accepted to Accepted again. After that, the highest throughput time is caused by a status being set from Accepted to Closed.

The project that yields the highest throughput time is the Company AE – A Project, with a throughput time of 811.46 days. Note that this is the average total throughput time, meaning that the total throughput time of each variant from the project is taken, and then divided by the number of variants from the project. The project has 5 different ticket IDs and has 4 process variants. All process variants have a high throughput time due to some step that involves the Closed status.

The company that yields the highest throughput time is Company AE, with a throughput time of 811.46 days, caused by the Company AE – A Project as mentioned above. This is the only project recorded for Company AE. The company that yields the second highest throughput time is Company AF, with a throughput time of 279.06 days. Note that this is again the average total throughput time, taking the total throughput time for each variant, and then dividing it by the total number of variants. The company has 5 different projects, where the high throughput time is mainly caused by 3 big projects.

*TVQ3: Are tickets solved and closed by different employees?*

In many cases, tickets are solved and closed by the same employee. No clear percentage for this could be determined, but by filtering on the Solved and Closed statuses, filtering out unknown employees and selecting process variants that involve Solved and Closed, it was clear that many tickets have been solved and closed by the same employee.

*TVQ4: Which project, service, and company holds the highest number of tickets?*

The highest number of tickets was issued for the OO – B Project. The total number of tickets issued for this project is 2 039. The OO – B Project is a project from Company OO, and several types of services were provided for this project, e.g., service alert and access control. Company OO is also by far the company for which the most tickets were issued, namely 10 105 out of the total of 25 507. After that, most tickets were issued for Company TT, namely 3 788. The type of services for which most tickets were issued is RFC, with 6 521 tickets and 179 projects.

*TVQ5: What could be a desired ticket process?*

A desired ticket process is given in *Figure 33 Reference Model Case Study 2*. It only shows the desired variant, meaning that e.g., the activity 'Not Solved' is not present in the model. As can be seen, a ticket should first be registered as New, after which it should be registered as Accepted. In case no request to the company has to be made, the supporting action will be performed, after which the ticket status can be set to Solved, and then Closed, the end of the process. In case a request for approval has to be made to the company, the ticket status after Accepted should be Request for Approval, followed by Request Approved, and then To Test. If this is done, the Solved and Closed statuses can be used again. This desired ticket process was created together with the Support Process Expert.

A best practice ticket process model was created with the Process Expert as well. This model can be seen in *Figure 35 Best Practice Process Model Case Study 2*. First, a ticket should be registered as New, after which it will either be Rejected or Accepted. In case a ticket is Rejected, the ticket should be Closed. In case a ticket is Accepted, three activities are possible: (i) a request for approval has to be made, (ii) it is needed to wait for the customer, e.g., for more information, (iii) information from a third party is required. These activities are represented by the statuses Request for Approval, Wait for Customer, and Require Information from Third Party, respectively. In the last two cases, the ticket may be Rejected or Accepted again. In case a request for approval has to be made, the request may be either rejected or accepted, represented by the statuses Request Declined, and Request Approved, respectively. In case the request is declined, the ticket will be Closed. In case the request is approved, the next status registration is To Test, followed by either Solved or Not Solved. In both cases, the ticket is Closed afterwards.

*TVQ6: How does the current ticket process differ from the desired ticket process?*

With conformance checking, it was found that 8% of all cases conforms to the desired ticket process. These cases follow the 2nd most common variant, where a ticket is set from New to Accepted to Solved to Closed. With respect to the best practice ticket process model, it was found that 9% of all cases conforms. This percentage is composed of cases that follow the 2nd most common variant, and cases that follow the 8th most common variant.

To verify the PM results, three checks, as formulated by [33], p, 42, were executed:

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

All three checks could be verified for the Celonis models.

To validate the PM results, the degree of fitness, precision, generalization, and simplicity of the model were checked. The Celonis models were constructed using the Fuzzy Miner, which aims to balance these four criteria [4]. With respect to the degree of fitness, 70% of the cases was included in the evaluated models, and no additional filters were applied. Thus, it seems that the degree of observed behaviour is well captured in the process model, so the degree of fitness seems good. The degree of precision seems to be okay for the models that show the 12 most common variants. In case all variants are shown, the model appears to allow for too much behaviour. In case fewer variants are shown, the model appears to allow for too little behaviour. Regarding the degree of generalization, the model represents the manner in which supporting actions are performed, and thus describes the system. However, not all activities are present in the model, e.g., it is not registered when a person has some difficulty solving the ticket. Nevertheless, the degree of generalization is expected to be sufficient. The degree of simplicity is good for the model showing the 12 most common variants. In case more variants are shown, the model becomes too complex, and in case less variants are shown, the model becomes too simple. In conclusion, the models appear to be sufficiently valid considering all four criteria.

The PM project objective was to gain more insight into the support ticket process. This project objective has been met; many insights were gained. Moreover, all formulated questions could be answered. First of all, it was possible to view the current variants of activities in the ticket process. Secondly, throughput times and the number of tickets could be analysed with respect to several factors, such as the company and the project. Thirdly, it was possible to investigate whether tickets are solved and closed by different employees. Lastly, a desired as well as a best practice process model could be created and evaluated with respect to the current ticket process.

### 7.1.8 Process Improvement & Presenting

Based on the findings, several improvement actions may be formulated. First of all, it should be investigated why Company OO has so many tickets. It could be the case that many incidents happen in the Company OO which cause the high number of tickets. The cause of this should be investigated, because if it is possible to treat the cause, many incidents may be prevented. It could also be the case that Company OO has many more tickets due to them registering supporting actions more quickly compared to other companies. This should be explored. Secondly, the number of tickets per project should be checked, because a high number of tickets for a project probably indicates that a project is not running smoothly. This may be due to a cause that can be resolved. Thirdly, it should be examined how it is possible that the same employee solves and closes a ticket. Since it is a rule that a ticket should be solved and closed by a different person, the importance of obedience to this rule

should be checked and it should be seen that the rule is then followed. Fourthly, the throughput times of the companies should be examined further. The reason for this, is that high throughput times are generally undesirable for companies, because it means that it takes long before a supporting action has been completed. Fifthly, the throughput times for the service types should be checked, because it might e.g., indicate that certain types of services should receive some more attention. Last but not least, it should be seen whether the best practice process model as created during this project is beneficial. Several other improvement actions may be defined, but these are expected to be the most important ones.

In *Table 17 Findings, Action Points, Stakeholders & Benefits Case Study 2*, the findings, corresponding action points, responsible stakeholder(s), and potential benefits are summarized.

*Table 17 Findings, Action Points, Stakeholders & Benefits Case Study 2*

| Finding | Action Point | Responsible Stakeholder(s) | Benefit |
|---|---|---|---|
| Company OO has many more tickets as compared to other companies. | Investigate why Company OO has many more tickets, e.g., through investigating the type and content of the tickets. | Support Customer Contact | The Company OO eMagiz platform could be improved, resulting in less incidents in the future.<br><br>The registration of incidents at Company OO or other companies may be adapted, ensuring better and fairer documentation. |
| Certain projects have a higher number of tickets as compared to other projects. | Investigate why certain projects have a higher number of tickets, e.g., through investigating the types of projects. | Support Customer Contact | The projects may be improved, resulting in less incidents in the future. |
| Tickets are solved and closed by the same employee. | Investigate why tickets are solved and closed by the same employee, e.g., through user interviews. | Support Technical Contact | The ticket registration system may be adapted to prevent the possibility of solving and closing a ticket by the same person.<br><br>The rule that a ticket has to be solved and closed by the same person may be raised. |
| Certain companies have a higher total ticket throughput time as compared to other companies. | Investigate why certain companies have a higher total ticket throughput time, e.g., through investigating per company which | Support Customer Contact | The manner in which incidents are managed could be improved, resulting in tickets being resolved more quickly in the future. |

| | | | |
|---|---|---|---|
| | tickets cause a high throughput time. | | |
| Certain services have a higher total ticket throughput time as compared to other services. | Investigate why certain services have a higher total ticket throughput time, e.g., through investigating per service which tickets cause a high throughput time. | Support Customer Contact | The prioritization of resolving tickets that belong to certain types of services may be changed, resulting in certain tickets being resolved more quickly. |
| A new ticket process model has been created. | Investigate whether the new ticket process model is usable and efficient, e.g., through user interviews. | Business Consultant | The process model may give guidance to users in the ticket registration process.<br><br>The process model may be adapted such that it is more usable and efficient. |

As mentioned in the previous case study and the methodology, two other improvement actions can be formulated, namely (i) consider the integration of event logs with systems, and (ii) consider the distribution of event logs through a centralized portal. The integration of event logs with systems could be achieved by ensuring that event logs are automatically generated from the CAPE Service Point system, the system that is used by Support. The distribution of event logs through a centralized portal is a companywide activity.

The results were communicated to the involved stakeholders during several meetings.

### 7.1.9 Change Management

To ensure that PM is a continuous activity in the organization, the improvement actions as defined in the previous phase should be executed. This can best be done by the Support team. Apart from following the improvement actions, it could also be investigated whether other processes can be useful for process mining. These could be processes from Support, but also processes from eMagiz, as suggested in *Phase 9: Change Management*. With respect to change management practices, existing practices may be followed, as mentioned previously (*Phase 9: Change Management*).

## 7.2 Expert and Practitioner Evaluations

To validate PROMISE further, expert and practitioner evaluations were obtained. The participants consisted of (i) people that have completed at least one PM project (experts), and (ii) practitioners within an SME that have a basic level of knowledge, experience, and skills in the field of data and business processes (practitioners). The evaluations were obtained by providing the experts and practitioners with (i) an assessment form, (ii) a user guide, given in Appendix *F User Guide*, (iii) the PROMISE visualization, PROMISE pillars and PROMISE steps, and (iv) the results from the second case study. The user guide provides some basic knowledge about PM and gives some general information about the methodology.

### 7.2.1 Evaluation Details and Results

The assessment forms that were given to the practitioners differed from the assessment forms that were given to the experts with respect to a few statements. The reason for this, is that people already familiar with PM are likely to have a different view about the

methodology. For example, the experts were asked to evaluate whether PROMISE includes all necessary PM activities, which would be difficult to evaluate for practitioners. So, the results from practitioners will be presented distinctly from the results from experts.

The statements for the assessment forms were formulated based on the constructs of the Technology Acceptance Model (TAM) [21]. These are perceived usefulness, perceived ease of use, and intention to use. Applying these constructs to our study, their meanings can be interpreted as follows: (i) perceived usefulness: the degree to which a person believes that PROMISE would enhance the performance of PM activities, (ii) perceived ease of use: the degree to which a person believes that using PROMISE would be free of effort, and (iii) intention to use: the degree to which a person intends to use PROMISE.

For each of the constructs, statements were formulated. To identify the level of agreement for each statement, a five-point Likert scale was used, indicating the options strongly disagree, disagree, neutral, agree, and strongly agree.

Participant details are given in *Table 18 Evaluations Participant Details*.

*Table 18 Evaluations Participant Details*

| Reference | Type of Participant | Country | Current Position |
|---|---|---|---|
| **E1** | Expert | The Netherlands | EngD candidate & researcher on PM, author of [60] |
| **E2** | Expert | The Netherlands | BI Consultant |
| **E3** | Expert | The Netherlands | BI Consultant |
| **P1** | Practitioner | The Netherlands | CTO |
| **P2** | Practitioner | The Netherlands | Support Technical Contact |
| **P3** | Practitioner | The Netherlands | Business Consultant |
| **P4** | Practitioner | The Netherlands | Team Lead |
| **P5** | Practitioner | The Netherlands | Manager Expert Services |

The individual results from the experts and practitioners are given in Appendix *G Evaluation Results Experts* and Appendix *H Evaluation Results Practitioners*, respectively.

The results from experts with averages are given in *Table 19 Results Expert Evaluations*.

*Table 19 Results Expert Evaluations*

| Construct | Code | Statement | Average Rate | Average Rate per Construct | Result |
|---|---|---|---|---|---|
| Perceived Usefulness | PU1 | I think that PROMISE covers all necessary PM activities to set up a PM project. | 4.67 | 4.25 | Agree |
| | PU2 | I think that PROMISE can help practitioners to start up a PM project. | 4.33 | | |
| | PU3 | I think that PROMISE can help to better understand the PM activities that are needed to start up a PM project. | 4.67 | | |
| | PU4 | I think that PROMISE gives good practical guidance. | 3.33 | | |

| | | | | | |
|---|---|---|---|---|---|
| Perceived Ease of Use | PEU1 | I think that the PROMISE steps are clear and easy to understand. | 5.00 | 4.33 | Agree |
| | PEU2 | I think that the PROMISE visualization is clear and easy to understand. | 4.33 | | |
| | PEU3 | I think that the PROMISE pillars are clear and easy to understand. | 5.00 | | |
| | PEU4 | I would be able to start a PM project based on PROMISE. | 4.33 | | |
| | PEU5 | I think that I would **not** need technical help to use PROMISE. | 3.33 | | |
| | PEU6 | Overall, I think that PROMISE is easy to use. | 4.00 | | |
| Intention to Use | ITU1 | I would use PROMISE when executing a PM project. | 3.67 | 3.50 | Agree |
| | ITU2 | I would refer to PROMISE if I got lost in a PM project. | 3.33 | | |

The results from practitioners with averages are given in *Table 20 Results Practitioner Evaluations*.

*Table 20 Results Practitioner Evaluations*

| **Construct** | **Code** | **Statement** | **Average Rate** | **Average Rate per Construct** | **Result** |
|---|---|---|---|---|---|
| Perceived Usefulness | PU1 | I think that I would be able to start up a PM project based on PROMISE. | 3.80 | 4.00 | Agree |
| | PU2 | I think that PROMISE can help to better understand the PM activities that are needed to start up a PM project. | 4.60 | | |
| | PU3 | I think that PROMISE gives good practical guidance. | 3.60 | | |
| Perceived Ease of Use | PEU1 | I think that the PROMISE steps are clear and easy to understand. | 4.20 | 3.73 | Agree |
| | PEU2 | I think that the PROMISE visualization is clear and easy to understand. | 4.00 | | |
| | PEU3 | I think that the PROMISE pillars are clear and easy to understand. | 4.00 | | |
| | PEU4 | I would be able to start a PM project based on PROMISE. | 3.60 | | |
| | PEU5 | I think that I would **not** need technical help to use PROMISE. | 2.80 | | |
| | PEU6 | Overall, I think that PROMISE is easy to use. | 3.80 | | |
| Intention to Use | ITU1 | I would use PROMISE when executing a PM project. | 4.20 | 4.30 | Agree |
| | ITU2 | I would refer to PROMISE if I got lost in a PM project. | 4.40 | | |

### 7.2.2 Expert Evaluations Discussion

The experts agreed on all three constructs, as well as all 12 statements. The construct ‘perceived ease of use’ received the highest score, with an average of 4.33. The statements PEU1 and PEU3, concerning the clarity and understandability of the PROMISE steps and the clarity and understandability of the PROMISE pillars obtained the highest average scores, both 5.00. The lowest average scores belong to PU4, PEU5, and ITU2, concerning the practical guidance of PROMISE, the need of technical help, and referencing to PROMISE, respectively, all with scores 3.33. The score for ITU2 was quite diverse, two experts gave a 4, and one expert gave a 2. The reason for a score of 2 having been given, is that this expert would rather refer to previous projects to base next steps on (E3). It was also stated that this expert would refer to PROMISE to give others an understanding of the material (E3). No other statements received a score of 2.

Overall, the experts indicated that PROMISE seems to cover all necessary PM activities and that it is clear and understandable. The pillars were evaluated to be a good addition to the visualization (E2), and the steps are clear and well explained (E1). The main threats to PROMISE appear to be that (i) it may be difficult for (non-data driven) employees to put the methodology into practice (E1, E2, E3), mainly with respect to data and tools, and (ii) the methodology describes the “perfect” process, meaning that difficulties may arise when challenges are encountered (E3). With respect to the first threat, it seems that it is important to involve a PM expert in the process, to support the data extraction and tooling activities if needed. This has also been indicated in PROMISE, see *Table 4 Stakeholders*. Moreover, existing studies on PM tools may be used to support the choice of a desired PM tool. Regarding the challenges that may be encountered, it is likely that practitioners will encounter challenges that have not been described in PROMISE. However, the methodology should be used as a guideline. Describing all potential challenges would likely cause the methodology to be too extensive. The case study descriptions in this study describe the PM activities undertaken in the case studies, which may also help to overcome any challenges. During the case studies, no significant challenges, which could not be resolved while following PROMISE, were encountered.

### 7.2.3 Practitioner Evaluations Discussion

The practitioners agreed on all three constructs, and 10 out of the 11 statements. The construct ‘intention to use’ received the highest score, with an average of 4.30. The statement that obtained the highest average score is PU2, concerning the statement that PROMISE can help to better understand the PM activities that are needed to start up a PM project. The statement on which the practitioners disagreed is PEU5, which concerns the need for technical help, with an average of 2.80. Specifically, three practitioners **dis**agreed that they would **not** need technical help to use PROMISE, one agreed that they would **not** need technical help, and one gave a neutral score. It was already predicted by experts that technical help would be desired, so this supports those predictions.

Overall, the practitioners indicated that PROMISE is easy to understand, that the methodology helps to better understand the needed PM activities to set up a PM project, and that they would refer to PROMISE if they got lost in a PM project. The results from the second case study helped the practitioners to better understand PROMISE (P5, P1). However, the practitioners indicated that they would need technical help to use PROMISE, especially on the part of data analysis (P5). Moreover, they stated that the methodology is likely not to be easy for most non-experts (P4), and that the understandability depends on the role, level, and experience of the person reading it (P4). This is also a requirement for PROMISE, namely that the user has a basic level of knowledge, experience, and skills in the

field of data and businesses. Moreover, as mentioned previously, it is advised to involve a PM expert in the process and to review existing studies on PM tools.

## 7.3 Validation

Based on the findings from applying PROMISE in the case study, several conclusions can be drawn. First of all, the steps gave good guidance, it was clear what needed to be done. Not all steps were performed in the defined order, e.g., a process was selected much earlier than defined in the methodology. The reason for this, is that it was immediately clear what process should be analysed. Nevertheless, in general, the order seemed to be right, and differences such as described above are likely to remain in practice. Moreover, all steps seemed to be necessary, so, while slight changes in order may occur in practice, it is important to check all the steps. Furthermore, a difference with respect to the defined PM techniques occurred. While enhancement is usually done after conformance, in this case, enhancement was done first, because no process model was present yet. However, this is expected to be an exception, so the PM techniques cycle should remain the same, since it describes the most logical order of applying PM techniques.

For a validation of PROMISE, the methodology should be validated against the requirements, as defined in *Table 6 Requirements*.

*R1. The PM (adapted) methodology must be usable in any SME.*

The methodology developed in this study addresses aspects to implement PM which were found to be important and relevant in two case studies. While the order in which steps have to be executed may differ a little per organization, no irrelevant steps were found. Moreover, no SME type or domain specific aspects are addressed in the methodology. Thus, it can be assumed that PROMISE is applicable in any type of SME. Since PM can be applied in any organization [5], it can be concluded that the PM (adapted) methodology is usable in any SME.

*R2. The PM (adapted) methodology must be understandable and usable by any practitioner in an SME, who has a basic level of knowledge, experience, and skills in the field of data and business processes.*

To determine the degree in which this requirement has been met, expert and practitioner evaluations were obtained. From these evaluations, it was clear that this requirement has been met. The experts and practitioners indicated that the PROMISE visualization, steps, and pillars are clear and easy to understand. Moreover, the practitioners expressed that the results of the second case study helped to better understand the methodology. It was also evaluated that technical help to support some of the PM activities is desired. This need for technical help will depend on the level of knowledge, experience, and skills of the user. It is advised to appoint a PM expert that may support some of the PM activities if needed.

To ease the understanding of PROMISE, a user guide was created. This user guide provides some basic knowledge about PM and gives some general information about the methodology. The user guide was given to the experts and practitioners, and can be found in Appendix *F User Guide*.

*R3. The PM methodology must address features that extract, integrate, and combine process-related data.*

Features that extract, integrate, and combine process-related data concern receiving, extracting, and transforming process data, gathering and analysing big data, and transforming data into process logs [12]. These aspects are described in the following

phases: Phase 2: Plan, Scope & Define, Phase 3: Data Exploration & Understanding, and Phase 4: Event Log Creation. So, R3 is satisfied.

*R4. The PM methodology must address features that visualize process execution.*

Features that visualize process execution encompass the visualization of processes and adaptations of those visualizations based on different criteria [12]. These features are addressed in Phase 5: Process Model Creation and Phase 6: Analysis & Enhancement. Thus, R4 is met.

*R5. The PM methodology must address features that generate various process related KPIs.*

Features that generate various process related KPIs may concern defining customized KPI and comparing 'as is' against 'to be' process models. These factors are addressed in Phase 2: Plan, Scope & Define, Phase 5: Process Model Creation and Phase 6: Analysis & Enhancement. However, defining KPIs was found to be less important for SMEs. For SMEs question-driven project are advised due to their limited number of resources. Moreover, especially for SMEs starting their first PM project, question-driven projects are expected to be more useful. This was also found by one of the authors of [60]. Thus, while it has been stated in PROMISE that KPIs can be formulated, the focus lies on the formulation of research questions.

In *Table 21 Phases & Requirements*, an overview of R3, R4 and R5 and the phases in which these requirements are addressed is given.

*Table 21 Phases & Requirements*

| **METHODOLOGY PHASE** | **REQUIREMENTS** |
|---|---|
| PHASE 2: PLAN, SCOPE & DEFINE | R3 & R5 |
| PHASE 3: DATA EXPLORATION & UNDERSTANDING | R3 |
| PHASE 4: EVENT LOG CREATION | R3 |
| PHASE 5: PROCESS MODEL CREATION | R4 & R5 |
| PHASE 6: ANALYSIS & ENHANCEMENT | R4 & R5 |

*R6. The PM methodology must address all PM project elements defined in* [60].

In *Table 22 Phases, PM Project Elements & Steps*, the PM project elements from [60] are categorized into the phases from PROMISE. The steps in which a project element is addressed have been added as well. As can be seen, all project elements defined in [60] have been addressed.

*Table 22 Phases, PM Project Elements & Steps*

| **METHODOLOGY PHASE** | **PM PROJECT ELEMENTS FROM** [60] |
|---|---|
| ACROSS PHASES | - Iterative nature |
| PHASE 1: BUSINESS UNDERSTANDING | - Organizational willingness (S1, S2)<br>- Stakeholder involvement (S2, S3)<br>- Process selection (S4, S5)<br>- Familiarity with process mining (S3)<br>- Data availability (S4, S5) |
| PHASE 2: PLAN, SCOPE & DEFINE | - Organizational willingness (S9)<br>- Stakeholder involvement (S10, S11, S12) |

| | |
|---|---|
| | - Linking business goals to PM projects (S6)<br>- Process selection (S7, S8)<br>- Project goal (S6, S13)<br>- Desired insights and KPI selection (S6, S13)<br>- Data availability (S7, S8, S9) |
| PHASE 3: DATA EXPLORATION & UNDERSTANDING | - Linking business goals to PM projects (S14)<br>- Vendor selection (S15)<br>- Project goal (S14)<br>- Desired insights and KPI selection (S14)<br>- Data availability (S16) |
| PHASE 4: EVENT LOG CREATION | - Data availability (S17)<br>- Data extraction and preparation (S18, S19, S20, S21) |
| PHASE 5: PROCESS MODEL CREATION | - Creation of process dashboards (S22, S23, S24, S25) |
| PHASE 6: ANALYSIS & ENHANCEMENT | - Creation of process dashboards (S27)<br>- Analysis of dashboard (S26) |
| PHASE 7: EVALUATION | - Interpretation and conclusion (S28, S31)<br>- Validation (S29, S30) |
| PHASE 8: PROCESS IMPROVEMENT & PRESENTING | - Defining improvement actions (S32, S33)<br>- Quantify, select, monitor improvements (S34, S35)<br>- Communicating quick wins (S32, S36) |
| PHASE 9: CHANGE MANAGEMENT | - Quantify, select, monitor improvements (S37, S38)<br>- Continuous effort (S37, S38) |

## 7.4 Summary

To summarize, this chapter concerned an application of the refined PM methodology in another case study at eMagiz. For each of the phases, the results have been described. After completion of the case study, expert and practitioner evaluations were obtained to further validate PROMISE. It was found that PROMISE is perceived to be understandable and usable by experts as well as practitioners, but that practitioners may need some technical help. Moreover, experts, but mainly practitioners intend to use the methodology. After the evaluations were obtained, PROMISE was validated against the requirements. It could be concluded that all requirements have been met.

# 8 Discussion

In this chapter, the contributions of this thesis will be presented, recommendations to methodology adopters and the organization at which the case studies were performed will be given, and the key takeaways will be presented. The contribution of this research to science is described in Section *8.1 Contribution to Science*, and the contribution to practice is discussed in Section *8.2 Contribution to Practice*. Recommendations are given in Section *8.3 Recommendations for Methodology Adopters* and Section *8.4 Recommendations for the Company*. Lastly, Section *8.5 Key Takeaways* covers the key takeaways.

## 8.1 Contribution to Science

Existing studies on PM lack research on how PM should be used in practice, especially research on the use of PM in SMEs is limited. This study contributes to this research gap by providing a methodology on the application of PM in SMEs. A methodology visualization was developed to give an overview of the phases that should be followed. In addition, methodology pillars were created to provide information with respect to the goals and deliverables for each phase. Lastly, specific steps that represent the PM activities that should be executed have been formulated.

Before the development of the methodology, an SLR was performed to give an overview of the existing landscape of research and literature on the topic of PM in organizations. In this SLR it was evaluated which methods/techniques/approaches/findings regarding the application of PM in SMEs have been published in the last decade (2012 to 2022). Furthermore, empirical evidence and evaluation approaches for these findings were assessed. Based on the findings, it was suggested to further validate results on the use of PM in organizations in future research, which has been achieved in this study.

The PM methodology was designed based on an evaluation of existing PM methodologies. Specifically, PM project elements found by [60] that were evaluated to be necessary when applying PM were used as steppingstones for the methodology. Moreover, guidelines for each of these elements were created based on findings from literature about PM in SMEs and large organizations. Additionally, the methodology was refined and validated through two case studies.

## 8.2 Contribution to Practice

The main practical contribution of this research is that it provides a methodology on the application of PM in SMEs, with practical guidelines, so that it is immediately applicable by practitioners that have a basic level of experience, knowledge and skills in the field of data and business processes. Since PM has proven to be valuable for organizations, and it appears that practitioners miss guidance in their application of PM, PROMISE is expected to be significant for practitioners. It was also evaluated by practitioners that the methodology is clear and easy to understand, and that it would be useful when starting a PM project.

To ease the understanding of PROMISE, a user guide was developed. This user guide provides basic knowledge about PM and some information about PROMISE, such that any practitioner with a relevant background should be able to apply it. Apart from this user guide, the designed steps should be of great help. While it was found that existing PM methodologies do not appear to provide practical guidelines, the steps from PROMISE show exactly which PM activities should be executed, and in which order. While it may differ slightly per organization which steps should be executed, all steps appeared to be relevant in two case studies. Moreover, the methodology was improved after one case study, refining the order in which the steps should be executed. Lastly, the results from both case studies have been included in this study, providing examples of successful PM projects for

practitioners. Nevertheless, it was found that practitioners expect to need technical help when using PROMISE, mainly during tool selection and application.

Apart from a more generic practical contribution, the study has shown its value to eMagiz, the SME at which the two case studies were performed. The reason for this, is that both PM projects were completed successfully, providing eMagiz with findings and potential improvements for two processes. Apart from this, the potential of PM for eMagiz will be further examined due to the success of the case studies.

## 8.3 Recommendations for Methodology Adopters

Several recommendations for PM practitioners who wish to adopt PROMISE can be made. First of all, it is required that the practitioner has a basic level of experience, knowledge and skills in the field of data and business processes. While this should be sufficient to understand PROMISE, and a user guide has been developed to ease the understandability of the methodology, it would be beneficial if the practitioner has some more knowledge on PM before adopting the methodology. This may be achieved by following a few PM courses and should enhance the understandability of the methodology. Moreover, it would be good if technical help can be obtained to support the PM activities. This was evident from the expert and practitioner evaluations. Such support is most likely to be needed when choosing and using a PM tool (E2).

Secondly, since SMEs have a limited number of resources, it is important to check for resource availability. Specifically, it should be checked whether the organization is able to give enough data and commitment, as described in PROMISE. These resources include stakeholders and PM tools. Stakeholders that may be required to execute the PM activities have been described in this study. Not much attention has been paid to PM tools, but several studies about PM tools have been performed, such as [9][22][35].

Thirdly, it is important to follow up on action points that may result from applying PROMISE. This will ensure that the results from the methodology are useful for the organization. For this, change management practices should be in place. Since change management practices were already in place at eMagiz, these have not been described in this study. However, several studies on this topic are available [36][44], which may be useful to check before applying PROMISE.

## 8.4 Recommendations for the Company

Several recommendations for eMagiz can be made. First of all, it is recommended to follow up on the action points formulated in this study. This will ensure that the findings from the case studies will have the desired impact and may be achieved by following existing change management practices. Secondly, it is advised to evaluate the possibility for future PM projects within the company. For this, PROMISE may be applied. Lastly, it could be considered to offer PM as a service to customers. eMagiz is an integration platform, through which much customer data flows. Currently, this data is not stored. However, it could be possible to store this data and apply PM to it, such that useful insights for customers can be gained. Since customer data is already flowing to the platform, it would only have to be stored, after which PM can be applied.

## 8.5 Key Takeaways

This thesis presents a methodology, PROMISE, on the application of PM in SMEs. It is directed towards practitioners who have a basic level of knowledge, experience, and skills in the field of data and business processes. The methodology should guide practitioners when starting up a PM project in an SME. The PROMISE visualization illustrates the phases that have to be followed when executing a PM project. The PROMISE pillars define the purpose

and deliverables for each phase. The PROMISE steps outline the specific PM activities that have to be executed.

The methodology was developed following the Design Science Methodology (DSM) from Wieringa [57]. The problem investigation was completed by investigating the difference between PM in SMEs versus large organizations, analysing existing PM methodologies, and defining the stakeholders and goals for the study. The treatment design was achieved through formulating requirements for the methodology, mapping methodology phases and elements, establishing guidelines based on the literature, and developing a first version of PROMISE. It was deemed possible to execute two case studies to evaluate PROMISE, so it was decided to add a treatment refinement, to refine the methodology. For the treatment refinement, a case study was conducted at eMagiz, concerning a release process. Afterwards, PROMISE was improved according to the results from the case study. The treatment validation was accomplished by conducting another case study. This case study was executed at Support, a supporting unit from eMagiz, and concerned their ticket process. After this case study, the results were evaluated, and the methodology was validated with respect to the requirements as formulated in the treatment design. It was concluded PROMISE is understandable by experts and practitioners, and most would use the methodology when executing a PM project. However, to execute the PM activities, some technical help is desired.

## 8.6 Summary

To summarize, the main contribution to science is that this study bridges the research gap concerning how PM should be used in practice. The main contribution to practice is that this study provides a PM methodology with practical guidelines, so that it is immediately applicable by practitioners with a basic level of knowledge, experience, and skills in the field of data and business processes. Recommendations for methodology adopters concern obtaining knowledge on PM, ensuring that PM expert advice may be obtained if needed, checking for resource availability, and following up on action points. Recommendations for the company are to follow up on the formulated action points, to evaluate the possibility for future PM projects, and to consider offering PM to customers as a service.

# 9 Conclusion

In this chapter, the main findings of this research are addressed. First, the main conclusions of the research questions are given. Then, limitations are discussed, and finally, suggestions for future research are presented.

## 9.1 Main Conclusions

The goal of this research is to provide an answer to the following main research question:

*How to design a methodology on the use of PM*

*that gives practical guidelines*

*so that PM can be implemented optimally*

*in SMEs?*

To answer this question, several sub-questions have been formulated. These will be answered in the following section.

### 9.1.1 Answers to Research Questions

**RQ1**: What empirical evidence on PM in SMEs is available?

In the SLR conducted in this study, 11 studies were identified that focus on PM in SMEs. Most of these studies are analyses, but also e.g., frameworks, methods, and implementations were produced. The studies in which a method was developed were not focussed on the use of PM in SMEs. Rather, these focussed on attaining a wider organizational goal using PM techniques, such as supporting collaborative processes in SMEs.

From the 11 studies that were found through the SLR, 10 studies were evaluated to provide empirical evidence on PM in SMEs. Most of this empirical evidence consisted of case studies, namely 8 out of the 10 studies. The remaining two studies produced empirical evidence through an illustrative example and a focus group study.

With respect to evaluation approaches to validate the methods, techniques, or other findings from the studies, it was found that only one study provides a specific validation method.

Since the number of studies that were found to focus on PM in SMEs was very low, studies focussing on PM in large organizations or other organizational contexts were also included. This resulted in a total of 21 studies. From these studies, one methodology on PM was found. With respect to empirical evidence, 9 out of the 10 added studies provided empirical evidence, again mainly through case studies. Regarding validation, 3 studies were found to provide a specific validation method.

Based on the findings, it was concluded that more research on the use of PM in SMEs is needed.

**RQ2**: How does PM in SMEs differ from PM in large organizations?

SMEs mainly differ from large organizations with respect to their size and annual turnover. Moreover, SMEs generally have less process maturity, lower managerial skills and employees tend to have multiple roles to fulfil. As a result of their lack of resources, SMEs sometimes choose to have an out-of-date process. So, based on the outcomes from PM, suggested improvements are not always implemented, making PM ineffective. SMEs do appear to house deeper IS/IT knowledge, which could help to better understand the PM activities.

Four PM challenges for SMEs were found by [26]. These are (i) preparation of event log data, (ii) poor documentation quality, (iii) creation of awareness for PM, and (iv) shifting manpower to fulfil PM tasks. Another PM challenge found from the literature is the choice of an appropriate case ID. However, since this challenge does not seem to arise from differences between SMEs and large organizations, such as a limited number of resources, this challenge is likely to arise in large organizations as well.

Some SME specific PM guidelines could be formulated as well. First of all, because SMEs generally have a lower budget and PM is not the highest priority, it is important to move across phases quickly to present an MVP such that trust is gained. Secondly, an open culture should be maintained, and employees should be involved to help create awareness for the benefits and costs for SMEs. Thirdly, it should be established who should be involved in the PM activities and who is responsible for the activities, since SMEs have limited resources. This should also help to ensure that the PM project can be completed and that it can be a continuous activity within the organization. Fourthly, to ensure that process evolutions can be made, it needs to be clear that people are available to help with PM activities during process selection. Fifthly, since SMEs typically have less process maturity, it is advised to start with a question-driven project to help scope the project and guide data extraction efforts. Lastly, because SMEs tend to have poor documentation quality, a comprehensive knowledge base needs to be established such that conformance checking can be achieved.

A more detailed description about the differences of applying PM in SMEs versus large organizations can be found in Section *4.1 PM in SMEs Versus Large Organizations*. Further explanations about PM guidelines specifically for SMEs can be found in Section *5.3 Phases & Elements Descriptions*.

**RQ3**: How useful are existing PM methodologies when starting with PM in an SME?

In total, six PM methodologies were evaluated. Four of these methodologies were assessed in [60], and mainly these methodologies were considered while structuring the methodology. Some guidelines were also retrieved from these methodologies, but they were mainly used for the structure of our methodology.

First of all, a mapping of the phases of these four PM methodologies was made. This resulted in seven phases, with a short description for each phase, based on the descriptions of all four methodologies. Then, the PM project elements from [60] were mapped into the phases, taking into account the descriptions. The categorization of these elements into phases was verified by one of the authors of [60]. During the further development of the methodology, phases were added and adapted, but the main structure is based on those four methodologies.

For the further development of the methodology, findings from literature were used, not only findings from the evaluated PM methodologies.

The evaluation of existing PM methodologies can be found in *Section 4.2 Existing PM Methodologies*.

**RQ4**: What are requirements for PM methodologies to be effective?

No requirements for methodologies could be found in the literature. However, several requirements were formulated according to several aspects.

Based on our definition of a methodology, being that it should deal with general principles, one requirement could be formulated. This requirement defines that the PM (adapted) methodology must be usable in any SME (R1).

Considering our main research question, two more requirements could be formulated. These concern (i) practical guidelines, and (ii) optimal implementation. With respect to the practical guidelines, this means that practitioners should be able to understand and use the methodology. Specifically, the PM (adapted) methodology must be understandable and usable by any practitioner in an SME, who has a basic level of knowledge, experience, and skills in the field of data and business processes (R2).

Regarding the optimal implementation, this is interpreted that by applying the methodology, the highest business values that can be obtained by applying PM can be achieved. It was found in the literature that, for this to be achieved, three key features of PM should be addressed. These are data connectivity, process visualization and process analytics. Thus, three requirements were formulated with respect to features that the PM methodology should address. Namely, the PM methodology must address features that extract, integrate, and combine process-related data (R3), features that visualize process execution (R4), and features that generate various process-related KPIs (R5).

The last requirement concerns more specific elements that the methodology should address. In [60], PM project elements are formulated that should be addressed in PM methodologies. Thus, it was decided that the PM methodology must address all PM project elements in [60] (R6).

The requirements were formulated in Section *5.1 Methodology Requirements*, which can be referred to for further explanations.

**RQ5**: What elements should a methodology on the use of PM in SMEs address?

As mentioned previously, the PM methodology should address all PM project elements defined in [60]. While no definitions of these elements were given in [60], our interpretations of the elements were discussed with one of the authors of [60]. Next to these elements, three types of features should be addressed, concerning data connectivity, process visualization and process analytics. Several project elements addressed these features, so this caused no difficulties.

Guidelines for the elements were formulated based on the literature. For the validation of the methodology, it was evaluated whether all elements had been addressed, which is the case.

All PM project elements including their descriptions can be found in Section *5.3 Phases & Elements Descriptions*. An overview of where the elements are addressed in the final methodology can be found in *Table 22 Phases, PM Project Elements & Steps*.

**RQ6**: To what extent can the proposed PM methodology be validated against the requirements?

It was found that the PM methodology, PROMISE, could be validated against all six requirements. First of all, since the methodology addresses steps that were found to be relevant in two case studies, does not address SME type or domain specific aspects, and PM can be applied in any organization as found from the literature, it was evaluated that the PM (adapted) methodology is usable in any SME. Secondly, to ensure that the PM (adapted) methodology is understandable and usable by any practitioner in an SME, who has a basic level of knowledge, experience, and skills in the field of data and business processes, a user guide was created. Moreover, the results from the case studies may help to increase the

usability and understandability as well. PROMISE was also evaluated by experts and practitioners with respect to its usability and understandability, where it was found that the methodology is clear and easy to understand and is expected to be useful for practitioners. Thirdly, it was concluded that PROMISE addresses all required features, namely features that extract, integrate, and combine process-related data, features that visualize process execution, and features that generate various process-related KPIs. Lastly, PROMISE addresses all PM project elements from [60], which was evaluated by categorizing all elements into phases of the methodology.

A more detailed validation against the requirements is given in Section *7.3 Validation*.

### 9.1.2 Concluding Remarks

Through the research questions as given above, the answer to the main research question could be found. This thesis shows how to design a methodology on the use of PM that gives practical guidelines so that PM can be implemented optimally in SMEs. This has been achieved through three main activities.

First, literature was reviewed to establish requirements for the methodology and to obtain findings on the use of PM in SMEs. Based on existing PM methodologies, an initial structure for PROMISE was developed, and elements and features that should be addressed in the methodology were established. Findings from the literature on the use of PM in SMEs were used to formulate guidelines for all phases of PROMISE, addressing the given elements and features.

Secondly, PROMISE was refined through a case study. The formulated phases and guidelines were evaluated through the implementation of a PM project. Based on the findings from this case study, PROMISE was improved. The main improvements concerned a transformation from guidelines into steps, a reordering of the newly defined steps, the addition of PM techniques to the methodology visualization, and a new design of the PROMISE visualization and pillars.

Thirdly, the refined methodology was validated through a second case study. This was achieved by executing a new PM project. Moreover, PROMISE was evaluated by experts and practitioners with respect to its perceived usefulness, perceived ease of use, and intention to use. Based on the findings from this case study and the evaluations, the methodology was validated against the requirements.

## 9.2 Limitations

The two main limitations of this research are that (i) the methodology has not been applied by another researcher or practitioner, and (ii) the methodology has not been applied at an SME other than eMagiz. However, the first limitation was mitigated by an evaluation of the methodology by experts and practitioners, and the creation of a user guide. The second limitation was mitigated by conducting two case studies. One of these case studies was completed for support, a supporting unit for eMagiz. So, while this is not a different company, several aspects were different, such as the employees that were involved.

Another limitation is that the PM project elements that were found to be important for PM methodologies were found in only one study. However, the elements in that study were found through a critical review of four existing PM methodologies and 27 interviews with PM experts and professionals. Moreover, guidelines belonging to the elements were formulated based on several studies found from the literature.

Additionally, the case studies were completed using only one PM tool. While most PM tools provide a similar set of functionalities, some tools may e.g., not provide the possibility to

create a PM dashboard. Thus, depending on the type of PM tool, certain activities may not be possible, or additional functionalities may be in place. Nevertheless, no limitations for the Celonis tool were found, and several studies on PM tools exist, which may be studied before choosing a PM tool.

Lastly, it may be the case that not all relevant literature has been included. Yet, an extensive literature review has been performed for the years 2012 to 2022. For this literature review, the type of result of each study was evaluated as well, also considering methodologies.

## 9.3 Future Research

For future research, it would be good to apply PROMISE in several different SMEs. This way, it would be possible to further validate that the methodology is applicable in any SME. Moreover, PROMISE should be applied by a different researcher or practitioner, to further validate that the methodology is applicable by any practitioner who has a basic level of experience, knowledge, and skills in the field of data and business processes. Furthermore, PROMISE could be applied in a country other than the Netherlands to validate that the methodology applies to SMEs in any country.

Additionally, PROMISE could be applied using different PM tools, to ensure that the methodology is usable using any PM tool. Moreover, it could be investigated which PM tool works best when following the PROMISE steps. Another idea for future research with respect to PM tools is to develop a tool that provides customized dashboards and KPIs, in other words, a tool that automizes certain aspects of the methodology.

PROMISE could also be validated further through a more extensive expert evaluation. For example, in depth interviews with PM experts and professionals could be conducted to validate whether PROMISE encompasses all necessary PM steps.

Lastly, several suggestions for future research can be formulated based on the results from the SLR. First of all, it would be interesting to investigate to what extent results from PM in large organizations can be generalized to SMEs. Secondly, the exact value of PM for SMEs, e.g., an exact monetary value, was not clear from the literature, so this may be a topic for future research. Thirdly, it was found in the SLR that no studies were a continuation of another study. It may be good to validate existing results from PM studies further.

## 9.4 Summary

To summarize, this study presents how to design a methodology on the use of PM in SMEs that gives practical guidelines so that PM can be implemented optimally in SMEs. All research questions could be answered, and the methodology developed in the study was refined through one case study, validated through another case study, and validated through expert and practitioner evaluations. The main limitations of this study are that the methodology has not been applied by another researcher or practitioner, and that the methodology has not been applied at an SME other than eMagiz. Suggestions for future research concerning the methodology are to apply the methodology in several different SMEs, to apply the methodology using different PM tools, and to obtain more extensive expert evaluations.

# References


[1] W. van der Aalst, Process mining: discovery, conformance and enhancement of business processes, Springer, 2011.

[2] W. van der Aalst, Process mining: Overview and opportunities, ACM Transactions on Management Information Systems (TMIS). 3 (2012) 1–17.

[3] W. van der Aalst, A. Adriansyah, A.K.A. de Medeiros, F. Arcieri, T. Baier, T. Blickle, J.C. Bose, P. van den Brand, R. Brandtjen, J. Buijs, Process mining manifesto, in: International Conference on Business Process Management, Springer, 2012: pp. 169–194.

[4] W.M.P. van der Aalst, Process mining: discovering and improving Spaghetti and Lasagna processes, in: 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2011: pp. 1–7.

[5] W.M.P. van der Aalst, Process mining: a 360 degree overview, in: Process Mining Handbook, Springer, 2022: pp. 3–34.

[6] W.M.P. van der Aalst, J. Carmona, Scaling Process Mining to Turn Insights into Actions, in: Process Mining Handbook, Springer, 2022: pp. 495–502.

[7] W.M.P. van der Aalst, K.M. van Hee, A.H.M. ter Hofstede, N. Sidorova, H.M.W. Verbeek, M. Voorhoeve, M.T. Wynn, Soundness of workflow nets: classification, decidability, and analysis, Formal Aspects of Computing. 23 (2011) 333–363.

[8] S. Aguirre, C. Parra, M. Sepúlveda, Methodological proposal for process mining projects, International Journal of Business Process Integration and Management. 8 (2017) 102–113.

[9] I.-M. Ailenei, Process mining tools: A comparative analysis, Partial Fulfillment of the Requirements for the Degree of Master of Science in Business Information Systems, Eindhoven University of Technology, Netherlands. (2011).

[10] I.F. Alexander, A taxonomy of stakeholders: Human roles in system development, International Journal of Technology and Human Interaction (IJTHI). 1 (2005) 23–59.

[11] S. Astromskis, A. Janes, M. Mairegger, A process mining approach to measure how users interact with software: an industrial case study, in: Proceedings of the 2015 International Conference on Software and System Process, 2015: pp. 137–141.

[12] P. Badakhshan, B. Wurm, T. Grisold, J. Geyer-Klingeberg, J. Mendling, J. vom Brocke, Creating business value with process mining, The Journal of Strategic Information Systems. 31 (2022) 101745.

[13] F.R. Blum, Mining software process lines, in: 2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C), IEEE, 2016: pp. 839–842.

[14] A. Bryman, Of methods and methodology, Qualitative Research in Organizations and Management: An International Journal. 3 (2008) 159–168.

[15] F.A. Bukhsh, Z.A. Bukhsh, M. Daneva, A systematic literature review on requirement prioritization techniques and their empirical evaluation, Comput Stand Interfaces. 69 (2020) 103389.

[16] R.T. By, Organisational change management: A critical review, Journal of Change Management. 5 (2005) 369–380.

[17] M.M. Caldeira, J.M. Ward, Using resource-based theory to interpret the successful adoption and use of information systems and technology in manufacturing small and medium-sized enterprises, European Journal of Information Systems. 12 (2003) 127–141.

[18] O. Çela, A. Front, D. Rieu, CEFOP: A method for the Continual Evolution of Organisational Processes, in: 2017 11th International Conference on Research Challenges in Information Science (RCIS), IEEE, 2017: pp. 33–43.

[19] A. Corallo, M. Lazoi, F. Striani, Process mining and industrial applications: A systematic literature review, Knowledge and Process Management. 27 (2020) 225–233.

[20] D. Dakic, S. Sladojevic, T. Lolic, D. Stefanovic, Process mining possibilities and challenges: a case study, in: 2019 IEEE 17th International Symposium on Intelligent Systems and Informatics (SISY), IEEE, 2019: pp. 161–166.

[21] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly. (1989) 319–340.

[22] P. Drakoulogkonas, D. Apostolou, On the selection of process mining tools, Electronics (Basel). 10 (2021) 451.

[23] M.L. Van Eck, X. Lu, S.J.J. Leemans, W.M.P. Van Der Aalst, PM $$^ 2$$: a process mining project methodology, in: International Conference on Advanced Information Systems Engineering, Springer, 2015: pp. 297–313.

[24] R. Eden, R. Syed, S.J.J. Leemans, J.A.C.M. Buijs, A Case Study of Inconsistency in Process Mining Use: Implications for the Theory of Effective Use, in: International Conference on Business Process Management, Springer, 2021: pp. 363–379.

[25] J. Eggers, A. Hein, Turning Big Data into Value: A Literature Review on Business Value Realization from Process Mining., in: ECIS, 2020.

[26] M. Eggert, J. Dyong, Applying process mining in small and medium sized IT enterprises–challenges and guidelines, in: International Conference on Business Process Management, Springer, 2022: pp. 125–142.

[27] P. Feldbacher, P. Suppan, C. Schweiger, R. Singer, Business process management: A survey among small and medium sized enterprises, in: S-BPM ONE-Learning by Doing-Doing by Learning: Third International Conference, S-BPM ONE 2011, Ingolstadt, Germany, September 29-30, 2011. Proceedings 3, Springer, 2011: pp. 296–312.

[28] P. Gering, Digitalisation toolkit for SMEs, CEUR Workshop Proc. (2020).

[29] T. Graafmans, O. Turetken, H. Poppelaars, D. Fahland, Process mining for six sigma: a guideline and tool support, Business & Information Systems Engineering. 63 (2021) 277–300.

[30] T. Grisold, J. Mendling, M. Otto, J. vom Brocke, Adoption, use and management of process mining in practice, Business Process Management Journal. (2020).

[31] A. Guzzo, A. Rullo, E. Vocaturo, Process mining applications in the healthcare domain: A comprehensive review, Wiley Interdiscip Rev Data Min Knowl Discov. 12 (2022) e1442.

[32] M. Heidt, J.P. Gerlach, P. Buxmann, Investigating the security divide between SME and large companies: How SME characteristics influence organizational IT security investments, Information Systems Frontiers. 21 (2019) 1285–1305.

[33] T.H.C. Van der Heijden, Process mining project methodology: developing a general approach to apply process mining in practice, Master of Science in Operations Management and Logistics. Netherlands: TUE. School of Industrial Engineering. (2012).

[34] A. Janes, Test case generation and prioritization: a process-mining approach, in: 2017 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW), IEEE, 2017: pp. 38–39.

[35] M. Kebede, M. Dumas, Comparative evaluation of process mining tools, University of Tartu. (2015).

[36] T. Lauer, Change management, Springer, 2010.

[37] J. Lee, S. Lee, J. Kim, I. Choi, Dynamic human resource selection for business process exceptions, Knowledge and Process Management. 26 (2019) 23–31.

[38] M. de Leoni, W.M.P. van der Aalst, M. Dees, A general framework for correlating business process characteristics, in: Business Process Management: 12th International Conference, BPM 2014, Haifa, Israel, September 7-11, 2014. Proceedings 12, Springer, 2014: pp. 250–266.

[39] R. Lorenz, J. Senoner, W. Sihn, T. Netland, Using process mining to improve productivity in make-to-stock manufacturing, Int J Prod Res. 59 (2021) 4869–4880.

[40] N. Mackenzie, S. Knipe, Research dilemmas: Paradigms, methods and methodology., Issues in Educational Research. 16 (2006) 193–205.

[41] N. Martin, D.A. Fischer, G.D. Kerpedzhiev, K. Goel, S.J.J. Leemans, M. Röglinger, W.M.P. van der Aalst, M. Dumas, M. la Rosa, M.T. Wynn, Opportunities and challenges for process mining in organizations: results of a Delphi study, Business & Information Systems Engineering. 63 (2021) 511–527.

[42] S.L.T. McGregor, J.A. Murnane, Paradigm, methodology and method: Intellectual integrity in consumer scholarship, Int J Consum Stud. 34 (2010) 419–427.

[43] N. Nguyen Chan, D.L. Nguyen Vo, C. Pham-Nguyen, T. le Dinh, N.A.K. Dam, T.T. Pham Thi, M.H. Vu Thi, Design and deployment of a customer journey management system: the CJMA approach, in: The 5th International Conference on Future Networks & Distributed Systems, 2021: pp. 8–16.

[44] J.S. Oakland, S. Tanner, Successful change management, Total Quality Management & Business Excellence. 18 (2007) 1–19.

[45] Z. Paszkiewicz, W. Cellary, Computer-Supported Collaboration of SMEs in Transnational Market, Journal of Transnational Management. 17 (2012) 294–313.

[46] L. Raymond, Organizational characteristics and MIS success in the context of small business, MIS Quarterly. (1985) 37–52.

[47] H. Rbigui, C. Cho, Purchasing process analysis with process mining of a heavy manufacturing industry, in: 2018 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, 2018: pp. 495–498.

[48] L. Reinkemeyer, Process mining in action, Process Mining in Action Principles, Use Cases and Outloook. (2020).

[49] M. Ronny, R. Hajo, B. Hans, B. Wasana, P. Rogier, Business process mining success, in: 21st European Conference on Information Systems, ECIS 2013, 2013.

[50] A. Rozinat, A.K.A. de Medeiros, C.W. Günther, A. Weijters, W.M.P. van der Aalst, Towards an evaluation framework for process mining algorithms, BPM Center Report BPM-07-06, BPMcenter. Org. 123 (2007) 142.

[51] S. Siccardi, C. Sebastiani, Business Process Measurement in small enterprises after the installation of an ERP software., in: SIMPDA, Citeseer, 2014: pp. 46–59.

[52] K. Smit, J. Mens, Process mining in the rail industry: A qualitative analysis of success factors and remaining challenges, (2019).

[53] S. Stephan, J. Lahann, P. Fettke, A Case Study on the Application of Process Mining in Combination with Journal Entry Tests for Financial Auditing, (2021).

[54] F. Stertz, J. Mangler, B. Scheibel, S. Rinderle-Ma, Expectations vs. experiences–process mining in small and medium sized manufacturing companies, in: International Conference on Business Process Management, Springer, 2021: pp. 195–211.

[55] R. Syed, S.J.J. Leemans, R. Eden, J.A.C.M. Buijs, Process mining adoption, in: International Conference on Business Process Management, Springer, 2020: pp. 229–245.

[56] M. Thiede, D. Fuerstenau, A.P.B. Barquet, How is process mining technology used by organizations? A systematic literature review of empirical studies, Business Process Management Journal. (2018).

[57] R.J. Wieringa, R.J. Wieringa, What is design science?, Design Science Methodology for Information Systems and Software Engineering. (2014) 3–11.

[58] F. Zerbato, P. Soffer, B. Weber, Process Mining Practices: Evidence from Interviews, in: International Conference on Business Process Management, Springer, 2022: pp. 268–285.

[59] P. Zerbino, A. Stefanini, D. Aloini, Process science in action: A literature review on process mining in business management, Technol Forecast Soc Change. 172 (2021) 121021.

[60] E. Zuidema-Tempel, R. Effing, J. van Hillegersberg, Bridging the Gap between Process Mining Methodologies and Process Mining Practices, in: International Conference on Business Process Management, Springer, 2022: pp. 70–86.

[61] Institut für Mittelstandsforschung Bonn: SME definition of the IfM Bonn, Https://Www.Ifm-Bonn.Org/En/Index/Definitions/Sme-Definition-of-the-Ifm-Bonn. (2022).

[62] EU recommendation 2003/361, Https://Eur-Lex.Europa.Eu/Legal-Content/EN/TXT/?Uri=CELEX%3A32003H0361. (2023).

# Appendices

## A Executive Summary

Process mining (PM) is a growing field of research that focusses on understanding and improving business processes. It relates the fields of process science and data science by mining data from information systems to investigate processes. The growing interest can be explained by the constant increase in the amount of data that is recorded in information systems. The goal of PM is to transform data into insights and actions. Various PM tools exist to support this transformation. These tools require event data as input and produce process models. Several tools also provide the possibility to create dashboards that can show graphs, charts, and other visuals to depict the data.

Three types of PM techniques can be applied in a PM project. First of all, process discovery can be used to gain insight into the current process. Secondly, conformance checking can be performed by comparing the current process to the desired process to identify bottlenecks. Lastly, enhancement techniques may be adopted to optimize the process by providing additional information. By applying these PM techniques in an organization, business values can be obtained in terms of monetary values, non-monetary values and process efficiency.

While PM can bring much value to organizations, it is unclear how organizations should apply PM to generate business values. It appears that practitioners miss guidance in their PM activities. Specifically, research on how PM can be applied in small to medium sized organizations (SMEs) is limited. Thus, research on the use of PM in SMEs is needed. The current research addresses this gap by proposing a methodology, PROMISE, on the use of PM in SMEs. PROMISE is an overall approach on how to start with PM in SMEs, including practical guidelines. PROMISE is directed towards practitioners who have a basic level of knowledge, experience, and skills in the field of data and business processes.

To evaluate the existing empirical evidence on PM in SMEs, a systematic literature review (SLR) was conducted. Following this, requirements for the methodology were formulated, and a first version of PROMISE was developed using literature on PM in SMEs and existing PM methodologies. This initial version of PROMISE was applied in a case study at eMagiz, an SME in Enschede, the Netherlands. After the case study, PROMISE was refined based on the findings. Thereafter, an additional case study was performed at Support, a supporting unit of eMagiz, using the refined methodology. The purpose of this second case study was to validate PROMISE. To achieve this, the methodology was also evaluated by experts in the field of PM methodologies, as well as practitioners experienced with data and business processes. The validation was completed by validating the methodology against the requirements formulated in this study.

PROMISE includes a visualization, depicting all phases of the methodology, pillars for each phase, and specific steps that have to be followed. The visualization is circular to emphasize the importance of PM being a continuous activity. The pillars define the purpose and deliverables for each phase, and the steps outline the specific PM activities that have to be executed. The results from the case studies have been described in this study as well, which can give some additional support to understanding the methodology.

In conclusion, the methodology developed in this study bridges the aforementioned gap by describing all PM activities that should be conducted when applying PM in an SME. PROMISE was evaluated to be useful and understandable by experts and practitioners, but

practitioners indicated that technical help would be appreciated when applying the methodology. For future research, PROMISE may be applied at SMEs other than eMagiz. Moreover, the methodology could be applied in PM projects by other practitioners to evaluate its practicality further.

# B Quality Scores

*Table 23 SLR Quality Scores*

| REFERENCE | QC1 | QC2 | QC3 | QC4 | QC5 | QC6 | QC7 | TOTAL SCORE |
|---|---|---|---|---|---|---|---|---|
| [53] | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 5 |
| [51] | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 2 |
| [45] | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 4 |
| [28] | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| [6] | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| [11] | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| [13] | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 3 |
| [37] | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 3 |
| [43] | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 3 |
| [34] | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| [18] | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 5 |
| [54] | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 4 |
| [26] | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 6 |
| [47]* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 2 |
| [39]* | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 5 |
| [52]* | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 3 |
| [20]* | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 4 |
| [30]* | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 5 |
| [41]* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 6 |
| [49]* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| [55]* | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 5 |
| [25]* | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 3 |
| [24]* | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 4 |

## C Overview of Results

*Table 24 SLR Overview of Results*

| REFERENCE | TYPE OF RESULT | EMPIRICAL EVIDENCE | EMPIRICAL SETTING | PURPOSE OF STUDY |
|---|---|---|---|---|
| [53] | Implementation | Case study | German medium-sized audit firm | Effectively combine journal entry tests with process mining to capture a more comprehensive view within a company's audit. |
| [51] | Analysis | Case study | 6 small Italian enterprises | Get a better understanding of the processes and a rationalization of the operations after the installation of an ERP software. |
| [45] | Method | Illustrative example | Partner selection process | Develop a method for supporting collaborative processes based on process mining techniques. |
| [6] | Reflection | None | None | Reflect on the current status of the process mining discipline and provide an outlook on upcoming developments and challenges. |
| [11] | Analysis | Case study | Small company located in Italy | Investigate how users interact with an enterprise resource planning software using process mining. |
| [13] | Framework | Case study | 5 Chilean SSEs (small software enterprises) | Develop a framework that allows the use of process mining techniques in families of processes within the software development domain (i.e., Software Process Lines). |
| [37] | Implementation | Case study and simulation | Purchasing process (not clear what data?) | Propose a systematic approach that analyses event logs to select |

| | | | | |
|---|---|---|---|---|
| | | | | suitable substitutes if the initial human resources become unavailable. |
| [43] | System | Case study | Dataset from Google Merchandise Store | Present an approach for designing and deploying a customer journey management system. |
| [18] | Method | Case study | French start-up | Present a method for analysing, diagnosing and evolving the organisational processes. |
| [54] | Analysis | Focus group study | 2 SSMCs (small and medium sized manufacturing companies) | Present expectations on and experiences with the introduction of process mining in SMMC, including employees in different positions, e.g., process supervisors and shopfloor workers, and exposure to process mining. |
| [26] | Analysis | Case study | Medium-sized IT vendor in Germany | Investigate the application of process mining and shed light into the particular challenges of an IT SME. |
| [47]* | Analysis | Case study | Manufacturing company (no size info) | Present an industrial application of process mining in a real purchasing process of a heavy manufacturing industry. |
| [39]* | Procedure | Case study | Large manufacturing company | Propose a data-driven procedure to improve productivity in make-to-stock manufacturing. |
| [52]* | Analysis | Case study | Largest rail organization in the Netherlands | Identify success factors and remaining challenges relevant to the practice of process mining in the rail industry. |

| | | | | |
|---|---|---|---|---|
| [20]* | Methodology | Case study | Dataset from 4TU.Centre for Research Data repository | Create a methodology to meet the challenge of the lack of representative benchmarks and process mining methodologies. |
| [30]* | Analysis | Focus group study | 22 participants from workshop organized by Signavio | Explore how process managers perceive the adoption, use and management of process mining in practice. |
| [41]* | Analysis | Delphi study | 40 international experts from academia and industry | Provide a holistic view of opportunities and challenges for process mining in organizations. |
| [49]* | Model | Case study | 4 industrial process mining projects (different, large companies) | Advance a model, tailored to the characteristics of process mining projects, which identifies and relates success factors and measures. |
| [55]* | Analysis | Case study | Large provider of services to pension funds in the Netherlands | Report findings on an exploratory case study of the early stages of the adoption of process mining. |
| [25]* | Analysis | None | None | Synthesize the existing knowledge on business value realization from process mining. |
| [24]* | Analysis | Case study | Large provider of services to pension funds in the Netherlands | Understand the effective use of business intelligence systems, specifically process mining. |

## D PROMISE Phases & Guidelines

**Phase 1: Business Understanding**

**G1**: Verify findings after each phase and move to previous phases if required.

**G2**: Move across phases quickly to present an MVP as soon as possible.

**G16**: Ensure that all who are involved with steps of the PM project have a basic understanding of PM.

**G3**: Involve top management as well as employees that have knowledge about existing data and processes.

**G4**: Convince the company of the importance of PM, e.g., by providing success stories of PM on example processes.

**G5**: It needs to be established who should be involved in PM activities and who is responsible for managing them, taking into account the continuity of PM.

**G6**: Involve data protection stakeholders.

**G7**: Ensure that employees agree on a transparent data policy.

**Phase 2: Plan, Scope & Define**

**G8**: Link the strategy of the company to the PM goals.

**G9**: Check whether the company is able to give enough data and commitment.

**G13**: Start with a question-driven project, or, if it is clear how to use PM to achieve a certain goal, a goal-driven project may be set up.

**G11**: Begin with a simple process, with a minimum of 3 process steps.

**G12**: Select a process by considering process properties, stakeholders and strategic goals.

**G14**: Consider the types of PM techniques that are needed with respect to the PM project goals and scope.

**G15**: Formulate KPIs for a goal-driven project and formulate questions for a question-driven project.

**G10**: Choose a suitable PM tool by considering the core functionalities of existing tools and the desired functionalities.

**Phase 3: Data Exploration & Understanding**

**G17**: Make sure that data is available, and that the data can be extracted.

**G18**: Ensure that the data contains a case ID, activity, and timestamp.

**G21**: Ensure that the selected data can help in answering the research questions.

**G19**: Check whether the data quality is sufficient and strive for the highest possible quality level of event logs.

**G20**: Find the right balance for the number of events in the data set.

**Phase 4: Event Log Creation**

**G22**: Take time to choose an appropriate case ID, ensuring to evaluate all possible case IDs.

**G23**: Select a case ID, activity and timestamp by taking into account the questions that should be answered by the PM analysis and ensuring that the selected timestamp belongs to the selected activity.

**G24**: Extract the data once the required dataset has been constructed.

**G25**: Prepare the extracted dataset so that it is suitable for further processing.

**G26**: Apply filtering if it is needed to reduce complexity or to focus on a specific part of the dataset.

**Phase 5: Process Model Creation**

**G27**: Create a process model with the chosen PM tool using its desired algorithms and techniques.

**G28**: Ensure that the process models highlight the aspects that are relevant for a particular type of user and take into account differing process versions.

**G29**: Create a comprehensive knowledge base with involved stakeholders for conformance checking.

**G30**: Create PM dashboards to gain insight into the results from the PM activities.

**Phase 6: Analysis & Enhancement**

**G31**: Conduct analyses in close collaboration with process analysts and business experts in a highly iterative and interactive manner.

**G32**: Make sure to compare the discovered process to the desired process, check for performance indicators and bottlenecks, and analyse the relationship between resources and activities.

**G33**: Enhance the process model by adding e.g., additional visual analytics and digital animations.

**Phase 7: Evaluation**

**G34**: Check whether the results are interpreted correctly and draw conclusions with respect to the research questions.

**G35**: Suggest actions for improvements and think about other possible elaborations of PM in the organization.

**G36**: Verify the PM results by investigating the correctness of the findings, e.g., by checking the soundness of the models.

**G37**: Validate the results by checking the degree of fitness, precision, generalization, and simplicity of the model.

**G38**: Evaluate the degree in which the results meet the PM project objectives.

**Phase 8: Process Improvement & Presenting**

**G39**: Define improvement actions concerning one or more of the following activities: redesign, adjust, intervene, and support.

**G40**: Consider the integration of event logs with systems and the distribution of event logs through a centralized portal.

**G41**: Prioritise improvement actions and decide on the manner in which improvements will be monitored.

**G42**: Communicate the recommendations, predictions, and other results to the involved stakeholders.

**Phase 9: Change Management**

**G43**: Ensure that PM is a continuous activity in the organization, e.g., by elaborating the PM project, following up on improvement actions, and measuring improvements.

**G44**: Implement change management to ensure that improvements are realized.

## E PROMISE Steps

**For all phases:**

1. Verify the findings after each phase and move to previous phases if required.

2. Move across phases quickly to present an MVP as soon as possible.

**Phase 1: Business Understanding**

1. Make sure to have an understanding of the company.
2. Convince the company, especially top management, of the importance of PM, e.g., by providing success stories of PM on example processes.
3. Make sure that employees have a basic understanding of PM.
4. Investigate what processes could be analysed.
5. Investigate what data could be analysed.

**Phase 2: Plan, Scope & Define**

6. Define the goal of the PM project, including the benefits it may bring to the company.
7. Investigate processes and data further.
    a. Find a case ID, activity, and timestamp. Make sure that these are valid, e.g., the selected timestamp should belong to the selected activity.
    b. Make sure that the data is available, and that the data can be extracted.
8. Choose which process to study, as well as the data that defines the process. Begin with a simple process, with a minimum of three types of activities.
9. Check whether the company is able to give enough data and commitment.
10. Establish who should be involved in the PM activities and who is responsible for managing them, taking into account the continuity of PM.
11. Involve data protection stakeholders.
12. Make sure that employees agree on a transparent data policy.
13. Formulate research questions, and possibly KPIs.

**Phase 3: Data Exploration & Understanding**

14. Determine which types of PM techniques will be used taking into account the PM project goals and the scope.
15. Determine which PM tool(s) will be used by considering the core functionalities of existing tools and the desired functionalities.
16. Check whether the data quality is sufficient and strive for the highest possible quality of event logs.
    a. Evaluate its trustworthiness.
    b. Evaluate its completeness.
    c. Evaluate its semantics.
    d. Evaluate its safeness.

**Phase 4: Event Log Creation**

17. Define the case ID, activity, and timestamp.
18. Extract the data.
19. Prepare the extracted dataset so that it is suitable for further processing.
20. Filter the data if it is needed to reduce complexity or to focus on a specific part of the dataset.
21. Find the right balance for the number of events in the data set.

**Phase 5: Process Model Creation**

22. Create a process model with the chosen PM tool using its desired algorithms and techniques.
23. Ensure that the process models highlight the aspects that are relevant for a particular type of user and take into account differing process versions.
24. Do conformance checking, making sure that reference models are well documented.
25. Create PM dashboards to gain insight into the results from the PM activities.

**Phase 6: Analysis & Enhancement**

26. Analyse the findings from the PM activities with process experts and PM experts in a highly iterative and interactive manner.
    a. Compare the discovered process to the desired process.
    b. Check for performance indicators and bottlenecks.
    c. Analyse the relationship between resources and activities.
27. Enhance the process model(s) or dashboard(s) by adding e.g., additional visual analytics and digital animations.

**Phase 7: Evaluation**

28. Check whether the results are interpreted correctly and draw conclusions with respect to the research questions.
29. Verify the results, by checking whether
    a. Each activity has a next activity or is connected to the end place,
    b. Each activity has a former activity or is connected to the start place,
    c. All activities between the start and end activity have the same frequencies for in- and outgoing arrows.
30. Validate the results, by evaluating the degree of fitness, precision, generalization, and simplicity.
31. Evaluate the degree in which the results meet the PM project objective.

**Phase 8: Process Improvement & Presenting**

32. Suggest actions for improvements and think about other possible elaborations of PM in the organization.
33. Define improvement actions concerning one or more of the following activities: redesign, adjust, intervene, and support.
34. Prioritise improvement actions and decide on the manner in which improvements will be monitored.
35. Consider the integration of event logs with systems and the distribution of event logs through a centralized portal.
36. Communicate the recommendations, predictions, and other results to the involved stakeholders.

**Phase 9: Change Management**

37. Ensure that PM is a continuous activity in the organization, e.g., by elaborating the PM project, following up on improvement actions, and measuring improvements.
38. Implement change management practices to ensure that improvements are realized.

# F User Guide

This user guide is meant for any practitioner who would like to start with process mining (PM) in a small to medium sized organization (SME), and has a basic level of knowledge, experience, and skills in the field of data and business processes. The goal of this user guide is to enhance the understandability of the PM methodology developed in this study, PROMISE, such that it can be applied most efficiently.

PM is about understanding and improving business processes and bridges the gap between process science and data science. PM tools are available to support PM activities by turning data into process models. This can be achieved if the data contains a case ID, activity, and timestamp. A case ID is a unique identifier of a process instances, activities are task or operations that belong to the case, and the timestamps belong to the activities. For example, a case ID may be an order ID, activities could be 'order placed', 'order received', 'order shipped', and the timestamps are the times that an order is placed, received, and shipped.

Three types of PM techniques exist: discovery, conformance, and enhancement. Discovery techniques are used to create a process model based on data. Conformance techniques provide the possibility to compare the generated process model with a documented model. Enhancement techniques are about providing additional information to improve an existing process model.

The PROMISE visualization is given in *Figure 26 Refined PROMISE Visualization Enhanced*. Additional information about the methodology, explaining the purpose, steps to be taken (indicated with a capital S) and deliverables for each phase, is given in *Figure 27 PROMISE Pillars*. The steps that need to be taken are given in Appendix *E PROMISE Steps*.

PROMISE consists of 9 phases, starting with Business Understanding. The phase Change Management is the last phase of a PM project. One of three directions may be taken after the Change Management phase, namely (i) the PM activities may be terminated, e.g. if not enough resources are available for it anymore, (ii) the Business Understanding phase may be entered, only if the strategy of the company has changed, such that the Business Understanding needs to be adapted, and (iii) the Plan, Scope & Define phase may be entered, to set up a new PM project.

Depending on the company's resources, it may be decided to apply only one or two types of PM techniques. The circle within the PROMISE visualization indicates which PM technique is addressed in each of the phases. So, in case a particular PM technique has been chosen, only those phases addressing that technique have to be followed. E.g., if only discovery techniques should be applied, the first four phases, and partly the fifth phase, have to be followed.

The Chapter *7 Treatment Validation* provides a complete workout of a case study in which PROMISE is applied. This chapter may help to understand the methodology more thoroughly.

## G Evaluation Results Experts

| Reference | Type of Participant | Country | Current Position |
|---|---|---|---|
| **E1** | Expert | The Netherlands | EngD candidate & researcher on PM, author of [60] |
| **E2** | Expert | The Netherlands | BI Consultant |
| **E3** | Expert | The Netherlands | BI Consultant |

| Element | Code | Statement | Rate<br>*1 = completely disagree*<br>*2 = disagree*<br>*3 = neutral*<br>*4 = agree*<br>*5 = completely agree* | Comments<br>*If you would like to add something, please add it here.* |
|---|---|---|---|---|
| Perceived Usefulness | PU1 | I think that PROMISE covers all necessary PM activities to set up a PM project. | 5 5 4 | See feedback per email regarding V&V, evaluation and team composition. |
| | PU2 | I think that PROMISE can help practitioners to start up a PM project. | 4 4 5 | Certainly 'agree', but in practice it is still quite difficult for non-data driven employees to get an idea of what data they need here. You explain the variables well, but in practice the average person encounters this.<br>I think that it is very theoretical and that it might help to put it more into practice.<br>The methodology is clear and understandable.<br>Extra tip: you write about business understanding but how can companies realize this? Maybe via workshops, courses, or… |
| | PU3 | I think that PROMISE can help to better understand the PM activities that are needed to start up a PM project. | 5 5 4 | Agree |
| | PU4 | I think that PROMISE gives good practical guidance. | 3 3 4 | You do this well when explaining the variables, but I think you could do this more.<br>The process as described within the methodology is the "perfect" process. What |

| | | | | |
|---|---|---|---|---|
| | | | | challenges do you face when putting it into practice? What are pitfalls?<br>Yes, see above |
| Perceived Ease of Use | PEU1 | I think that the PROMISE steps are clear and easy to understand. | 5 5 5 | The descriptions and steps are a good addition to the visualization.<br>Every step is clear and well explained |
| | PEU2 | I think that the PROMISE visualization is clear and easy to understand. | 4 5 4 | The steps are all logical.<br>Tip: the colours are all shades of purple, which might be a bit difficult to distinguish. |
| | PEU3 | I think that the PROMISE pillars are clear and easy to understand. | 5 5 5 | Very clear |
| | PEU4 | I would be able to start a PM project based on PROMISE. | 5 4 4 | I can certainly extract valuable information from this, but this is still complicated for an average purchasing or sales department.<br>Yes, see comments that I have in the email and in this feedback doc. |
| | PEU5 | I think that I would **not** need technical help to use PROMISE. | 3 4 3 | This is what I was referring to earlier, I think you give any company a lot of good tools to get started with PM in practice, but in terms of data, choosing and using a tool, they will generally still need quite a bit of support.<br>Depending on the data available and the ease of distracting data from the systems. Thereby, you will need tooling to define the visualization and create an analysis based on that.<br>This depends, I have knowledge about process mining dashboarding, but process mining often involves a combi of an analyst and a business process owner. I |

| | | | | |
|---|---|---|---|---|
| | | | | consider myself to be the process owner. An analyst is needed to help people that do not know much about PM is recommended |
| | PEU6 | Overall, I think that PROMISE is easy to use. | 4 4 4 | The methodology by itself is.<br>Yes |
| Intention to Use | ITU1 | I would use PROMISE when executing a PM project. | 5 3 3 | I can definitely gain useful insights from this.<br>Maybe, maybe not. Depending on the project and data available.<br>Less relevant for me |
| | ITU2 | I would refer to PROMISE if I got lost in a PM project. | 4 2 4 | From the practical side, I would rather refer to previous projects and base my next steps based on that. To give others an understanding of the material, I could refer to the methodology.<br>Yes, it serves as a guideline. |

## H Evaluation Results Practitioners

| Reference | Type of Participant | Country | Current Position |
|---|---|---|---|
| **P1** | Practitioner | The Netherlands | CTO |
| **P2** | Practitioner | The Netherlands | Support Technical Contact |
| **P3** | Practitioner | The Netherlands | Business Consultant |
| **P4** | Practitioner | The Netherlands | Manager Expert Services |
| **P5** | Practitioner | The Netherlands | Team Lead |

| Element | Code | Statement | Rate<br>*1 = completely disagree*<br>*2 = disagree*<br>*3 = neutral*<br>*4 = agree*<br>*5 = completely agree* | Comments<br>*If you would like to add something, please add it here.* |
|---|---|---|---|---|
| Perceived Usefulness | PU1 | I think that I would be able to start up a PM project based on PROMISE. | 4 4 4 3 4 | Instructions are clear and make it seem easy enough.<br>I think I could not start alone, but gives a good summary and starting point.<br>Starting up a project would be possible, however don't know how realistic it is to finish it |
| | PU2 | I think that PROMISE can help to better understand the PM activities that are needed to start up a PM project. | 5 5 4 4 5 | |
| | PU3 | I think that PROMISE gives good practical guidance. | 4 3 4 4 3 | But I had to also read the use case, to better understand it. |
| Perceived Ease of Use | PEU1 | I think that the PROMISE steps are clear and easy to understand. | 4 5 4 4 4 | |
| | PEU2 | I think that the PROMISE visualization is clear and easy to understand. | 4 4 4 4 4 | |
| | PEU3 | I think that the PROMISE pillars are clear and easy to understand. | 4 5 4 4 3 | Depends really on role, level and experience of person reading it |

| | | | | |
|---|---|---|---|---|
| | PEU4 | I would be able to start a PM project based on PROMISE. | 4 4 3 3 4 | I'm not that smart<br>The part of data analysis needs help from someone that knows the tech part. Is this not the same as PU1? |
| | PEU5 | I think that I would **not** need technical help to use PROMISE. | 2 2 2 5 3 | Feel like it would be possible but it would make it much easier<br>Don't know, cannot oversee this |
| | PEU6 | Overall, I think that PROMISE is easy to use. | 4 4 4 4 3 | I estimate it will not be easy for most non-experts |
| Intention to Use | ITU1 | I would use PROMISE when executing a PM project. | 5 4 4 3 5 | I'm interested to have a further look in the future, since I'm working to define/improve process at eMagiz. When to know that you need to start a PM project is the challenge also.<br>For sure, this methodology gives me a direction and without it I would not be able to start smoothly |
| | ITU2 | I would refer to PROMISE if I got lost in a PM project. | 5 5 4 4 4 | Yes, very useful to have some clear documentation<br>As a starting point seems very good, is there anything similar if you search on internet? |

**Additional Comments:**

I think I would struggle the most with the practical application of the first two phases; convincing people and understanding the company is easier said than done and some more pointers how to approach that would certainly be welcome.

The case study helped me quite a bit, so I gave scores based on both the methodology and the case study documents.