The α and ω of Process Mining

Vinicius Stein Dani

The *Alpha* and *Omega* of Process Mining

Vinicius Stein Dani

SIKS Dissertation Series No. 2024-32 The research reported in this thesis has been carried out under auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

The Alpha and Omega of Process Mining
ISBN: 978-94-6510-042-5
DOI: 10.33540/2528

Druk: ProefschriftMaken
Cover image: the author

The *Alpha* and *Omega* of Process Mining

De *Alfa* en *Omega* van Process Mining

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op donderdag 21 november 2024 des ochtends te 10.15 uur door

Vinicius Stein Dani

geboren op 26 maart 1984
te Porto Alegre, Rio Grande do Sul, Brazilië

Promotoren:
Prof. dr. ir. Hajo A. Reijers
Prof. dr. Henrik Leopold

Co-promotor:
Dr. ir. Jan Martijn E. M. van der Werf

Beoordelingscommissie:
Prof. dr. Fabiano Dalpiaz
Dr. Kathrin Figl
Prof. dr. Mieke Jans
Prof. dr. Manuel Resinas
Prof. dr. Oktay Türetken

Dit proefschrift werd (mede) mogelijk gemaakt door financiële steun van NWO in het AutoDrive project [project code 16672]

Back off, man. I'm a scientist.

Peter Venkman.

I'm not superstitious, but
I am a little stitious.

Michael Scott.

No matter how thin you slice it,
there will always be two sides.

Baruch Spinoza.

I'll be back.

The Terminator.

Acknowledgements

I must admit that I do not have enough words (not even in Portuguese!) to thank my dream team of promotors: Hajo Alexander Reijers, Henrik Leopold, and Jan Martinus Evert Maria van der Werf (also known as Jan Martijn). Their knowledge, wisdom, patience, and support have been invaluable in guiding me through the process of working on and writing this thesis and in growing both professionally and personally. I extend my thanks to Henrik, whom I for sure bothered the most during these years; to Hajo, for pushing me mentally (and physically during our fitness sessions!); and to Jan Martijn, for sharing his love for trains and for indirectly encouraging me to continue nurturing my passion for music production. Thank you, guys, for this opportunity, for believing in my potential, and for being there for me along this journey!

I would also like to express my gratitude to the friends I made in the Utrecht Process Science research group: Iris Beerepoot, Xixi Lu, Ünal Aksu, Jelmer Koorn, Inge van der Weerd, Suhwan Lee (special thanks for the life-related philosophical discussions at 't Oude Pothuys), Wouter van der Waal, Nico Brand, and Jens Gulden. Whether directly or indirectly, you all have contributed to my achievements during these years. I extend a special thanks to Iris and Xixi; alongside my promotors, they are role models for the type of scientist I aspire to be one day. Additionally, I am thankful to Mahendra ER and Marcus Dees for our collaborations, to my BSc, MBA, and MSc supervisors for their guidance that led me toward pursuing a PhD, and to the assessment committee of this thesis for the invaluable feedback in shaping it.

I am also grateful to my closest family: Delcio and Adely, Amadeus, Paulo and Elenir, Francine and Ariel†, and Aunt Adriana and Mariana. A special thanks goes to my wife, Francine, for her loving kindness and affection, and for teaching me the importance of finding the positive aspects in everything. I would also like to thank my friends Gustavo, Rodrigo, Rogers, Cleber, Danny, Melina, and Júnior for our online meetings. Additionally, I thank Bira, Nilton and Ary, Gustavo and Simone, Luciano and Leise, Pedro and Gabi, Mateus and Ariele, Patrick and Marília, and Antoine and Dani for the in-person contact in the cinematographically beautiful Netherlands over these years.

As I always say, although I am not a religious person, I believe that there is some sort of a “divine” force, usually referred to as “God”, and to whoever or whatever “s/he” is, I dedicate my deepest gratitude, particularly for allowing me to meet all of these (and many more) amazing people along this journey! Last but not least, I also thank myself, and I must say that I am very proud of this thesis that you are about to (hopefully) read :)

Abstract

Process mining is a discipline that enables organizations to visualize, analyze, and improve their work processes. To leverage process mining capabilities, an organization first needs to prepare and extract its data in a specific format, the event log. An example of the work involved in event log extraction is the definition of the data sources from which the data will be extracted. Various methodologies and techniques to apply process mining in organizations have been developed, particularly regarding process discovery and analysis. After insights are acquired through these techniques, an organization needs to work towards translating the acquired insights into actual process improvements. However, these methodologies do not provide support in this last, but not least, step. What remains unclear is the work that it takes to go from data within the organization, i.e., before a process mining initiative can start, to process improvements, i.e., translating the results of the initiative into action. Therefore, we study the following main research question:

How effective is process mining in supporting organizations?

We answer this research question in two parts: the *alpha* (α) and *omega* (ω) of process mining. In part α, we systematically identify the variety of manual tasks still involved in event log extraction. We provide a view on the work that it takes to go from data generated by information systems into an event log, to enable process discovery and analysis. After an event log is generated, process mining tools and techniques for process discovery and analysis can be used to acquire insights about the work processes of the organization under scrutiny.

In part ω, we study the diverse space of actions and objects of actions (e.g., update document, increase resource, conduct root-cause analysis, etc.) triggered by process mining insights. This provides a view on the work that

it takes to go from the acquired insights to actual process improvements. Moreover, we unveil challenges and specific causes that hinder organizations in effectively adopting process mining to improve their work processes.

Combined, our findings provide researchers and practitioners in the process mining field with a knowledge base to (1) understand the variety of work required for carrying out a process mining initiative and (2) proactively counter challenges that may arise during the course of the initiative.

Samenvatting

Process mining stelt organisaties in staat om hun werkprocessen te visualiseren, analyseren en verbeteren. Om dit te kunnen realiseren moet een organisie eerst haar informatiesystemen en bijbehorende databronnen geschikt maken, om vervolgens de data te kunnen extraheren in een specifiek formaat voor process mining: het event log. Om een voorbeeld te geven, een organisatie moet eerst de databronnen definiëren, voordat extractie plaats kan vinden. In de afgelopen jaren zijn verschillende methoden en technieken ontwikkeld om process mining toe te passen. Deze methoden richten zich echter vooral op process mining zelf: het ontdekken van processen en het uitvoeren van procesanalyses. Hoe de resultaten ingezet binnen een organisatie moeten worden om tot verbeterprocessen te komen, is een essentiële stap die in deze methoden weinig tot geen aandacht krijgen.

Tot nu toe is het niet duidelijk hoeveel tijd en werk het kost om van databronnen tot procesverbeteringen te komen. Met andere woorden, het is niet bekend wat er allemaal bij komt kijken om een process mining-initatief te starten, noch wat er gedaan moet worden om inzichten te vertalen naar verbeterstappen. Daarom bestuderen we in deze dissertatie de volgende onderzoeksvraag:

Hoe effectief is process mining in het ondersteunen van organisaties?

We beantwoorden deze vraag in twee delen: de *alfa* (α) en *omega* (ω) van process mining. In deel α identificeren we systematisch de verscheidenheid aan handmatige taken die nodig zijn om event logs te kunnen extraheren. We tonen de stappen die nodig zijn om data, die vaak door verschillende informatiesystemen gegenereerd worden, te transformeren in een event log om process mining mogelijk te maken en de werkprocessen van de onderzochte organisatie inzichtelijk te maken.

In deel ω bestuderen we in hoeverre inzichten verkregen door process mining

worden opgepakt en leiden tot mogelijke verbeteringen, zoals het bijwerken van documenten, het vergroten van middelen, het uitvoeren van analyses van de hoofdoorzaken, etc. Dit onderzoek geeft inzicht in het werk dat nodig is om van verworven inzichten naar daadwerkelijke procesverbeteringen te komen. Bovendien identificeren we uitdagingen en belemmeringen die organisaties ervaren bij het effectief adopteren van process mining-technieken om hun werkprocessen te verbeteren.

De uitkomsten van deze dissertatie bieden onderzoekers in en gebruikers van process mining een basis om (1) de verscheidenheid aan werkzaamheden te begrijpen die nodig zijn voor het uitvoeren van een process mining-initiatief en (2) om proactief potentiële problemen die tijdens een process mining-initiatief kunnen ontstaan het hoofd te bieden of zelfs kunnen voorkomen om zo process mining effectiever te kunnen inzetten.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

Organizations typically have various information systems with different data sources. These systems support a variety of organizational processes. To differentiate themselves from their competitors, in terms of quality and speed of delivery of their service or product, organizations need to continuously improve their processes [36].

Process mining is a discipline that supports organizations in using their own data to extract insights about their work processes. The ultimate objective is to use these insights for process improvement [107, 133]. Process mining requires a so-called event log as input. Among others, such an event log enables organizations to create a process model of a considered work process and conduct process analysis. By using process mining, organizations can identify conformance indicators and answer different questions about their work processes [110]. Many different discovery and analysis techniques have been proposed, along with dedicated process mining methodologies for how to systematically obtain relevant insights using these discovery and analysis techniques [16, 138]. These methodologies typically focus on applying process mining techniques on a given event log and on reporting the insights for process improvement. Although these methodologies highlight the importance of progressing the acquired insights into process improvement, they nonetheless consider this is an aspect out of scope.

While there is evidence that process mining can support organizations to improve their processes [9, 67], there is also evidence indicating otherwise [86]. A variety of studies have investigated the application of process mining and associated challenges [37, 89, 149]. For example, in [149], the authors identified a set of challenges regarding process mining analysis, highlighting the need for enhanced support for acquiring process mining insights. While in [37], the authors focused on the link between process mining and process improvement by investigating how process mining can support improving process awareness in organizations. What remains an open question is how effective is process mining in supporting organizations?

Recognizing this, we need to understand the pipeline that leads from data within the organization to process improvements for the organization. If we understand this pipeline, its associated tasks and challenges, we can study the effectiveness of process mining techniques and tools. In this thesis, we want to study how organizations apply this pipeline to understand its effectiveness. We first explore the characteristics of such a process mining pipeline. Next, we present the research questions and research methods used to guide and structure the research conducted in this thesis. Then, we provide a reflection on the research contributions and a list of publications that have been used as a foundation for this thesis. Finally, we delineate the thesis outline.

1.1 Process Mining Pipeline

Typically, in applying process mining, three major stages can be distinguished. These are as highlighted in Figure 1.1. The first stage, which we refer to as the *alpha* (α) of process mining, concerns *event log extraction*. Once the event log is obtained, the second stage, the actual application of *process mining* starts. The third and last stage, which we refer to as the *omega* (ω) of process mining, concerns to *progressing the insights obtained from process mining results into process improvements*. Many studies explore methods and techniques for process discovery and analysis, i.e., they consider the α and ω of process

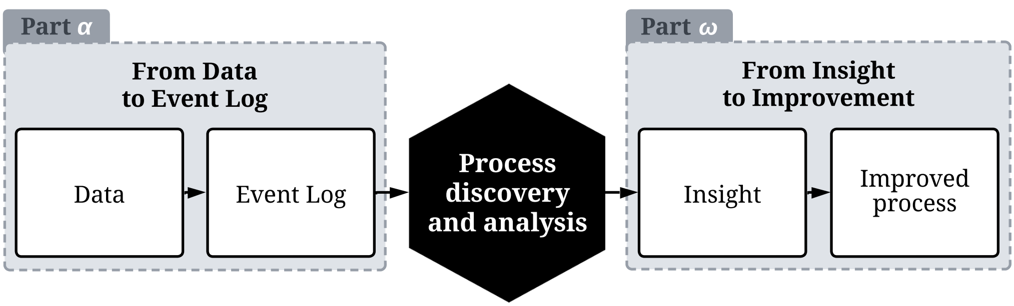


Figure 1.1: Overview of the pipeline from data to process improvement. Highlighted are the aspects we focus on this thesis.

mining as black boxes or out of scope. In this thesis, we turn it around: we focus on parts α and ω, and consider the application of process mining as a black box.

In Part α of this thesis, we turn the spotlight on the stage before process discovery and analysis techniques can be used: based on data generated by information systems, an organization needs to identify, extract, and make sense of event data in order to generate an event log. This involves different tasks, as we unveil in this thesis, and is time-consuming. After an event log is generated, widely available process mining tools and techniques for process discovery and analysis can be used to acquire insights about the work processes of the organization under scrutiny.

In Part ω of this thesis, we illuminate the stage after process discovery and analysis techniques are used: different insights can be unveiled based on process discovery and analysis and are expected to trigger process improvement initiatives, which is not always the case, as we show in this work.

1.2 Research Questions

Based on the introduction of this thesis, we derived the following main research question:

MRQ. How effective is process mining in supporting organizations?

Process mining supports organizations throughout the entire span of the process discovery and analysis endeavor, including before it starts, during its execution, and after it ends. Most process mining methods and techniques do not focus on extracting an event log (before) or checking whether the derived process mining insights lead to improvement (after). Thus, next, we present the questions we derived based on our main research question to better understand the work involved before and after process discovery and analysis occur.

Part α: to understand how effective process mining can be in the stage before process discovery and analysis, we need to understand the different work involved in event log extraction. Specifically, our focus is on the manual tasks involved. With this understanding, we can identify the potential for (semi-)automation of manual tasks, which could ultimately reduce the time and costs of an event log extraction. To do so, we derived the following two sub-research questions.

RQ1. What are the specific manual tasks that humans perform in the context of event log extraction?

RQ2. How to link a reference model and the underlying database to support the event log extraction?

Part ω: to understand how effective process mining can be in the stage after process discovery and analysis, we need to understand the work that is triggered by process mining insights. Specifically, we focus on triggered actions and its objects of action (e.g., update document, increase resource, conduct root-cause analysis, etc.). Also, we need to understand to which extent the triggered actions are recommended and performed. Thus, we can identify specific causes that lead to recommended actions not being performed and challenges involved in progressing process mining insights into process improvement. To do so, we derived the following three sub-research questions.

RQ3. What are the actions organizations can take towards process improvement?

RQ4. To what extent are recommended actions also performed?

RQ5. Which challenges do organizations face when translating process mining insights into process improvements?

1.3 Research Methods

In this thesis, we have used different qualitative research methods to gather and analyze data from both literature and practitioners in the field of process mining. In Table 1.1, we present an overview of the research methods employed in each chapter of this thesis along with a link to the research questions they address. Specifically, to address RQ1, we conducted a systematic literature review, which is reported in Chapter 2. To support answering RQ2, we conducted an exploratory literature review, reported in Chapter 3. For RQ3 and RQ4, we performed a second systematic literature review, with insights reported in Chapters 4 and 5. Additionally, to support answering both RQ4 and RQ5, we conducted semi-structured interviews, with results discussed in Chapters 5 and 6. To further support answering RQ5, we performed a second exploratory literature review, also outlined in Chapter 6. Across all Chapters, we used qualitative coding to systematically analyze and interpret data, which allowed us to identify underlying patterns that contributed to answering our research questions.

Table 1.1: Research methods and questions mapping to the thesis chapters.

Chapter	**Research questions**	**Research method** Literature review	Qualitative coding	Interview
2	RQ1	✓	✓	
3	RQ2	✓	✓	
4	RQ3	✓	✓	
5	RQ4	✓	✓	✓
6	RQ5	✓	✓	✓

1.4 Contributions

Our contributions can be organized into two main categories:

- the understanding and classification of the work involved in extracting event logs;
- the understanding and classification of the work involved in translating process mining insights into process improvement.

Next, we present our contributions in more detail.

Understanding and classification of the work involved in extracting event logs. Combining a structured literature review with qualitative data coding, we derived a *taxonomy of manual tasks in event log extraction*. With this taxonomy, we classified the main manual tasks that we have identified in event log extraction, answering RQ1. In addition, we proposed a first step towards *identifying a set of potential mappings between process model activities and database tables*, answering RQ2. With this approach, we believe, we can reduce the burden of the process analyst during the extraction of an event log, saving a considerable amount of manual work.

Our taxonomy can serve as input for future automation efforts and the enhancement of process mining methodologies, providing a comprehensive overview of human involvement in the event log extraction phase of process mining initiatives. In addition, this understanding can serve as a basis for future research on human-computer interaction aiming at designing more user-friendly event log extraction and process mining applications.

Our mapping approach of relational database and process model can potentially reduce the manual effort and technical expertise required to map database tables to process model activities. This approach can lead to more accessible and scalable process mining projects across different domains and organization sizes. In addition to that, it can serve as inspiration for future research exploring automation techniques in event log extraction in the context of process mining. We believe this work encourages further development of methodologies for (semi-)automated event log extraction and, potentially, advancing the integration of process mining with other information systems.

Understanding and classification of the work involved in translating process mining insights into process improvement. Once an event log is available, an organization can conduct process discovery and analysis. These tasks usually are performed with the intention of deriving insights for process improvement. In this context, we provided a classification of the intervention space of actions triggered by process mining insights, answering RQ3, and an understanding of the challenges to translate process mining insights into process improvement, answering RQ5. Based on these challenges, we derived a set of recommendations for process mining practitioners to consider when starting a process mining initiative. Finally, we also identified a set of specific causes why process mining initiatives are halted, answering RQ4. Thus, we believe the contributions of this thesis can serve as a basis for the enhancement of process mining methodologies to support both practitioners and researchers in process mining initiatives.

Our overview of the recommended and performed actions triggered by process mining insights can serve as basis to inform future research and practitioners. It can support new process mining initiatives by creating awareness about the work required after process mining insight have been acquired. We provide material for organizations to understand how diverse are the actions they may need to take based on process mining insights. This understanding of the different actions that can potentially be triggered by process mining insights is a first step towards bridging the gap between process analysis and process improvement implementation. In addition, such understanding also serves as a basis to, for example, strengthen the link between process mining and business process management in future research. Moreover, we identify gaps in the literature regarding the translation of process mining insights into process improvement, unveiling further areas for future research.

Our recommendations to overcome challenges in translating process mining insights into process improvements can serve as guidance to help organizations realizing tangible benefits from their process mining initiatives. Moreover, we also highlight the need for a deeper understanding of the organizational factors that influence the success of process improvement projects in the context of process mining. Moreover, we uncovered five key specific causes of why organizations fail to execute recommended actions towards process improvement. These findings can serve both researchers and practitioners in developing strategies to address and mitigate the occurrence of these causes, thus improving the likelihood of process mining insights being translated into process improvement.

Thus, we believe that our work can serve as a basis for a view of process improvement triggered by process mining insights that include not only technological but also organizational, managerial, and human factors, therefore promoting interdisciplinary research to address the different challenges related to each of these dimensions in the future.

1.5 Publications

This thesis builds upon a number of research outcomes published in peer-reviewed venues. Each work contributes to understanding either the effort to extract event logs or the effort to translate process mining insights into process improvement.

- [119] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, X. Lu, I. Beerepoot, J. J. Koorn, and H. A. Reijers. *Towards Understanding the Role of the Human in Event Log Extraction*. In International Conference on Business Process Management Workshops (BPM Workshops), 2021.
- [122] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, and H. A. Reijers. *Supporting Event Log Extraction based on Matching*. In International Conference on Business Process Management Workshops (BPM Workshops), 2022.
- [120] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, I. Beerepoot, and H. A. Reijers. *From Process Mining Insights to Process Improvement: All Talk and No Action?* In International Conference on Cooperative Information Systems (CoopIS), 2023.
- [121] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, and H. A. Reijers. *Progressing from Process Mining Insights to Process Improvement: Challenges and Recommendations*. In International Conference on Enterprise Design, Operations, and Computing (EDOC), 2023.
- [123] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, I. Beerepoot, and H. A. Reijers [Accepted]. *From Loss of Interest to Denial: A Study on the Terminators of Process Mining Initiatives*. In International Conference on Advanced Information Systems Engineering (CAiSE), 2024.

Other publications yet not included in this thesis are enlisted next.

- [116] V. Stein Dani. *Event Log Extraction: How to Minimize the Effort of the Human-in-the-Loop?* (Extended Abstract). In International Conference on Process Mining Doctoral Consortium and Tool Demonstration Track (ICPM Doctoral Consortium), 2020.
- [117] V. Stein Dani, M. ER, J. J. Koorn, J. M. E. M. van der Werf, H. Leopold, and H. A. Reijers. *Pair Modeling: Does One Plus One Add Up?* In International Conference on Business Process Management Workshops (BPM Workshops), 2021.
- [70] J. J. Koorn, I. Beerepoot, V. Stein Dani, X. Lu, I. van de Weerd, H. Leopold, and H. A. Reijers. *Bringing Rigor to the Qualitative Evaluation*

of Process Mining Findings: An Analysis and a Proposal. In International Conference Process Mining (ICPM), 2021.

- [27] J. V. de Camargo, N. M. Bohnenberger, V. Stein Dani, J. P. M. de Oliveira, E. Sosa, G. Polančič, and L. H. Thom. *A Complementary Analysis of the Behavior of BPMN Tools Regarding Process Modeling Problems*. In Enterprise, Business-Process and Information Systems Modeling Working Conference (BPMDS), 2022.
- [118] V. Stein Dani, C. M. dal S. Freitas, and L. H. Thom. *Recommendations for visual feedback about problems within BPMN process models*. Software and Systems Modeling, Volume 21, 2022.
- [83] Y. Liu, V. Stein Dani, I. Beerepoot, and X. Lu [Accepted/In press]. *Turning Logs Into Lumber: Preprocessing Tasks in Process Mining*. In International Conference on Process Mining Workshops (ICPM Workshops), 2023.

1.6 Thesis Outline

Figure 1.2 depicts the outline of this thesis structure. In part α, we focus on event log extraction; while in part ω, we focus on progressing from process mining insights to process improvement.

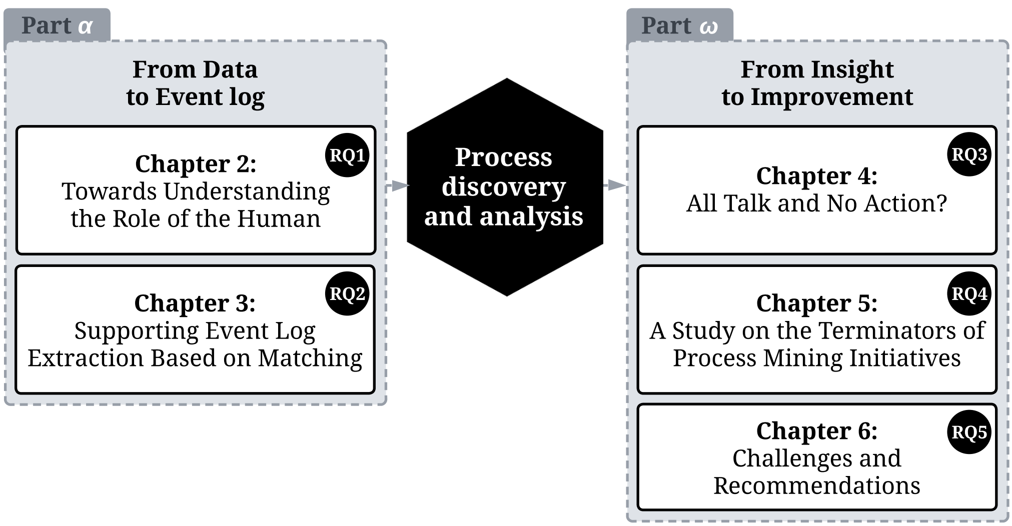


Figure 1.2: Overview of this thesis outline.

Part α: From Data to Event Log

- **Chapter 2: Towards Understanding the Role of the Human.** In this chapter, we answer RQ1. To do so, we develop an understanding of how humans are involved in event log extraction. Based on a structured literature review and qualitative data coding, we derive a taxonomy of human tasks in event log extraction. This taxonomy can serve as input for both future automation efforts, as well as for process mining methodologies.

- **Chapter 3: Supporting Event Log Extraction Based on Matching.** In this chapter, we answer RQ2. To do so, we take a novel angle at supporting event log extraction. The core idea presented in this chapter is to use an existing process model as a starting point and automatically identify to which database tables the activities of the considered process model relate to. Based on the resulting mapping, an event log can then be extracted in an automated fashion. In this chapter, we define a first approach that is able to identify such a mapping between a process model and a database.

Part ω: From Insights to Improvement

- **Chapter 4: All Talk and No Action?** In this chapter, we answer RQ3. To do so, we investigate which types of actions have been taken in existing studies and to which insights these actions are linked. Our findings show that a large variety of actions exists. Many of these actions relate to changes to the investigated process and also to the associated information systems, the process documentation, the communication between staff members, and personnel training. Understanding the diversity of the actions triggered by process mining insights is important to instigate future research on the different aspects of translating process mining insights into process improvement.

- **Chapter 5: A Study on the Terminators of Process Mining Initiatives** In this chapter, we answer RQ4. To do so, we combine a systematic literature review with semi-structured interviews of process mining experts to develop a better understanding of the extent to which recommended actions are actually performed, as well as the causes hampering the progress from recommended to performed actions. Based on our analysis, we discover specific causes why organizations do not perform recommended actions. These findings are crucial for both researchers and organizations to develop measures to anticipate and mitigate these causes.

- **Chapter 6: Challenges and Recommendations** In this last chapter, we answer RQ5. To do so, we explore the challenges involved in progressing process mining insights into process improvement. By conducting a qualitative study based on semi-structured interviews, we identify seven challenges pertaining to translating process mining insights into process improvements. Furthermore, we provide five specific recommendations for practitioners and stakeholders that should be considered before starting a new process mining initiative. By doing so, we aim to close the gap between insights and action and help organizations use process mining to realize process improvements effectively.

Part α

From Data to Event Log

Chapter 2

Towards Understanding the Role of the Human

Abstract

Process mining is widely used to visualize, analyze, and improve business processes. However, often its application is hindered by the considerable preparation effort that needs to be conducted by humans. One of the key tasks required in this context is obtaining the input artifact for process mining techniques: the event log. The data that is required for building such an event log typically needs to be collected from several databases and then transformed into a suitable format. While it has become clear to both academics and practitioners that the amount of human work is substantial, there is no deep understanding of the exact activities humans need to perform. Therefore, we use this paper to develop a precise understanding of how humans are involved in event log extraction. Based on a structured literature review and qualitative data coding, we derive a taxonomy of human tasks in event log extraction. This taxonomy can serve as input for both future automation efforts, as well as for process mining methodologies.

Published in V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, X. Lu, I. Beerepoot, J. J. Koorn, and H. A. Reijers. Towards understanding the role of the human in event log extraction. In International Conference on Business Process Management Workshops (BPM Workshops), 2022 [119]

2.1 Introduction

Many organizations use process mining techniques to visualize, analyze, and improve their business processes. Among others, process mining has been applied in auditing [61], production planning [40], and in the healthcare domain [110]. While process mining has been shown to come with many benefits [51], its application is still often hindered by the considerable preparation effort that needs to be conducted by humans [34]. One of the key challenges in this context is to obtain the input artifact for process mining techniques, the so-called *event log*.

Depending on the specific IT landscape, the data that is required for building an event log must be collected and extracted from several databases and then transformed into an appropriate format that can be processed by available process mining tools. While a variety of automated techniques have been proposed to support organizations in obtaining event logs (e.g., [20, 29, 95]), human support is still required at several stages. Unfortunately, there is currently no deep understanding of the *exact activities* humans are performing in this context. While this may seem surprising, this can be explained by the fact that available techniques for event log extraction often focus on rather isolated technical aspects and not on the full range of tasks that are required in a practical setting.

Recognizing that the human involvement in event log extraction comes with considerable time and cost, we set out to develop a precise understanding of how humans are involved in the context of event log extraction. We believe that such an understanding is vital input for both future automation efforts as well as for process mining methodologies. Therefore, the research question of this paper is "*What are the specific manual tasks that humans perform in the context of event log extraction?*". To answer our research question, we first conduct a structured literature review targeting process mining case studies. Based on qualitative data coding, we then use the identified papers to derive a taxonomy of human tasks in event log extraction.

The remainder of the paper is organized as follows. Section 2.2 discusses the background of this paper. Section 2.3 describes our research method. Section 2.4 presents our taxonomy of manual tasks in even log extraction. Section 2.5 reflects on the implications and limitations before Section 2.6

No.	Case identifier	Event	Timestamp
1	123	Receive order	2021-05-11 11:35
2	123	Collect order items	2021-05-11 12:41
3	234	Receive order	2021-05-12 09:03
4	123	Send invoice	2021-05-12 13:49
5	234	Collect order items	2021-05-12 17:12
6	123	Ship order	2021-05-13 08:54

Table 2.1: Excerpt from an event log of an ordering process.

concludes the paper.

2.2 Background

Process mining is a family of data analysis techniques that aims to discover, monitor, and improve organizational processes by analyzing data from so-called event logs [135]. Such event logs can be obtained from various IT systems and provide insights into how organizational processes are executed. Table 2.1 shows an excerpt of a simple event log of an ordering process. It shows that each entry of an event log must have at least three attributes: a case identifier, an event, and a timestamp. The event column reveals *what* happened. The timestamp column shows *when* the event occurred. The case identifier relates each event to a particular process execution, often referred to as *case*. In the example, we only observe two different cases (123 and 234).

Event logs like the one from Table 2.1 are a prerequisite for any available process mining technique. The process of obtaining event logs is called *event log extraction*. It is a complex and time-intensive process, which requires human involvement at several stages. The data required for constructing an event log often resides in a variety of sources. One of the main challenges is that many information systems are not process-centric and, therefore, do not record events or case identifiers explicitly. This means that we may need to identify and extract event data from a variety of different databases and transform them to a process-centric event log [34].

To provide automated support for this endeavor, many automated techniques have been developed. Recognizing the large variety of potential data sources and requirements in practice, available techniques differ considerably with respect to required inputs, their output, and also their limitations.

Among others, there are extraction techniques that build on ontologies [20], redo logs [29], and database objects [95]. All these techniques require human intervention or input at some stage. Unfortunately, the exact role of the human is not always clear. Since many of these techniques address rather specific problems of event log extraction, the required human involvement can often only be understood when these techniques are applied in practice.

Given that the human involvement in event log extraction is both time and cost-intensive, we use this paper to develop a precise understanding of the respective human tasks. We argue that such an understanding is a prerequisite for further automation efforts, as well as for developing process mining methodologies. Therefore, we define our research question as: *"What are the specific manual tasks that humans perform in the context of event log extraction?"*.

2.3 Research Method

To answer our research question, we followed a two-step approach. We first conducted a structured literature review. Based on the result, we then derived a taxonomy of human tasks via qualitative data coding. In the sections 2.3.1 and 2.3.2, we describe both steps in detail.

2.3.1 Literature Study

To identify which manual tasks have been performed in the context of event log extraction, we decided to focus on papers conducting case studies in process mining projects. In Scopus, we used the search string: ("process mining" AND "case stud*"). We included peer-reviewed papers that were published in journals or conferences between 2000 and 2020. Next, we filtered out duplicate papers, papers that were not in English, and papers that merely mentioned process mining. The search resulted in 191 papers. We identified papers that could contain tasks related to event log extraction by making a first read of the papers. This resulted in a set of 120 papers that were assessed more closely to extract the actual tasks related to manual efforts during log extraction.

As a next step in the study, we went through the 120 papers to identify human tasks related to event log extraction. To label a task as manual, we defined the following criteria:

1. It is explicitly mentioned that the task was performed manually.
2. There was no explicit counter statement that the task was performed (semi-) automatically.
3. We assume that conceptual models, such as data models and process models, were created manually unless the authors explicitly stated that they were created in an automated fashion.

With this in mind, the final set of papers from which we identified manual efforts is composed of 46 papers. On these papers we performed the coding as presented in the next section.

2.3.2 Coding and Taxonomy Derivation

To derive a taxonomy of manual tasks, we coded the 46 papers resulting from our literature review based on the coding of qualitative data [113]. We performed three specific steps: First, we performed hypothesis coding. This means that we devised a set of codes we expected to find in the data without actually conducting any further analysis. Second, we performed a holistic coding to identify possible categories that could emerge from and represent the data. Third, we compared both coding schemes and discussed how they relate to each other to achieve a more concise representation of the identified categories. We identified that we could directly match the holistic coding entries to our hypothesis coding, which led us to the taxonomy presented in Section 2.4.

For example, the following codes emerged from the holistic approach: i) "search into the data to select case perspective", supported by "*[...] It is therefore possible to consider the data from at least three different 'case' perspectives, i.e. an incident may be considered as a case, each patient may be considered as a case, or each response unit may be considered as a case. [...]*" [5]; and, ii) "identify, from discussion with domain expert, which case should be considered", supported by "*[...] in this case we constructed three different event logs according to the different perspectives. [...] After several discussions, it appeared that both the team flow and the document type flow represented the business process best [...]*" [143]. Both codes are related to the selection of a case, which can be summarized by our hypothesis code "Select case notion". As can be seen from these examples, we used lower-level coding in our holistic coding, and high-level coding in our hypothesis coding.

2.4 A Taxonomy for Manual Tasks in Event Log Extraction

In this section, we present the outcome of our research: a taxonomy for manual tasks in event log extraction. We first provide an overview of our taxonomy in Section 2.4.1. In the remainder of the section, we discuss the details of each of the five categories.

2.4.1 Overview

A visual representation of our taxonomy is shown in Figure 2.1. It consists of five categories, represented as four squares and one rectangle. These five categories resemble the main manual tasks we identified in event log extraction. Starting from the context and scope definition, the data source need to be assessed, and attributes need to be selected. Based on this selection, an event log can be extracted from the data source, and needs to be assessed before it can be used in process mining. Note that there is no strict flow. Some tasks might be executed repeatedly or in an arbitrary order. The numbers in brackets indicate in how many papers these task categories were mentioned.

2.4.2 Context and Scope Definition

The first step as indicated by several process mining methodologies (cf. [138]) is to define the overall context and scope (supported by 25 papers from our literature study). Naturally, this is a completely manual task that mainly consists of discussions between the process analyst and the stakeholders. The scope can vary between very exploratory to very specific. Once there is an agreement on the scope, this defines the subsequent data extraction steps. However, as pointed out in [143], the scope definition often needs to be adjusted or redefined:

> *"[...] it proved very difficult to mark out the boundaries of the process under investigation in the larger DMS [Document Management System].*

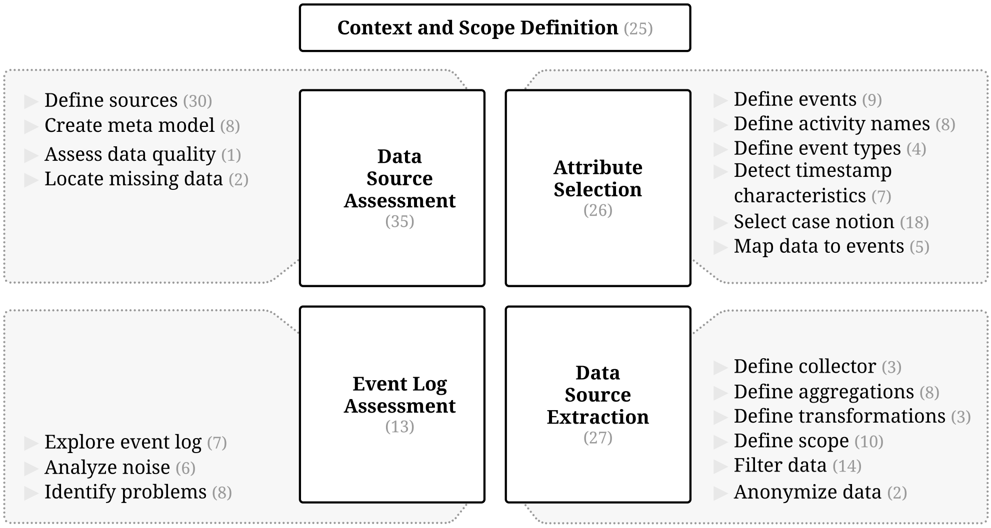


Figure 2.1: Taxonomy of manual tasks in event log extraction.

> *Therefore, the data scope had to be fine-tuned multiple times, going through the scope adjustment loop after each inspection, until a satisfactory data set was obtained.*" [143]

This excerpt highlights that the definition of context and scope is not a one-off task, but may require repeated manual interventions.

2.4.3 Data Source Assessment

Event logs are extracted from data sources. As a first step, these sources need to be assessed on their usability and quality. In our literature study, we identified four task categories: reverse engineer a meta model, define sources, audit data, and locate missing data. These categories are supported by 35 papers overall.

Define sources

Typically, there is a variety of potential data sources available. This means that a human has to understand, analyze, and decide which data sources contain the required data and, therefore, need to be considered for data extraction. That this is a complex manual task is, for example, highlighted in [142]:

> *"The selection of information on business events is an important challenge*

> *in event log extraction: while large sets of information may reduce the performance of process mining tools, too little information may affect the quality of the analysis and resulting conclusions.*" [142]

However, currently, there is no clear perspective on how this manual task can be supported or automated.

Create meta model

When data is extracted from multiple sources, the process analyst needs to decide on how to properly merge the data. This task can be supported by acquiring (or generating) meta models for each data source. This provides a better understanding of the different data sources [5], and allows the process analyst to already develop an understanding of how the data relates to specific process activities. In our literature study, we identified that authors leverage different meta models for this task, such as entity models [39] and hierarchy models [69].

Assess data quality

Prior to data extraction, there should be an assessment of the quality of the available data sources. It may be that data sources turn out to be unsuitable for process mining. An example has been reported in [64]:

> *"Clinical inspection of the data quality of the EHR [Electronic Health Record] revealed data that was considered too unreliable to use for process mining. Issues included unrecorded events and observations, recording on letters and paper records rather than the EHR, mis-diagnosis and inappropriate referrals.*" [64]

This example highlights that some manual effort is not only required for the assessment, but also for fixing the problem. In the case above, the authors decided to manually review large amounts of data, which they then manually transformed into a suitable format.

Locate missing data

Suppose the initially available sources do not contain all the necessary data for the process mining project. In that case, there is a need to locate the missing data in other data sources or extract it by other means. In practice, it is often necessary to obtain data managed by departments in the organization that otherwise would not be considered to take part in the process mining initiative. However, suppose the missing data is not available in *any* of the information systems of the organization. In that case, this data may be obtained by other means. In [64], for example, the authors filled the gap of missing data through interviews with domain experts. In any case, the detection and localization of missing data is so far still a manual and time-consuming effort.

2.4.4 Attribute Selection

Once the data sources have been assessed and selected, the next step is to select the attributes for log extraction. In this category, we identified six task categories, which are supported by 26 papers: define events, define activity names, define event types, detect timestamp characteristics, select the case notion, and map activities to events.

Define events

One of the key tasks here is to decide which attributes will together represent the events. Attributes can be more generic towards process mining, such as resources, actors, and activities. In some case studies, highly context-specific attributes were selected. An example of such a context-specific attribute is provided in [142]:

> "*Some logistics elements (e.g., the cargo type) can be selected as the instance attributes such that the logisticians could look at the processes from a wider perspective [...] The cargo type of the instance is useful information as it has a big impact on the cargo handling processes*" [142]

The majority of the studies report on their defined events or the relevance of a proper selection of events [47, 80, 98] without getting much into details on the decisions related to the event definition, nor how it was performed.

Define activity names

Defining activity names involves different actions, such as data correlation and abstraction [34]. In [52], the authors use names already available in the data source. While in [142], to perform this task, the authors use available documents. Independently of how this task is performed, it usually still requires manual efforts and more thorough discussions within the case studies in the literature.

Define event types

If event types are available (e.g., start, complete, etc.), it might be needed to define which will be necessary for the analysis. In some cases [105], there are so many event types available that there is the need to trim down the data by choosing only event data from a subset of event types. Otherwise, the data would generate spaghetti-like process models.

Detect timestamp characteristics

Timestamps are available in data sources in many different formats, sizes, and varieties. The detection of timestamp characteristics before the data extraction can anticipate different problems, such as in [87]:

> *"[...] we do not have any information about the actual timestamps of the start and completion of the service delivered. Consequently, the ordering of events which happen on the same day do not necessarily conform with the order in which events of that day were executed."* [87]

However, few studies reporting on the event log extraction report on the detection of timestamp characteristics in early stages [4, 32]. From our findings, we identified that the studies discuss timestamp characteristics only later on in the process of event log extraction [52].

Select case notion

In many settings, there are several options concerning the case notion selection. A typical example could be found in the healthcare domain:

> *"It is [...] possible to consider the data from at least three different 'case' perspectives, i.e. an incident may be considered as a case, each patient may be considered as a case, or each response unit may be considered as a case."* [4]

As also pointed out by several other authors, this decision is an important one to take and may require several analyses to understand the implications of a potential case identifier [47, 143].

Map data to events

In this task, process analysts need to identify which tables contain data that relate to events they wish to be part of the event log. What is more, they need to define which mappings provide the intended representation of the events. Often this mapping is guided by activities from a reference model resulting from the previous task category. It can also be directly performed by the process analysts [6].

2.4.5 Data Source Extraction

Once the attributes have been selected, the actual event logs can be extracted. In this category, we identified six task categories: define collector, define aggregations, define transformations, define scope, filter data, and anonymize data. These categories are supported by 27 papers from our literature study altogether.

Define collector

In some cases [4, 64], the process analyst requests and receives the desired data from a data expert, who represents an involved stakeholder. In other cases, the process analyst has access to the data either via direct access to the data sources [13] or via a dump from the needed data sources [141]. The actual role of the person who collects the data is often left implicit (as e.g. in [25, 88]).

Define aggregations

When data is too fine-grained, there is often the need for the process analyst alongside the domain expert to define which will be the aggregations that need to be manually performed [44, 125]. The importance of defining aggregations is reported by [111], which states that the aggregation definition leads the process analyst towards a high-level event log, which reduces the complexity of the resulting process model.

Define transformations

Different studies report on data transformation supported by domain knowledge [62, 99]. In [62], the authors report they transformed data scattered across a set of tables in a database into a unified format. In [99], the authors transformed some data elements using proxy timestamps whenever there were no previously recorded timestamps.

Transformations also refer to the decisions and manipulations performed to transform event data from non process-centric information systems into a process-centric event log. Independently of the transformations performed, it is still necessary to define and report on the data transformations to be consistent and keep documentation of what was performed over the data and why. In this way, it becomes repeatable.

Define scope

Depending on the available data, it is necessary to define the scope further. For instance, in [47], there is too much data about a process related to many different company branches. In such a case, the authors decided to trim the data by narrowing it down to one branch to "*produce a more focused analysis*" [47].

Filter data

After the different required definitions to perform the extraction, invariably, data filtering is performed. In many cases, the process analyst can make use of tools to support this task. However, domain knowledge and manual efforts are often driving the data filtering task [47, 64, 143]. From practice, we know that data filtering is an iterative process, which is often aligned with process discovery and conformance checking.

Anonymize data

If the process under analysis has data linkable to a specific person, this data should be anonymized (or at least pseudonymized) to preserve privacy [88, 134, 142]. The detection of the particular attributes that may incur in links to specific persons relies on domain knowledge. Such detection is often performed manually.

2.4.6 Event Log Assessment

As a final category in the extraction process, the extracted event log is assessed on its quality. In this category, we identified three task categories supported by 13 papers: explore event log, analyze noise, and identify problems.

Explore event log

The process analyst needs to explore the extracted event log to reflect back on the previous definitions and to attempt identifying problems that might occur because of remnants of noise, incomplete or imprecise data, or even inadequate definitions. Some studies are explicit about this task, such as in [47]:

> *"The next phase of event log exploration is intended to adjust the data set according to the time and scope of the work, to identify sanitization rules, and, to identify the different analysis dimensions."* [47]

During this task, the process analyst often leverages process discovery and conformance checking techniques to perform the exploration. However, data and scope adjustments are usually ad-hoc, and performed manually [143].

Analyze noise

If the definitions, mapping, and filtering previously mentioned are not performed thoroughly, different types of noise in data may still be detected after data is extracted, such as incomplete cases [47], and outliers [13]. And even with thorough definitions, one can still find, for example, infrequent process variants [125]. Although there are several tools to support data filtering, domain knowledge drives the noise analysis and is usually performed by a process analyst alongside a domain expert.

Identify problems

Once the event log is extracted, it can be assessed on potential problems. Examples of problems identified are: incomplete recordings of events [125], and incomplete cases [47]. This identification task is also driven by domain knowledge. It involves iterations between the process analyst and the domain expert, who will be redefining the scope or the events themselves, and performing data filtering.

2.5 Discussion

In this section, we discuss the implications and the limitations of our taxonomy for manual tasks in event log extraction.

As for the *implications* of our taxonomy, we argue that it provides an important starting point for reducing the extent of human involvement in event log extraction. We can identify two main use cases in this regard. The first use case concerns *automation*. Without exactly understanding where and how humans are involved, increasing the level of automation is hardly feasible. Our taxonomy, therefore, provides important input on the tasks that can be potentially automated. The second use case concerns *guidance*. Given the nature of some tasks (e.g. defining the scope or defining sources), it is clear that not all human tasks can be automated. However, what can still reduce the time investment and the overall effort is an increased level of guidance for manual tasks. Such guidance may include checklists or other artifacts that provide orientation and help to structure the task execution.

Naturally, our taxonomy is also subject to a number of *limitations*. The first limitation concerns the scope of our literature study. Since we exclusively targeted process mining case studies, we cannot claim to provide a complete picture. We, however, made this choice deliberately since we wanted to learn about the use of process mining in real-world environments. The second limitation concerns the use of a literature study in general. While studying the selected papers, we realized in several places that certain details are missing or not explicitly discussed. Hence, there might exist additional manual tasks that were simply not reported upon or discussed in the analyzed papers. Nonetheless, due to the extent of our literature study, we are confident that our taxonomy provides a rather comprehensive overview.

2.6 Conclusion

The extraction of event logs comes with substantial human effort. In this paper, we set out to develop a precise understanding of which manual tasks humans perform in the context of event log extraction. We conducted a structured literature review and applied qualitative data coding to systematically derive a taxonomy of manual tasks. Our taxonomy highlights that human work is required in various phases of the event log extraction process ranging from scope and context definition to data source assessment. As such, our taxonomy does not only provide a comprehensive overview but can also serve as input for future automation efforts and for methodological process mining support. In future research, we aim to follow up on the insights provided in this paper and develop techniques that reduce, simplify, or support human work in event log extraction.

Chapter

3

Supporting Event Log Extraction Based on Matching

Abstract

Process mining allows organizations to obtain relevant insights into the execution of their processes. However, the starting point of any process mining analysis is an event log, which is typically not readily available in practice. The extraction of event logs from the relevant databases is a manual and highly time-consuming task, and often a hurdle for the application of process mining altogether. Available support for event log extraction comes with different assumptions and requirements and only provides limited automated support. In this paper, we therefore take a novel angle at supporting event log extraction. The core idea of our paper is to use an existing process model as a starting point and automatically identify to which database tables the activities of the considered process model relate to. Based on the resulting mapping, an event log can then be extracted in an automated fashion. We use this paper to define a first approach that is able to identify such a mapping between a process model and a database. We evaluate our approach using three real-world databases and five process models from the purchase-to-pay domain. The results of our evaluation show that our approach has the potential to successfully support event log extraction based on matching.

Published in V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, and H. A. Reijers. Supporting event log extraction based on matching. In International Conference on Business Process Management Workshops (BPM Workshops), 2023 [122]

3.1 Introduction

Process mining is used in many different organizations for tasks such as analyzing, improving, and auditing business processes [36, 61, 104]. However, the application of process mining requires an event log [135], which is often not readily available in practice [34]. One of the main reasons is that the information systems supporting the execution of many business processes do not produce event logs that can be used for process mining. As a result, event logs need to be extracted manually by exploring the underlying databases of these information systems. In essence, every activity executed in the context of the business process must be manually related to specific tables in the database. This mapping is then used to extract the event log. This effort for event log extraction is very time-consuming and requires considerable manual work [119]. It, thus, creates a substantial hurdle for the application of process mining in practice [138].

Recognizing this, many researchers have developed techniques to support the extraction of event logs. However, they usually require creating an intermediate data model [30] or using instance data [82]. Furthermore, they do not automatically identify the mapping between the tables of a database and the activities of a considered process because they do not focus on extracting event logs that relate to an already known process flow.

In this paper, we propose a novel approach for supporting event log extraction that takes an existing process model as a starting point. The core idea is to automatically identify to which database tables the activities of a given process model relate to and, based on the resulting mapping, provide an effective alternative for event log extraction. In prior work, the problem of mapping entities from two different representations has been addressed in various contexts. Among others, researchers have proposed techniques for finding mappings between database schemas [84, 94], between ontologies [63, 72], or between process models [132, 144]. Such techniques for automatically deriving mappings between two different representations are commonly referred to as *matchers* [144]. However, to the best of our knowledge, there is no approach available that focuses on identifying a mapping between a database and a process model [119]. To accomplish this, we build on a two layer matching architecture and different notions of similarity.

The remainder of the paper is structured as follows. In Section 3.2, we illustrate the problem of and the challenges related to creating a mapping between database tables and process model activities. In Section 3.3, we describe our proposed approach to support event log extraction based on matching. Section 3.4 evaluates an implemented proof-of-concept. Finally, in Section 3.5, we discuss related work and in Section 3.6, we conclude this paper.

3.2 Problem Illustration and Challenges

In this paper, we approach the problem of event log extraction from a matching perspective. More specifically, we aim to develop an approach that automatically identifies a mapping between the tables of a database and the activities of a given process model. To illustrate the problem and the associated challenges, consider the example shown in Figure 3.1. It shows a simplified purchase-to-pay process model (extracted from [36]) and a corresponding exemplary database. The goal of our approach is to identify for each activity from a given model to which database table it relates (if any). Formally, such a mapping is a relation over the activities and tables, such that (a, t) maps activity a to table t. In other words, table t contains data of an event for activity a. A *potential mapping* is a candidate mapping that needs to be verified for correctness. Figure 3.1 depicts several potential mappings. The relations with a checkmark are correct mappings, whereas the mappings marked with a cross are incorrect. Automatically identifying the correct mappings comes with four main challenges:

1. *Large search space*: Given that databases often contain hundreds of tables, the search space for the mapping is typically very big. To illustrate this, consider the example from Figure 3.1. The combination of 6 activities and 26 tables already results in over 300 million possible mappings. A useful matching technique, therefore, must be able to effectively reduce the search space and precisely recognize which activity-table pairs represent correct mappings.

2. *Granularity differences*: Processes and databases dramatically differ in their level of granularity. While a process model typically only has a handful of activities [43], a database often has hundreds of tables. This

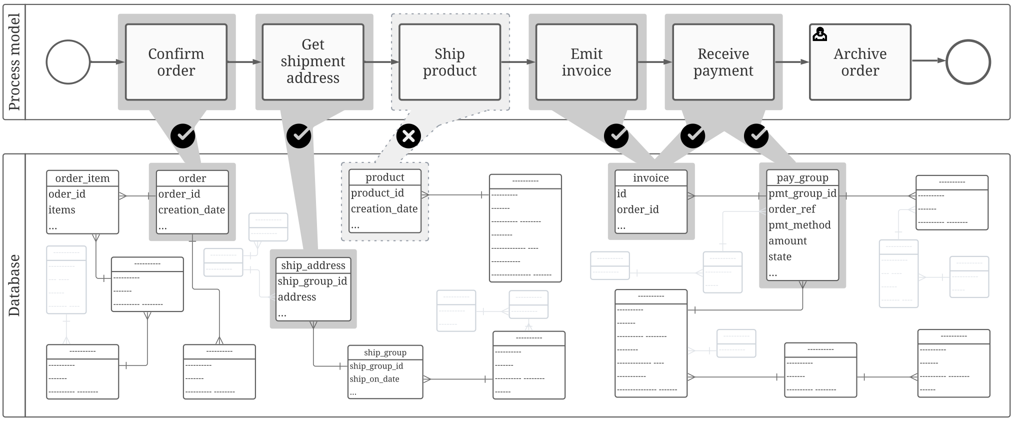


Figure 3.1: A process model, a database, and the mappings between them.

causes two related problems. First, this means that a single activity may have multiple corresponding tables. For example, in Figure 3.1, the activity "*Receive payment*" produces a payment entry for the database table "*pay_group*" while also producing an update of an entry in the table "*invoice*". Second, this means that a single table may have multiple corresponding activities. For example, the table "*invoice*" stores data about a newly created invoice produced by the activity "*Emit invoice*". The same table also reflects a payment status updated via the execution of the activity "*Receive payment*".

3. *Scope differences*: The scope of the process model and the database rarely overlap to a full extent. As a result, the mapping between process model and database is partial. This means that some activities do not have a correspondence to any table and, the other way around, many tables do not have a correspondence to any activity. For example, in Figure 3.1, the activity "*Archive order*" may be related to a manual status update executed on an external system managed by another department of the organization and, therefore, does not relate to any of the tables of the considered database.

4. *Ambiguous semantics*: Both process models and database tables typically have very short labels. As a result, it is often hard to identify which words from the considered labels carry the important semantics. To illustrate this, consider the activity "*Ship product*" from Figure 3.1. We can see

that this activity contains the action "*ship*" and the object "*product*". In Figure 3.1 it is, however, incorrectly mapped to the table "*product*" instead of "*ship_group*". The problem is that it is hard to evaluate which term should be used in this context to decide about the mapping since both "*ship*" and "*product*" are used in the database tables.

In this work, we make a first attempt to address these challenges. We propose an approach that identifies a set of potential mappings between process model activities and database tables. While this does not provide the user with a final set of correct mappings, the user is provided with a small set of potential mappings. From those, the user can simply select the correct mappings and, hence, no longer needs to look at all possible mappings and identify each mapping manually. We realize that this only represents a first step. We are, nonetheless, convinced that this already dramatically reduces the burden of the process analyst and saves a considerable amount of manual work. In the next section, we introduce our approach on a conceptual level.

3.3 Mapping Database Tables to Process Activities

In this section, we describe our matching approach to automatically map database tables to process model activities. We first present an overview of the architecture of our matching approach in Section 3.3.1. Then, in Section 3.3.2 and Section 3.3.3, we discuss the main components of our matcher in detail.

3.3.1 Overview

Figure 3.2 shows the architecture of our proposed approach. The first module is responsible for *preprocessing* and feeding input data into the matcher. Among others, the preprocessing component parses the input, removes irrelevant tokens (such as punctuation), and turns all strings into lower case. The input data includes a database and a process model. At this point, we expect that both have already been transformed into a textual format and are provided as CSV files. These files contain the table attributes from the database (e.g., tables names, descriptions, and columns with their names and descriptions), and the activity labels from the process model.

Inspired by [46], the *matcher* module consists of two main components: a first- and a second-line matcher (1LM and 2LM), where the 2LM builds on the output of the 1LM. The matcher automatically generates a set of potential mappings. To generate these potential mappings, we leverage natural language processing (NLP) techniques and the available input information. The main intuition behind relying on NLP techniques is that tables and activities with similar names are more likely to be conceptually similar and, therefore, related. In the following sections, we explain the details of the components from the matcher module.

3.3.2 First-Line Matcher (1LM)

Our approach starts with analyzing the set of activity labels A of the process model and tables T of the database using different similarity metrics. For each table, we consider all database table attributes, denoted by R. Then, for each activity a and database table attribute t_r, several similarity measures are calculated. This results in a set of similarity matrices M_s, for each similarity measure s.

Table 3.1 shows a cohort of the similarity matrix $M_{s(A \times R)}$ for the normalized Levenshtein-based similarity measure on a process model with two activities, “*Create order*” and “*Create invoice*”, and a database consisting of two tables, “*Order*” and “*Invoice*”. In this example, the table “*Order*” has two columns: “*id*”, and “*creation_date*” and, the table “*Invoice*” has three columns: ‘*id*”, “*id_order*”, and “*date*”.

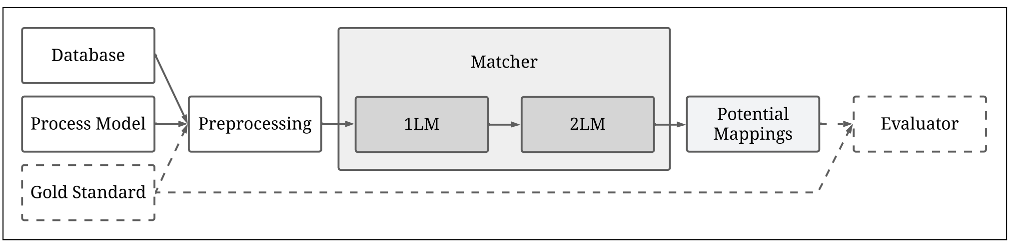


Figure 3.2: Architectural overview of our approach.

3.3.3 Second-Line Matcher (2LM)

The 2LM derives the set of potential mappings between tables and activities by using as input the similarity matrix M_s generated by the 1LM. Our approach maps exactly one database table t to one activity label a, and the inner workings of the 2LM adheres to the following rationale: First, considering all available similarity scores in M_s (cf., Table 3.1), the 2LM determines a similarity score to represent a table with respect to each activity. This is performed for each tuple (a, t). Second, for each activity, it selects one table as a potential mapping considering the similarity score assigned to the table. Many different mechanisms can be implemented to derive the table's similarity score from its attributes' similarity scores.

We developed a baseline 2LM inspired by [132], which selects the *Highest raw 1LM-based Scoring Table* as a potential mapping for an activity label. Based on the output of this 2LM for each 1LM similarity matrix, we performed an inductive content analysis with open coding [113]. Recurrent observations from the coding served as a basis for the definition of two new 2LM implementations: one based on *Word Frequency ($2LM_2$)*, and another based on the *Surface Measure of Overall Table Scores ($2LM_3$)*. Next, we further explain each implemented 2LM.

Table 3.1: Similarity matrix generated by the 1LM for the normalized Levenshtein-based similarity algorithm. The closer the similarity score is to 1, the higher the similarity between the two compared objects.

Database tables attributes t_k	**Process model activities** a	
	Create order $f_s(a, t_k)$	**Create invoice** $f_s(a, t_k)$
Order	0.590	0.210
id	0.140	0.120
creation_date	0.480	0.440
Invoice	0.210	0.670
id	0.140	0.120
id_order	0.500	0.180
date	0.380	0.330

Highest raw 1LM-based Scoring Table ($2LM_1$)

Each row in a similarity matrix M_s produced by the 1LM represents the similarity scores of an activity and all attributes t_r of all tables $t \in T$. $2LM_1$ selects for each activity and table combination the attribute with the highest similarity as *table score*. Then, for each activity, the table with the highest table score is selected as potential mapping.

Word Frequency ($2LM_2$)

This technique multiplies the table attributes similarity score by the number of activity label word repetition within the table attribute. This is done before the *table score* definition and, if there is no word repetition, the similarity score is kept as is. Hence, this matcher derives each of its potential mappings similarly to $2LM_1$.

Surface Measure of Overall Table Scores ($2LM_3$)

This technique is inspired by [58], and leverages all similarity scores of a table to build a radar chart, where each similarity score is an axis of the chart. The *table score* $S(a,t)$ is then determined by calculating its surface area, as shown in Equation 3.1, where R denotes the set of table attributes.

$$S(a,t) = \sin\left(\frac{\pi}{|R|}\right) \sum_{x \in R} \sum_{y \in R} \left(M_s(a,t_x) \cdot M_s(a,t_y)\right) \tag{3.1}$$

3.4 Evaluation

In this section, we present a quantitative evaluation of our approach. In Section 3.4.1 and Section 3.4.2, we describe the data and our setup. In Section 3.4.3, we report on the results and provide a discussion in Section 3.4.4.

3.4.1 Data

The evaluation builds on three inputs: 1) a set of databases, 2) a set of process models, and 3) a gold standard.
Databases. For the evaluation, we used three databases: 1) Odoo (former Open ERP), 2) Magento Commerce, and 3) Oracle ATG Webcommerce. The selected

databases cover two scenarios we want to evaluate: databases with and without textual descriptions of the tables and columns. Oracle is accompanied by a textual description, whereas Odoo and Magento are not. Additionally, these databases were selected considering two other factors: 1) they store purchase-to-pay data; and, 2) they are widely used. Table 3.2 summarizes the overall characteristics of the selected databases.

Table 3.2: Characteristics of the databases used in the evaluation of our approach.

Characteristic	Odoo	Magento	Oracle
Database			
Nº of tables	571	358	239
Nº of columns	6294	3561	1199
Nº of words	57297	42189	37051
Table			
Avg Nº of words per table name	2.794	3.502	2.838
Avg Nº of words per table description	2.356	3.815	14.197
Avg Nº of words per column name	1.978	2.216	1.952
Avg Nº of words per column description	1.974	2.311	12.009

Process models. We used five process models of a purchase-to-pay process of different sizes. The set of process models contains one small process model extracted from [36], and four medium-sized process models extracted from the BPM Academic Initiative (BPMAI) repository [145]. The BPMAI models were selected based on the following criteria: 1) it is modelled in English, 2) it contains at least 10 activities, and 3) it relates to a purchase-to-pay process. To make sure the latter is the case, we selected process models containing the business objects *"order"*, *"invoice"*, and *"shipment"*. Table 3.3 summarizes the overall characteristics of the selected process models.

Table 3.3: Characteristics of the process models used in the evaluation of our approach.

Characteristic	PM_1	PM_2	PM_3	PM_4	PM_5
Process model					
Nº of activities	6	10	11	12	14
Nº of words	13	34	33	34	50
Activity label					
Min Nº of words	2	1	2	1	2
Max Nº of words	3	6	5	5	6
Avg Nº of words	2.166	3.400	3.000	2.833	3.571

Gold standard. The gold standard G contains the true mappings between the database tables t and the process model activities a. It is a set of relations (a, t). To evaluate the quality of the output of our approach (i.e., the potential mappings), we compare it to G as we further explain in the next section. We manually compiled G based on prior experience and insights into which tables hold the information related to the considered activities. For activities we did not know the corresponding table, we consulted the documentation of the database. We fine-tuned G based on discussions until consensus.

3.4.2 Setup

For each combination of database and process model, we generated five similarity matrices $M_{s(A \times R)}$ via 1LM, one for each similarity algorithm $s \in S$, comprising different string-similarity scoring techniques, such as: edit-based (via Levenshtein, and a normalized Levenshtein-based algorithm), Jaccard, n-gram, and Cosine similarity. Then, we implemented the 2LMs as discussed in Section 3.3.3, and to assess the performance of our approach we use precision, recall, and F1-score. This is in line with evaluations from other matching papers from the BPM domain (see e.g. [132]). To calculate these metrics, we compare the output from our approach with the mappings from the gold standard G.

Given a combination of a process model containing the activities A and a database containing the tables T, we compare the set of mappings between A and T from the gold standard G with the set of potential mappings P automatically produced by our approach. Based on this comparison, we can identify: 1) the correct mappings (i.e., the true positives TP) via $G \cap P$, 2) the incorrect mappings (i.e., the false positives FP) via $P \setminus TP$, and, 3) the missing mappings (i.e., the false negatives FN) via $G \setminus TP$. Thus, we can calculate precision via $\frac{TP}{TP+FP}$ and recall via $\frac{TP}{TP+FN}$. The F1-score is the harmonic mean between precision and recall.

3.4.3 Results

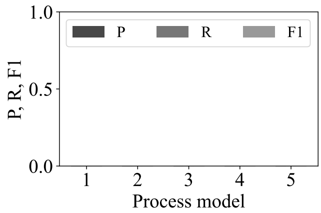
Table 3.4 summarizes the performance results of our approach in terms of precision, recall, and F1-score. For each database, the fourth column of this table presents the number of mappings in the gold standard G. This allows us to compare the number of mappings from G to the amount of correct mappings (TP) generated by each of the 2LMs.

Table 3.4: Evaluation summary with Precision, Recall, F1-scores, total true positives (TP), and false positives (FP) for the three different 2LM implementations. The baseline $2LM_1$ results are zero for Odoo and Magento because more than one table had the same highest similarity score for each activity and the baseline selects the first table with the highest similarity score as a potential mapping.

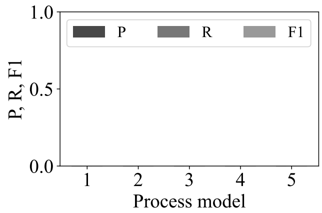
DB	**PM**	$\lvert A\rvert$	$\lvert G\rvert$	**2LM$_1$**					**2LM$_2$**					**2LM$_3$**				
				P	R	F1	TP	FP	P	R	F1	TP	FP	P	R	F1	TP	FP
Odoo	1	6	7	0.000	0.000	0.000	0	6	0.000	0.000	0.000	0	6	0.167	0.143	0.154	1	5
	2	10	9	0.000	0.000	0.000	0	10	0.100	0.111	0.105	1	9	0.100	0.111	0.105	1	9
	3	11	12	0.000	0.000	0.000	0	11	0.091	0.083	0.087	1	10	0.091	0.083	0.087	1	10
	4	12	6	0.000	0.000	0.000	0	12	0.000	0.000	0.000	0	12	0.000	0.000	0.000	0	12
	5	14	8	0.000	0.000	0.000	0	14	0.000	0.000	0.000	0	14	0.000	0.000	0.000	0	14
Magento	1	6	6	0.000	0.000	0.000	0	6	0.167	0.167	0.167	1	5	0.167	0.167	0.167	1	5
	2	10	5	0.000	0.000	0.000	0	10	0.100	0.167	0.125	1	9	0.000	0.000	0.000	0	10
	3	11	6	0.000	0.000	0.000	0	11	0.182	0.333	0.235	2	9	0.000	0.000	0.000	0	11
	4	12	4	0.000	0.000	0.000	0	12	0.000	0.000	0.000	0	12	0.000	0.000	0.000	0	12
	5	14	7	0.000	0.000	0.000	0	14	0.071	0.143	0.095	1	13	0.071	0.143	0.095	1	13
Oracle	1	6	8	0.000	0.000	0.000	0	6	0.167	0.125	0.143	1	5	0.834	0.625	0.714	5	1
	2	10	8	0.100	0.125	0.112	1	9	0.100	0.125	0.112	1	9	0.300	0.375	0.334	3	7
	3	11	11	0.091	0.091	0.091	1	10	0.273	0.273	0.273	3	8	0.454	0.454	0.454	5	6
	4	12	6	0.167	0.334	0.223	2	10	0.167	0.334	0.223	2	10	0.250	0.500	0.334	3	9
	5	14	8	0.071	0.125	0.091	1	13	0.143	0.250	0.182	2	12	0.357	0.625	0.454	5	9

On average the implemented $2LM_3$ finds 39% of the correct mappings for the databases with table and column descriptions. The implementations $2LM_2$ and $2LM_3$, perform similarly for a scenario where the database does not have useful textual descriptions, as shown in Figures 3.3d and 3.3g, for example. The $2LM_3$ implementation performs better than $2LM_2$ for a scenario where the database has textual descriptions, as shown in Figures 3.3f and 3.3i. In both scenarios, the $2LM_3$ performs well when the process model does not have too many similar activity labels, which is the case for the results related to PM_1. All the results presented in this work are based on a 1LM using cosine similarity, which is the similarity algorithm that performed best. Figures 3.3a to 3.3i depict, respectively, the output of our approach for the three different implemented 2LMs presented in Section 3.4.2. The first column of Figure 3.3 presents the output for Odoo, the second for Magento, and the third for Oracle.

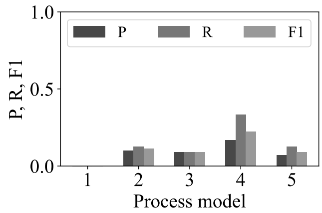
In summary, we can state that all 2LM implementations performed better on the scenario where textual descriptions were available for the tables and columns. Moreover, the $2LM_2$ and the $2LM_3$ improve consistently when compared to the baseline implementation on the Oracle database, which



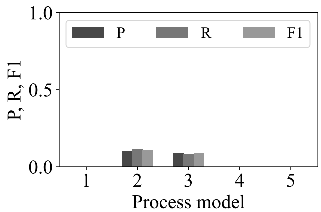
(a) Odoo $2LM_1$



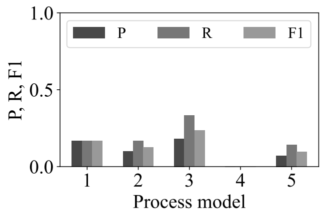
(b) Magento $2LM_1$



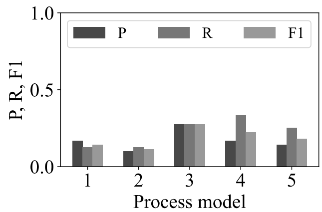
(c) Oracle $2LM_1$



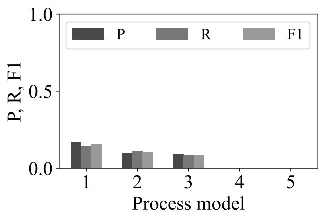
(d) Odoo $2LM_2$



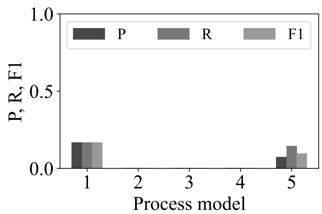
(e) Magento $2LM_2$



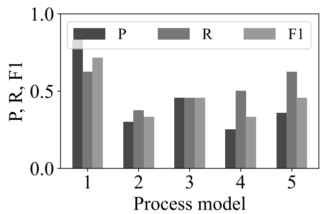
(f) Oracle $2LM_2$



(g) Odoo $2LM_3$



(h) Magento $2LM_3$



(i) Oracle $2LM_3$

Figure 3.3: Evaluation output for three databases and five process models used in this evaluation. The first row of figures shows the output for the baseline $2LM_1$, while the second and the third rows show the output for the other two implemented 2LMs. Figures 3.3a, 3.3d, and 3.3g refer to Odoo; Figures 3.3b, 3.3e, and 3.3h refer to Magento; and, Figures 3.3c, 3.3f, and 3.3i refer Oracle. On each figure, the vertical axis represents the value of precision, recall, and F1-score, for each of the five process models, shown in the horizontal axis.

is the database with textual descriptions for both tables and columns. For the databases without textual descriptions, the results deteriorate in general, showing the importance of additional textual information about the objects being mapped. The reason for this results deterioration is that multiple tables end up receiving the same similarity score, driven by similarly named columns throughout different tables.

3.4.4 Discussion

To generate the $2LM_2$ and the $2LM_3$, we derived improvement opportunities based on recurrent observations acquired via an inductive content analysis with open coding [113] performed over all outputs from the $2LM_1$. By doing so, we avoided optimizing a new 2LM to any particular scenario.

The performed content analysis supported the identification of commonalities and differences among all potential mappings (correctly and incorrectly identified mappings) versus the ones that should have been identified, but were not. We made the following key observations: First, the missing mappings (i.e., *FN*) often had repetition of words that were similar to the ones within the activity label, while it was not the case for the incorrect mappings. Second, the incorrect mappings often had multiple table elements with mild similarity scores, while the wrongly selected potential mapping had usually only one slightly higher score, which then misled the baseline mapping derivation. Therefore, the matcher should consider the table attributes scores altogether.

With the current work, we provide a first step towards supporting event log extraction based on a given process model. Our approach is able to identify a set of potential mappings, which then can be processed by a process analyst. While our technique can be further improved, we also provide some insights into how this can be accomplished (cf. $2LM_3$).

3.5 Related Work

While this paper is the first work on database to process model matching, there are three major research areas that are concerned with matching: schema matching, ontology matching, and process model matching.

Approaches for *schema matching* aim to identify matches between the elements of two different database schemata. The purpose of schema matching techniques include data integration, schema evolution, and maintenance [84, 94]. The matching strategies pursued by these techniques are similar to the ones presented in this paper. For example, in [84], the authors determine the similarity between two database schema elements using attributes, such as names and data types, and combine it with structural similarity. In [94], the authors leverage the results of a variety of basic matchers to determine whether two schema elements match.

Approaches for *ontology matching* are concerned with matching the elements of two ontologies [63, 72]. One of the key use cases for ontology matching is ontology merging, i.e., the combination of two ontologies. The matching strategies are again similar to one presented here. For example, in [63], the authors leverage lexical and structural characteristics of the considered ontologies to determine matching elements.

Approaches for *process model matching* aim to identify correspondences between the activities of two process models [81, 91, 144]. The main use case of process model matching is to detect differences and commonalities between two processes. Available approaches for process model matching exploit textual, structural, and behavioral features of the models. Early work mainly built on simple textual similarity features, such as the Levenshtein distance, and mainly focused on structural features [144]. Later, also semantic similarity measures and behavior were used to identify corresponding activities [81].

This brief review illustrates that existing matching approaches are closely related to our work. There is, however, a key difference: The works above focus on matching entities of the same type. While this does not guarantee that the to-be-matched entities are similar, they are at least comparable. In the setting addressed in this paper, we need to deal with the fact that the entities are very different in nature. A process model, for example, does not come with instance data and a database does not have a clear notion of control-flow or activities. Hence, while we partially build on matching strategies explored in previous work, the conceptual setting of our work differs considerably.

3.6 Conclusion

In this paper, we presented a new approach to support event log extraction based on matching. The main idea of our approach is to automatically identify the mappings between a database and a process model. Against the background of the challenges associated with this task, we focused on the automated identification of potential mappings in this paper. While this requires process analysts to select the correct mappings, it still saves them from a considerable amount of manual work. To evaluate our approach, we tested it using three different databases and five different process models related to a purchase-to-pay process. We found that textual information is highly important to improve the performance of our approach. At the same time, we also found that more sophisticated mechanisms are required to further improve our approach.

As for future work, we see several directions. First, we plan extend the idea from the syntactic level to a level that incorporates semantic relations as well [21, 22] between the activities and tables by, for example, leveraging bidirectional encoder representations from transformers. Second, we aim to take order relations between the database instance data and the process model activities into account. In this way, we can, for instance, exclude candidate matches if the order relations from the process model contradict the timestamps from the associated database tables. Third, we intend to incorporate feedback from humans. By letting the user select which potential mappings are correct, we can leverage a feedback loop to further improve the potential mappings generated by our approach.

Part ω

From Insights to Improvement

Chapter

4 All Talk and No Action?

Abstract

Organizations from various domains use process mining to better understand, analyze, and improve their business processes. While the overall value of process mining has been shown in several contexts, little is known about the specific actions that are taken to move from process mining insights to process improvement. In this work, we address this research gap by conducting a systematic literature review. Specifically, we investigate which types of actions have been taken in existing studies and to which insights these actions are linked. Our findings show that there exists a large variety of actions. Many of these actions do not only relate to changes to the investigated process but also to the associated information systems, the process documentation, the communication between staff members, and personnel training. Understanding the diversity of the actions triggered by process mining insights is important to instigate future research on the different aspects of translating process mining insights into process improvement. The insights-to-action realm presented in this work can inform and inspire new process mining initiatives and prepare for the effort required after acquiring process mining insights.

Published in V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, I. Beerepoot, and H. A. Reijers. From process mining insights to process improvement: All talk and no action? In International Conference on Cooperative Information Systems (CoopIS), 2023 [120]. Recipient of Best Paper Runner-up Award.

4.1 Introduction

Process mining techniques allow organizations to obtain insights that help them improve their processes [35]. The core idea of process mining is to exploit so-called *event logs*. These event logs are extracted from different IT systems that are used throughout the organization and, therefore, reveal how processes are actually executed [135]. Process mining has been successfully applied in various domains, including healthcare [93], auditing [59], and supply chain management [65].

Despite the success and popularity of process mining in practice, there is a limited understanding of how process mining insights eventually lead to process improvements. Specifically, it is unclear which *actions* organizations can consider based on the insights they have obtained through process mining. Existing process mining methodologies (e.g. [38]) provide structured guidance on how to use process mining to obtain relevant insights. However, they do not provide details on how to translate these insights into process improvements. We argue that understanding this *realm of actions* is a valuable aspect to complement existing process mining methodologies. By understanding which actions toward process improvement can be taken and how they are connected to the obtained insights, organizations can more easily identify the best path toward process improvement.

Against this background, we use this work to address the following research question: "*What are the actions organizations can take towards process improvement and how are they connected to process mining insights?*". To answer our research question, we conduct a systematic literature review. Based on the identified literature, we first investigate the different actions that are recommended or performed by process mining projects. Second, we investigate which insights lead to specific actions. Finally, we derive an overview of the actions triggered by process mining insights. Our contribution, therefore, is a systematic overview of actions, insights, and their connection. What is more, we identify the *intervention space*, i.e., the aspects of the organization that are affected by the actions, since process improvements actions may not only concern the process itself.

The remainder of this paper is structured as follows. In Section 4.2, we present the background and highlight the research gap. In Section 4.3, we describe our research method. In Section 4.4, we report our findings. In Section 4.5, we provide a reflection on our findings and, finally, Section 4.6 concludes the paper.

4.2 Background

In this section, we introduce the background for our work and highlight the research gap. First, we briefly explain what process mining is and what it offers to organizations. Second, we elaborate on process mining methodologies, i.e. works that specify which steps organizations need to take to successfully apply process mining. Third, we discuss the relationship between process mining and process improvement and argue that there is a missing link between the two.

Process Mining. Process mining is a family of techniques that facilitate the analysis of business processes based on so-called event logs [35, 133]. These event logs are extracted from different types of information systems that support the process execution and are usually captured using the dedicated and standardized format XES [140]. It is important to highlight that event logs are not available per se and that the extraction of event logs from information systems may require considerable manual effort [119]. Once an event log is available, different types of analyses can be performed. The three most prominent process mining use cases in practice include process discovery, conformance checking, and enhancement [133]. The goal of process discovery is to generate a process model from the given event log that appropriately captures the as-is process. In conformance checking, a normative process model (capturing the desired process) is compared against the event log to detect deviations. Enhancement relates to a variety of use cases where a process model (e.g., discovered by means of process discovery) is enriched with additional information such as execution time, resources, or costs. Among others, this facilitates predictions related to the remaining execution time or the chances of successful process completion.

Process Mining Methodologies. Different process mining methodologies have been developed [16, 133, 138] with the goal of supporting process mining initiatives in practice. They typically outline specific steps, such as defining scope, collecting data, applying process discovery or conformance-checking techniques, analyzing results, and improving processes. Although these methodologies generally follow a similar high-level flow, they often do not provide specific guidance on how to translate process mining insights into process improvements [38] nor do they outline the different actions that *could* be used to follow up on the obtained insights.

The authors of the Process Diagnostics Methodology [16] recognize the importance of the recommendation phase (i.e., results transfer) of a process mining project. However, they make clear that it is the organization's responsibility to interpret and take action based on the acquired process mining insights. Although the authors of PM^2 [138] recognize the importance of the process improvement phase, they argue that this is usually part of a separate project. The authors of L* [133] propose improvement actions (e.g., redesigning, intervening) to follow up on the acquired insights. However, they do not provide much details about these actions.

Process Mining for Process Improvement. A key driver behind the application of process mining for many organizations is the desire to improve their business processes. However, successfully using process mining for process improvement comes with several challenges. Recognizing this, several studies investigate how process mining is implemented and, among others, identify key success factors [86] and key challenges for the adoption of process mining [67, 89].

Other studies also more explicitly focus on the link between process mining and process improvement. For example, Eggers et al. [37] investigate how process mining can support improving process awareness in organizations. They identify seven mechanisms related to achieving increased process awareness pertaining to, for example, the inter-individual process level (i.e., when stakeholders share awareness of their sub-process within one department) or the inter-functional process level (i.e., when stakeholders share awareness of the end-to-end process across different departments). Lashkevich et al. [74] develop an analysis template to support identifying improvement opportunities based on process mining insights systematically. In their paper, they provide an example of a template relating to bottleneck analysis.

What is currently still missing is a comprehensive understanding of the actions that can be used to follow up on process mining insights. We believe that making these actions explicit can help organizations to understand the different options they can consider and, in this way, complement existing process mining methodologies.

4.3 Research Method

To answer our research question, we conducted a systematic literature review according to [68, 102], which involves four main stages: 1) literature review protocol definition, 2) study selection and data extraction processes execution, 3) data analysis, and 4) reporting. To ensure reproducibility, we involved several authors in these four stages. Three of the authors were involved in defining the literature review protocol. The search string and exclusion criteria were applied via the search engines by one of the authors, as defined in the review protocol. The inclusion criteria were defined and applied by two authors independently. Finally, two authors conducted the data extraction while discussing with the other authors the derivation of codes and themes reported in Section 4.4. We resolved disagreements through discussions among the authors. Below, we discuss the first three stages of our literature review. In Section 4.4, we report our findings.

4.3.1 Literature Review Protocol Definition

In this stage, we defined the research question and the study selection and data extraction processes. We were particularly interested in identifying which actions are performed after process mining insights have been acquired.

Based on our research question, we defined the following search string: "(process mining) AND ('case study' OR 'case studies') AND (application OR apply OR applied)", to focus on process mining application and not in, for example, the implementation of a new process discovery technique. Then, inspired by other literature review studies in the process mining field [35, 147], we defined the following set of search engines to apply our search string on: ACM Digital Libray, IEEE Xplore, Science Direct, Scopus, and Web of Science. We did consider including Springer Link in the set of search engines. Still, based on a pilot run of our study selection process, we identified that it would only add duplicates to the papers retrieved by the other search engines.

We defined exclusion and inclusion criteria to support our study selection process composed of four main stages: 1) application of search string into search engines, 2) application of exclusion criteria, 3) removal of duplicates, and 4) application of inclusion criteria. A study selection process of a systematic literature review determines how the exclusion and inclusion criteria will be applied to derive the final set of papers to be fully read [68]. The following exclusion criteria were defined: a) the paper is not written in English, b) the paper is not a conference paper, journal, or book chapter, c) the paper is not from computer science, decision sciences, business, management and accounting, healthcare, or social sciences. We further defined the following inclusion criteria: a) the paper is about the application of process mining or the use of process mining in a case study and b) the paper discusses what happens with process mining insights after they have been acquired.

The studies conforming to both inclusion criteria were kept and then further analyzed in the study selection and data extraction stage (cf. Section 4.3.2). The exclusion criteria supported us in filtering out papers directly from the search engines and the inclusion criteria supported us in deciding which papers were to be fully read, via a three-step application of the inclusion criteria, further detailed in the next section.

4.3.2 Study Selection and Data Extraction

Figure 4.1 presents our study selection process and shows the number of papers obtained from the execution of each stage. We applied the search string to the search engines, applied the exclusion criteria to the resulting papers, and removed duplicates. Then, we applied the inclusion criteria via a three-stage screening of the remainder papers: first, we screened the papers' titles and keywords (and their abstracts, when it was not yet clear if the paper should be excluded); second, we screened the abstracts of the remaining papers (and, in some cases, the conclusions); third and, finally, we screened the conclusions (and, in some cases, the methodology or the full text) to then reach to the final set of 57 selected papers to be fully read.

For the data extraction process, we imported the complete list of 57 selected papers into an evidence table where we kept track of the following features

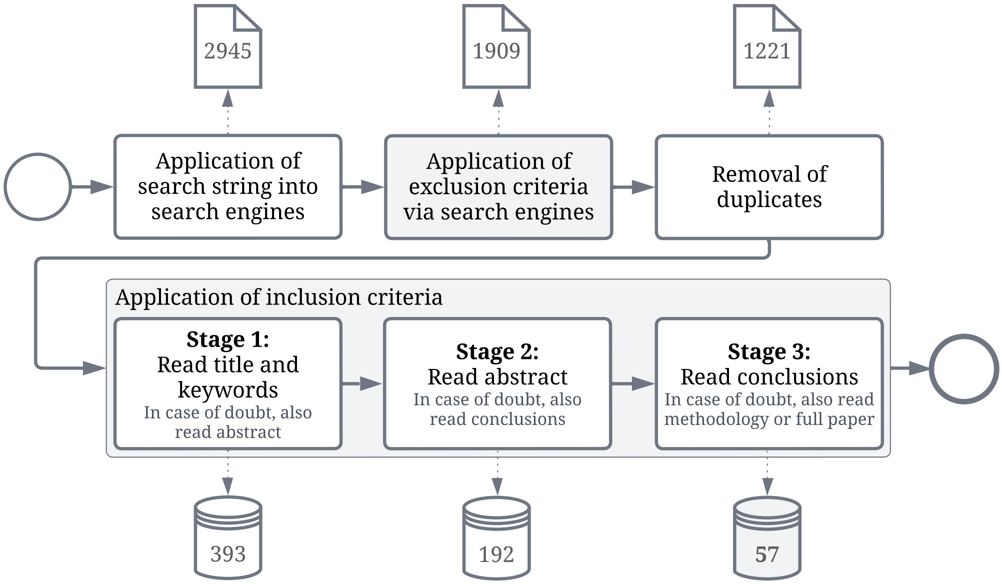


Figure 4.1: Study selection process with number of papers yielded per stage.

extracted from each paper: reported insight, quote (from which the action was coded), coded action (i.e., what happened -or was recommended- triggered by the reported insight), action sphere (i.e., the coded action was either performed or recommended).

4.3.3 Data Analysis

We conducted an inductive content analysis with open coding, inspired by [113], to make sense of the extracted data from the selected papers resulting from our study selection process. The generated codes were grouped into different higher-level categories. The themes naturally emerged from the categories. In Section 4.4, we present the themes, categories, subcategories, and codes derived from the selected papers of our literature study.

While performing the coding of the quotes extracted from each of the 57 selected papers, the codes naturally assumed the format *verb* + *object*, which then enabled us to categorize our findings in terms of "actions" (verbs) and "intervention space" (composed by objects target of the actions). An example of a code that emerged from our open coding is "update documentation" where we have the *verb* "update" and the *object* "documentation". Because of the high amount of different verbs related to the same objects (e.g., information system, process case, etc.) of the intervention space, we identified a clear pattern pointing out the important role the objects target of the actions themselves play in understanding the realm of actions related to translating process mining insights into process improvement.

In total, 156 quotes related to what happened with process mining insights after they had been acquired were extracted from the 57 papers fully read. Because each quote may derive one or more codes, summing up all supporting quotes for each category leads to a total of 226 supporting quotes. For example, we derived the codes "justify conduct" and "clarify conduct" out of the following quote from [85]: "*The use of both manual and online document approval by the director needs to be justified and clarified whether it will be a permanent practice* (...)". As another example, the codes "identify data quality issues" and "adjust data quality issues" were derived out of the following quote: "*The business improvement team will use the conformance checking results to identify and rectify potential* (...) *data quality issues*" [79].

4.4 Findings

In this section, we present the findings of our paper. In Section 4.4.1, we first provide a high-level overview. In Sections 4.4.2 through 4.4.4, we then take a detailed look into three themes we identified and discuss the specific actions for each theme. Finally, in Section 4.4.5, we discuss the most recurrent insights and the actions they trigger.

4.4.1 Overview

Studies reporting on what organizations do after they have acquired insights through process mining refer to both actions performed and recommended (i.e., actions to be performed). In this paper, to develop an overview of the realm of actions that can be triggered by process mining insights, we consider both kinds of reported actions simply as "action". The rationale behind this decision is that the recommendations are made by experienced professionals in the field and, therefore, can be considered as feasible. As a result, we identified three main themes of actions: i) supporting process understanding and documentation; ii) improving the involved information system supporting the investigated process; and iii) improving the investigated process. Each theme refers to one or more *intervention spaces*, such as *analysis* or *documentation*.

Figure 4.2 summarizes our results visually. It shows the main themes (dark gray), the intervention spaces (light gray), and the objects (white background) that are related to the intervention space. The numbers attached to the intervention spaces and the objects reveal the total number of supporting quotes from the analyzed papers. While these numbers should not be interpreted as a relevance factor, they do indicate how frequently a certain intervention space or object is the subject of an action after a process mining analysis.

In the next sections, we discuss each theme in more detail and provide a snapshot with respect to the identified actions.

4.4.2 Supporting Process Understanding and Documentation

This theme contains actions related to three intervention spaces: analysis, documentation, and communication and training. Next, we discuss each intervention space in detail.

Analysis. This intervention space contains actions related to different flavors of investigation that can be triggered by process mining insights. Several studies report on conducting or specifying follow-up investigations [2, 60, 139]. As an example, consider the domain expert checking if the identified relationships among members of collaboration groups match the designed procedures [60]. Other studies report on simulating or testing recommended proposed changes [90, 108, 131, 139]. Other kinds of follow-up investigation are related to investigating or discussing root-causes of the insights [24, 137, 146]. For example, in [24], the authors investigated the causes of a high ticket resolution time variance. Studies also report on investigating causality or correlation [2, 103, 112]. In [112], the authors investigated the causal relation between two different activities of interest, while in [2], the authors investigated (alongside experts) the correlation between the involvement of specific organizational units and the process performance achievement. Studies also reported on clarifying or justifying conduct related to unexpected

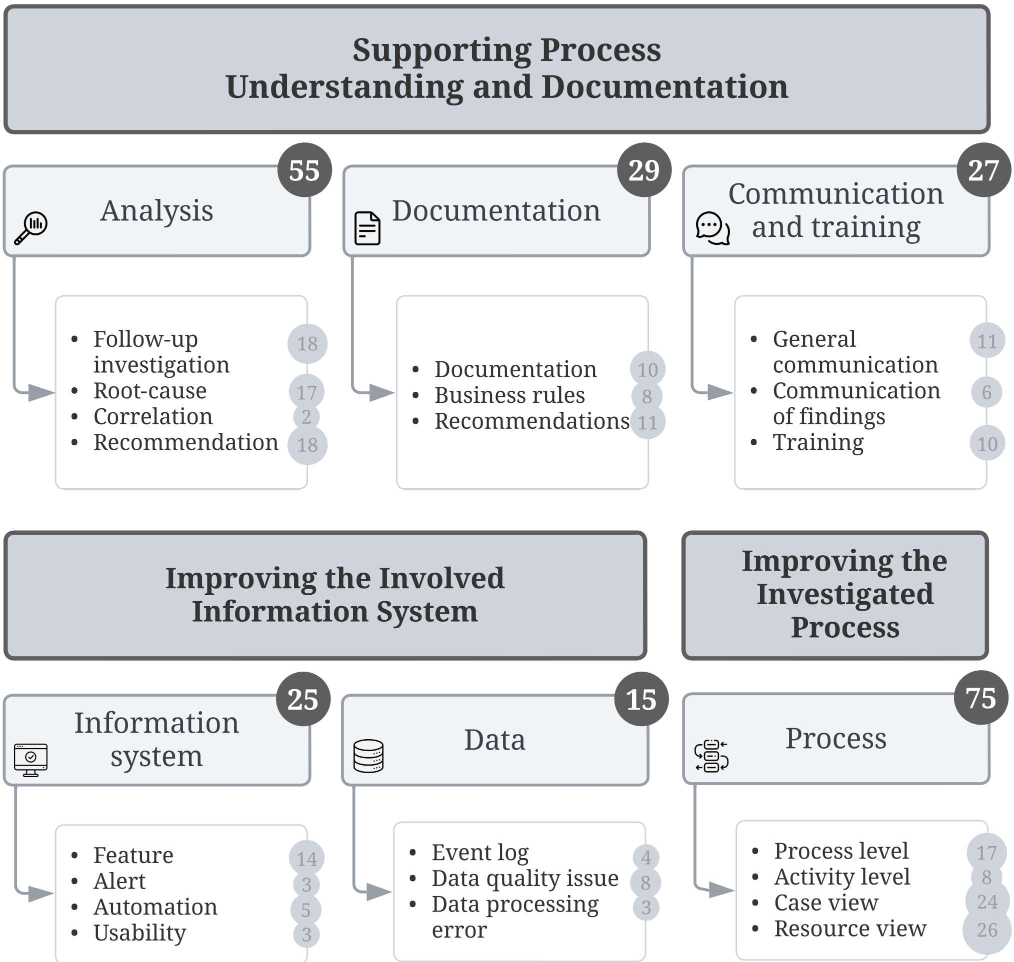


Figure 4.2: Themes and intervention spaces of actions triggered by process mining insights as reported in the literature.

behavior in process cases [85, 112], reviewing performance indicator [146], deriving background arguments to support decision making [50, 92] and deriving improvement initiatives [1, 33, 85].

Documentation. This intervention space contains actions directed at the documentation itself, business rules within a documentation, and reports (i.e., a specific type of documentation that will be used to report on process mining findings). Actions related to the documentation itself include: creating, reviewing, updating, and improving a documentation [3, 79, 101] or using specific techniques to organize information (e.g., ontologies) to add up to a documentation [3]. In [79], the authors mention that the directly-follows graphs were used to document processes with outdated or missing documentation. Regarding business rules and documentation, there are papers that refer to isolating, adapting, or reviewing business rules [24, 50, 59]. For example, in [59], the authors mention the need to review business rules in the sense of checking if they are being enforced. In addition, other papers refer to adjusting service level agreements [106, 115] based on process mining insights. Finally, regarding reporting, papers refer to writing, sharing, and presenting a report with the acquired insights [78, 99, 103], as well as formulating recommendations to be added to the report [139].

Communication and training. This intervention space contains actions concerned with communication, information sharing, and training. Actions regarding communication include challenging conventional beliefs, increasing awareness about the process, or creating an ad-hoc custom visualization for communicating findings [50, 77, 99]. Regarding information sharing, reported actions are related to providing feedback on performance measurements [61, 137]. For example, in [137], the authors provide feedback regarding specific detected loops in the process under investigation. Other papers report on discussing the likelihood of partial findings [99, 112], informing the manager about specific findings [137], or improving information sharing [2, 112, 146] to, for example, improve coordination between collaborating stakeholders [146]. Training may be used to reinforce internal controls or good practices [142, 143]. Other studies report on conducting training for staff members [48, 101, 143] leading to, for example, quality improvement, or reducing the need to perform a specific corrective activity [101]. Also, the discussion of potential training issues [79] was reported as an action triggered by process mining insights.

4.4.3 Improving Involved Information System

This theme contains the actions related to two intervention spaces: information system and data. Next, we discuss each intervention space in detail.

Information system. This intervention space contains actions directed at the information system(s) of the organization. They include creating, introducing, or testing a new feature [26, 50, 56, 85], testing or improving test scripts of a feature [112], or simply using existing features [2]. For example, in [85], the authors used process mining in the context of the adoption of a new Enterprise Resource Planning system. Based on the insights, they identified the need to reinforce testing new features to ensure they behave as expected. Other papers report on adjusting feature settings [24, 71, 76], creating alerts [53, 59], improving system usability [112, 129], identifying automation opportunities [10, 49], and adopting or implementing automation [42, 106, 143]. For example, in [49], the authors report on identifying automation opportunities by calculating the ratio between cases in which an activity was executed by a user and the total number of instances of the activity under investigation.

Data. This intervention space contains actions directed at the event log. They include filtering or re-collecting the event log [1, 56, 114], as well as identifying or rectifying data quality issues [79, 114, 128, 143]. In [143], the authors reported deriving recommendations for improving data quality issues of the event log based on the argument that the input data quality interferes with the quality of a process mining project. They recommended, for example, the verb-object naming style for activities and keeping track of both start and end timestamps of activities, as these are helpful for process analysis. Other papers report identifying, reviewing, or rectifying data processing errors [79, 146]. For example, in [79], the authors discuss that the business improvement team would use conformance checking-related insights to identify and rectify data processing errors and data quality issues. Note that the actions related to identifying or reviewing could also fit into the intervention space "Analysis". However, because the papers explicitly discuss these actions being directed at the event log itself, we included them in the "Data" intervention space.

4.4.4 Improving the Investigated Process

This theme represents the intervention space that contains actions toward the process. These actions can be more generic, such as redesigning, simplifying, changing, or standardizing the process [33, 45, 56, 85]. However, the actions can also be more specific toward a particular aspect of the process, such as isolating or checking potential deviation in cases [59, 60]. The actions can also refer to analyzing process cases [42, 61, 79] such as in [42], where the authors report on analyzing process cases containing high time-consuming tasks [42] or, as in [60], where the authors report on analyzing process cases from a resource perspective.

Some papers refer to actions directed at activities, such as parallelizing, removing, increasing the frequency of, or limiting, preventing or postponing the execution of an activity [2, 17, 31]. For example, in [31], the authors report preventing customers from going through a specific activity multiple times. Several papers refer to actions toward resources, such as waiting for, involving more actively, increasing, replacing, reallocating, aggregating, manually inspecting, increasing the visibility of, protecting, or restricting access to a resource [24, 53, 126]. In [126], the authors observed that the pattern separation of duty should be applied to restrict access to certain parts of the information systems only to specific employees. In [24], the authors reported that the stakeholders are considering increasing the number of developers to resolve a ticket resolution time issue. Other papers refer to actions toward specific detected patterns in the process (either desired, i.e. good practice, or undesired), such as adopting or removing specific patterns [33], defining or improving good practices [3, 146]. Other papers report on identifying or understanding specific patterns, or identifying the following –or possible exploitation– of good practices [61, 101]. Although these papers reporting on identifying or understanding specific patterns or good practices could fit the intervention space “Analysis”, we kept them under the “Process” intervention space because of the explicit relation to the intervention to the process.

4.4.5 Most Frequently Reported Insights and Actions

The most recurrent insights reported in the literature are related to:

1. *Low data quality*: Low data quality may refer to both the data from the databases as well as the event log itself. For example, in [114], the authors reported on missing fields in records and incorrect event sequences in the event log. In [57], the authors reported on identifying incomplete traces. As such data quality issues may compromise the validity of the obtained insights, they need to be addressed before any further action can be taken.

2. *High wait time*: High wait time is a common concern in different contexts. For example, in [129], the authors noticed that it was taking more time for a process participant to take over a specific task than to work on that task. In [66], the authors report on identifying the delivery of goods taking longer than the defined service standard.

3. *High amount of rework*: Rework is another frequent concern. For example, the authors of [143] identified rework caused by manually misclassified documents. In [112], the authors identified that a specific system feature-related data had not been cached, requiring the user to unnecessarily repeat the execution of another related task within the system.

4. *Discovered process model*: The discovered process model is used for a variety of purposes. If, however, the discovered process model does not allow the analyst to obtain the required insights, this might be addressed before any further action can be taken. For example, the authors of [128] obtained spaghetti-like process models, which did not allow them to conduct a proper analysis of the process. In [45], the authors discussed the suitability of the discovered process model to support the definition of a standard process.

5. *Non-compliant behavior*: Besides performance-related insights, non-compliant behavior, i.e. conformance violations, represent a very common trigger for actions. For example, in [85], the authors identified actions performed by process participants that were not conforming to the expected behavior. In [142], the authors reported on an inward cargo handling where they identified many instances of the process that did not properly complete according to a normative process model.

Other insights refer to, for example, high or low demand on a specific resource (e.g., process participant), high execution time of specific activity, low automation rate, lack of domain knowledge, among others.

Figure 4.3 presents the top five most frequently reported insights and actions in the literature, respectively, in terms of the number of supporting quotes from the selected papers, as described in Section 4.3.3. For the reader to distinguish with ease the connections between insights and the process-related artefacts of the intervention space, we chose to repeat both action verbs and

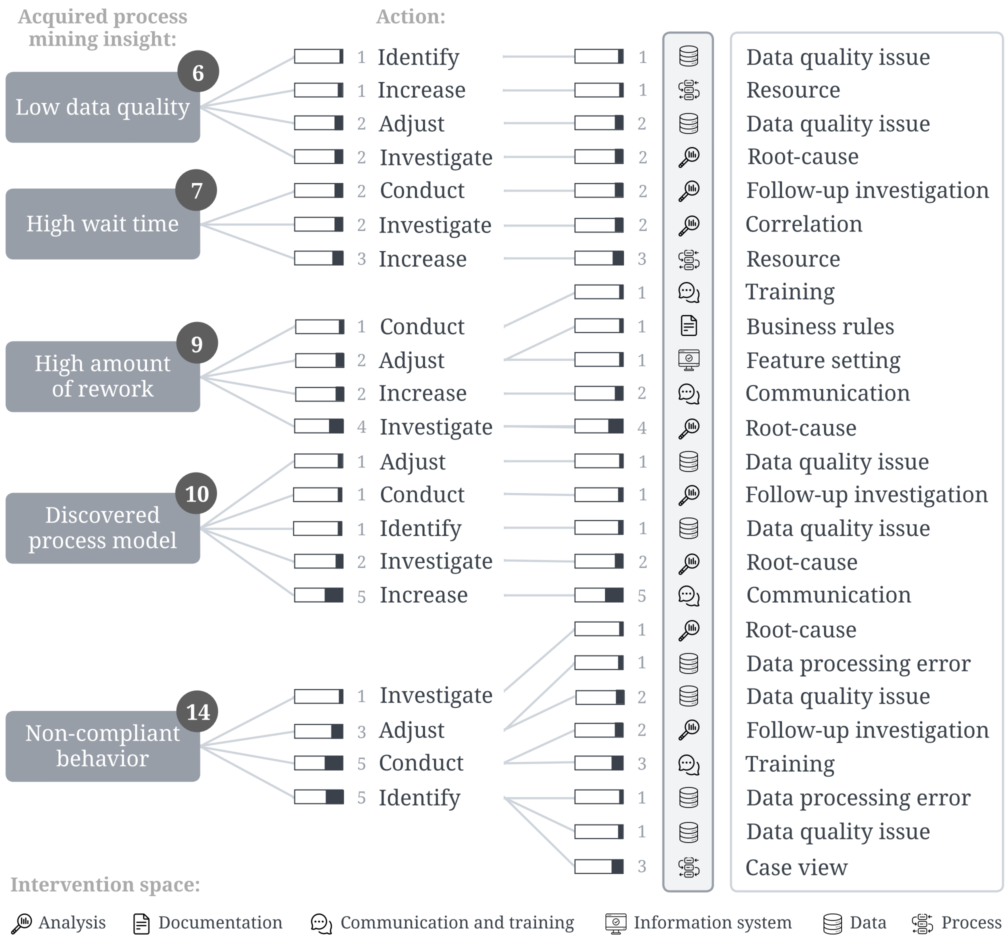


Figure 4.3: Most frequently reported process mining insights and triggered actions.

objects as triggered for each reportedly acquired process mining insight. Two aspects stand out from Figure 4.3. First, the intervention space is quite large, highlighting that process mining insights not only trigger interventions to the process under investigation, but also to process-related artefacts. Second, there are many-to-many relations between acquired process mining insights and triggered actions, as well as between insights and process-related artefacts, objects of the intervention space. Have in mind that because there is a wider variety of actions than insights reported and the same amount of supporting quotes connecting insights and actions, the amount of supporting quotes for the most recurrent reported action is lower than for the most recurrent reported insight.

4.5 Discussion

From our findings, it is clear that translating process mining insights into process improvement requires cooperation and coordination as different levels of knowledge and expertise are needed to support understanding and documenting the process, improving the involved information system supporting the investigated process, and improving the investigated process. We can reason that there is a need to properly operationalize the integration between technical and organizational workers to intervene in the process itself and in the underlying information system(s) while understanding and documenting the process, both coordinately and supported by domain and data knowledge.

While we do not claim to provide a comprehensive list of actions triggered by process mining insights, we provide an initial step toward understanding the diversity of the insight-to-action realm. We acknowledge that further research is needed to investigate the representativeness of the actions herein shown and a deeper understanding of each action, precisely detailing what they entail. Only then we'll be able to move towards ultimately recommending assertively follow-up actions from acquired process mining insights and triggering (semi-) automated actions not only related to prediction-based alert systems but also towards undertaking specific changes to the process or process-related artefacts.

Having this said, there are several important findings we were able to derive in this paper. Below, we discuss the three main points.

Actions are concerned with much more than the process itself. Intuitively, one would expect that process improvement is mostly about the process itself, especially when the basis for improvement are insights obtained through process mining. Our findings show that the investigated process is indeed subject to several actions, such as parallelizing or removing activities. We, however, could also show that there are several intervention spaces besides the process itself. Among others, we identified that actions are taken towards understanding or improving other process-related artefacts, such as documentation, communication, training, and supporting information systems. These findings highlight that process improvement requires a holistic view, including several facets, such as the IT infrastructure and the human resources that are involved in the process execution.

The relationship between insights and actions is highly complex. Our analysis revealed that there is a many-to-many relation between insights and actions. This means that one insight can trigger several actions and that one action can be triggered by several insights. While this is not totally unexpected, it helps to better understand the relationship between insights and actions. What is more, researchers and practitioners conducting a new process mining initiative can plan ahead for actions they may need to perform based on the insight they have obtained. In addition, they can also acquire a broader vision of potential insights to consider obtaining. Assume the actions they may need to perform are related to another insight that was not previously considered to be obtained. In this case, this not previously considered to be obtained insight could be added to the pool of insights to be acquired.

Gap between recommended and taken actions. During our analysis, we observed a gap between recommended and taken actions. For several insights, e.g., high wait time and high rework rate, we identified recommended but not any taken actions. This observation shows that certain actions seem to be either associated with too much effort or they are not considered for other reasons. While we cannot provide specific insights into why this gap exists, it is important to note that it is there. We believe that this represents an important direction for future work: understanding which actions are (not) performed to improve processes and why.

4.6 Conclusion

In this paper, we used a structured literature review to investigate which types of actions organizations have taken in the context of process mining initiatives and to which insights these actions are linked. We found that there exists a large variety of actions and that many of these actions do not only relate to changes to the investigated process but also to the associated information systems, the process documentation, the communication between staff members, and personnel training.

With these findings, our study provides an important step towards enhancing the implementation phase (as reported by Emamjome et al. [38]) of existing process mining methodologies. Specifically, the derived overview of actions triggered by process mining insights can serve as a catalog for practitioners that aim to translate process mining insights into actual process improvement. We believe that such a catalog can be particularly useful for novice process mining consultants and managers as it provides guidance on which actions they might consider given particular insights. Such catalog may also support practitioners sketching initial *plans of action* for their projects, supported by evidence from real-life case studies.

From an academic point of view, our results complement existing process mining methodologies, such as [16, 133, 138]. By including our findings, it is possible to devise a methodological framework for process mining that does not stop with obtaining insights but with realizing process improvements. Having this said, there are several aspects that require further investigation. First, it is interesting to conduct a deep investigation of what each action entails. For example, what are the different departments and personnel involved and what were the challenges faced while implementing a specific action. Second, it would be useful to conduct case studies with successful and unsuccessful process mining projects to highlight commonalities and differences between these projects and further understand which actions ultimately lead to a successful translation of process mining insights into process improvement.

In future work, we will conduct a survey with experts and a multiple case study to complement the intervention space taxonomy presented in this study. We will further investigate the relations between recommended and performed actions to move towards well-informed recommendations supported by performed actions. Finally, we will derive a catalog of the many-to-many relations between insights and the affected process or process-related artefacts.

Chapter 5

A Study on the Terminators of Process Mining Initiatives

Abstract

Process mining has been used to obtain insights into work processes in various industries. While there is plenty of evidence that process mining has helped a number of organizations to improve their processes, there are also a few studies indicating that it did not happen in other cases. An obvious yet frequently overlooked challenge in that context is that organizations actually need to take action based on the insights process mining tools and techniques provide. In practice, analysts typically use process mining insights to recommend actions, which then need to be performed and implemented, for example, by process owners or management. If, however, recommended actions are not performed, the insights will not help organizations to progress into process improvement either. Recognizing this, we use this paper to develop a better understanding of the extent to which recommended actions are actually performed, as well as the causes hampering the progress from recommended to performed actions. To this end, we combine a systematic literature review involving 57 papers with 17 semi-structured interviews of process mining experts. Based on our analysis, we discover specific causes why organizations do not perform recommended actions. These findings are crucial for both researchers and organizations to develop measures to anticipate and mitigate these causes.

Published in V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, I. Beerepoot, and H. A. Reijers. From loss of interest to denial: A study on the terminators of process mining initiatives. In International Conference on Advanced Information Systems Engineering (CAiSE), 2024 [123]

5.1 Introduction

Process mining is used to obtain insights into work processes in organizations. It has been successfully employed in various industries including manufacturing [14], finance [28], and healthcare [93]. Among others, process mining techniques support discovering, analyzing, and improving processes [107, 133]. While there is plenty of evidence that process mining can effectively help organizations improve their processes [9, 67], there are also a few studies indicating that it was not successful in other cases [86, 121].

Taking a closer look at scholarly work on process mining endeavors, we observe that it is often concerned with proposing process mining techniques to obtain specific *insights*. Recent examples include the work from Lashkevich et al. [73], who propose an approach providing insights into the causes of wait times and their impact. Another example is the work by Bozorgi et al. [18], who propose a prescriptive monitoring method that decides which process instances are worth intervening to obtain a desired outcome. What these and many other techniques have in common is that they require humans to take *action*. In practice, analysts typically use the insights to *recommend* actions, which then need to be *performed*, for example, by process owners or management. If, however, recommended actions are not performed, the insights will also not help organizations to improve their processes.

Recognizing this, we argue that we need to develop a better understanding of the extent to which recommended actions are performed, as well as actually the causes hampering the progress from recommended to performed actions. While existing work has studied the relation between process mining insights and actions on a general level [120], the distinction between recommended and performed actions has not been made. As a result, there is also no understanding of when or why recommended actions are not being followed up by performed actions. In this work, we address this research gap by posing the following two research questions:

RQ1. To what extent are recommended actions also performed?

RQ2. What are the causes for certain recommended actions not resulting in performed actions?

To answer these research questions, we combine a systematic literature review and semi-structured interviews. We analyzed a total of 57 papers, as well as 17 transcripts from semi-structured interviews with process mining experts. In this way, we cross-validate and corroborate our findings, enhancing reliability and validity of our research. We identify five specific causes why organizations do not perform recommended actions. These findings are crucial for researchers and organizations to develop strategies to anticipate and mitigate these causes.

The remainder of this paper is organized as follows. Section 5.2 discusses the background and related work. Section 5.3 introduces our methodology and research method. Section 5.4 presents the identified causes hampering the progress of recommended to performed actions, while Section 5.5 discusses the high-level relation of the identified causes to process mining projects. Finally, Section 5.6 concludes the paper.

5.2 Background

In this section, we introduce the background and related work of our research. We first provide a brief overview of process mining and process mining methodologies. We then highlight the research gap.

Process mining is a set of specialized data analysis techniques. It leverages so-called event logs to provide insights into the execution of work processes [133]. Among others, it allows organizations to automatically discover as-is processes and to detect violations against rules and regulations. To support the application of process mining in organization, different process mining methodologies have been proposed and adopted. Two widely known process mining methodologies are Process Diagnostics [16] and PM^2 [138]. Although these methodologies highlight the importance of progressing process mining insights into process improvement, they also acknowledge that this is an aspect they consider out of scope. That is, because they primarily focus on analytical techniques rather than the practical implementation of the acquired insights. According to Emamjome et al. [38], this lack of practical implementation of the acquired process mining insights has hindered process mining in delivering the promised outcomes, despite the growing adoption of process mining.

However, several studies also investigate the application of process mining and the associated challenges. In [86], the authors analyze reports of process mining case studies to identify critical success factors. In [149], the authors use semi-structured interviews to investigate the challenges regarding process mining analysis. They identified a set of over twenty challenges highlighting the need for enhanced support for acquiring process mining insights. Similarly, the authors in [121] used interviews with process mining experts to derive challenges organizations face when progressing process mining insights to process improvement. Finally, the authors in [120] have also analyzed which type of insights lead to what type of actions in the organizations. They introduce the notion of an *intervention space* to conceptualize what aspects of an organisation (e.g., the process itself or its underlying IT infrastructure) are affected by improvement actions.

What is currently still missing is an understanding of when and why certain actions triggered by process mining insights are only recommended but not performed. Insights into this phenomenon can serve as a basis for augmenting process mining methodologies. What is more, it can support organizations proactively mitigating issues that can potentially lead to the termination of their process mining initiatives.

5.3 Research Method

To answer our research questions, we conducted a systematic literature review and a series of semi-structured interviews with process mining experts to provide an integrated overview of literature and practice. Figure 5.1 presents an overview of our methodology composed by three main stages. First, we conducted a *data collection* (cf. Section 5.3.1) on two data sources: a systematic literature review and a series of semi-structured interviews. Second, we conducted a *data analysis* (cf., Section 5.3.2) on the data acquired from both the literature and the interviews. Third, we *synthesized* our findings (cf., Section 5.3.3) from both data sources into one to derive an integrated view of literature and practice.

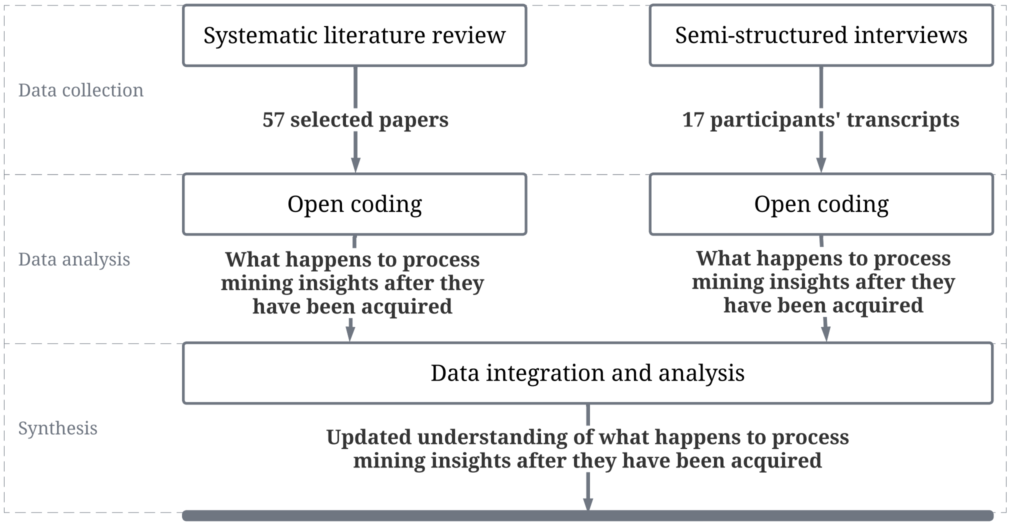


Figure 5.1: Methodology overview

5.3.1 Data collection

Our data collection consists of two steps. First, inspired by [68, 102], we conducted a systematic literature review. Then, based on [15, 19], we conducted a series of semi-structured interviews.

Systematic Literature Review. We conducted a systematic literature review by defining a review protocol that encompasses the research questions and the definition of the search string, search engines, and the papers selection criteria [68, 102]. Then, we applied the search string to the search engines, followed by the paper selection by applying our selection criteria. These stages were collaboratively conducted by several authors, discussing and resolving disagreements altogether in order to mitigate threats to validity regarding reproducibility.

Because we wanted to obtain insights into the application of process mining, we were particularly interested in case studies. Hence, we defined our search string as "(process mining) AND ('case study' OR 'case studies') AND (application OR apply OR applied)". To decide on the search engines to be used, we investigated other systematic literature reviews in the process mining field [35, 147] and adopted the following search engines: ACM Digital Libray, IEEE Xplore, Science Direct, Scopus, and Web of Science. We defined the following selection criteria: i) the paper is about the use of process mining in a case study; ii) the paper mentions what happens to process mining insights; iii) the paper is either from accounting, business, computer or decision or social sciences, healthcare, or management; iv) the paper is published either in conference proceedings or a journal; v) the paper is written in English. The final set of selected papers amounts to 57 peer-reviewed papers, which together provide an overview of the extent to which recommended actions are also performed from the literature perspective.

Semi-Structured Interviews. We conducted a series of semi-structured interviews with process mining experts as our target population and defined the interview protocol based on [15, 19]. We invited participants via email, reinforcing the invitation via LinkedIn. In total, 17 experts averaging seven years of experience in process mining were interviewed. Over 64% of the participants pursued their PhDs in the process mining field, on top of their years in industry. The interviewees are from four continents, and 82% of them are from organizations with over one thousand employees. They have used different process mining tools including Celonis, ARIS Process Mining, UiPath Process Mining, SAP Signavio, and Fluxicon Disco. The interviews took on average 54 minutes.

We asked participants about their role regarding process mining in their organization, and we asked them to share some examples of process mining initiatives that they had worked on. We wanted to see how recommended and performed actions would naturally emerge from the interviews. Therefore, we did not directly ask the interview participants what the recommended and performed actions were, nor which were the causes of a process mining project termination. For each example shared by the interviewees, we asked what happened to process mining insights after they had been acquired. The final set of collected interview transcripts together provides an overview of the extent to which recommended actions are also performed from the practitioners' perspective.

5.3.2 Data Analysis

Based on [113], we conducted separate open codings on the data acquired from both the literature and the interviews. To do so, our qualitative coding process was composed of three main stages: i) getting acquainted with the data by reading it and taking broad notes, ii) re-reading the data and generating codes, and iii) reviewing the codes to merge semantically similar ones into categories. We coded each of the selected research papers and interview transcripts whenever it discussed a recommended (R) or performed (P) action or a cause for not performing the recommended action. Then, we categorized the actions into different types of process mining insights, such as data quality, wait time, etc. This categorization allowed us to (1) identify specific instances where recommended actions did not materialize into actual process improvement and (2) determine their causes.

5.3.3 Synthesis

To provide an integrated view of literature and practice about the extent to which recommended actions are also performed and the causes for certain recommended actions not resulting in performed actions, we integrated the data from the systematic literature review with data from the interviews. Doing so allowed us to quantify the frequency of actions, resulting in an overview of recommended and performed actions across different process mining insights, as reported in the literature and interviews. Interestingly, the literature does not report on the causes for not performing recommended actions. Thus, we drew on the interviews to derive those causes.

5.4 Findings

In this section, we present the results of our study. Section 5.4.1 first provides an overview of the extent to which recommended actions are performed (RQ1). Sections 5.4.2 through 5.4.6 then elaborate on the reasons why recommended actions are not performed (RQ2).

5.4.1 Overview

In Table 5.1, we provide an overview of the extent to which recommended actions are performed. We grouped the reported actions into five types of insights that are typically acquired during process mining initiatives [120]:

- *Data quality*: Data quality refers to issues such as incomplete or inconsistent data. Several studies discuss that data quality is the starting point for recommended and performed actions [1, 148].
- *Wait time*: Wait time refers to insights indicating that waiting in the process leads to delays. As examples consider waiting for a resource to become available or waiting for a process participant to finish their task [2, 146].
- *Rework*: Process mining insights related to rework refer to, for example, an activity that is repeated unnecessarily [31, 112].
- *Discovered process*: In many cases, the discovered process is already an insight for the stakeholders. This is the case when, for example, the organization does not have their processes documented [50, 99].
- *Compliance*: Process mining insights related to compliance refer to, for example, a mismatch between the documented process and its execution [61, 79].

The top part of Table 5.1 shows the results based on the literature, the bottom part shows the results from the interviews. It illustrates that literature and interviews provide different viewpoints. Many papers report on recommended actions without discussing whether the recommendations were implemented or not. At the same time, case studies tend to report on successful process mining projects rather than unsuccessful ones. The interviews, by contrast, often discuss why actions were not performed and the particular reasons for that.

In the subsequent sections, we take a closer look at this phenomenon and discuss which are the causes that prevent organizations moving from recommended to performed actions. We refer to these causes as *terminators*. And because the terminators differ considerably depending on the insight that led to a recommended action in the first place, we use the five types of insights that are typically acquired during process mining initiatives.

5.4.2 Data quality

In many studies, insights related to data quality were acquired during the process mining initiative. Recommended actions include conducting a follow-up investigation to understand the root-cause of the data quality issues or

Table 5.1: Recommended (R) and performed (P) actions across the frequently reported process mining insights from literature and interviews

Data source	Process mining insights									
	Data quality		Wait time		Rework		Discovered process		Compliance	
	R	P	R	P	R	P	R	P	R	P
Literature	[41] [143]	[1] [114] [148]	[2] [42] [66] [92] [108] [129] [146]	[2] [112] [115] [148]	[10] [53] [106] [126] [143] [146]	[137] [2] [31] [66] [77] [112]	[1] [33] [139] [71] [78] [79] [124] [142] [146]	[23] [45] [50] [56] [79] [99] [103] [128]	[45] [48] [61] [79] [85] [112] [142]	[53] [60] [79] [101] [114]
Total literature support number for recommended or performed actions	2	3	7	4	6	6	9	8	7	5
Interview	I5 I9 I16	I12 I15	I8 I12 I17	I2 I3 I5 I6	I4 I17	I4 I6 I11 I15	I12 I13 I16	I3 I5 I8 I9 I11	I15	I2 I6 I14 I15
Total interviews support number for recommended or performed actions	3	2	3	4	2	4	3	5	1	4

adjusting the information system to start recording the end time of specific activities for further process analysis [1, 143]. However, these actions are not always performed. We identified two terminators related to the insight *data quality* that are causing such initiatives to stagnate.

Laborious data preparation. Interviewee I15 reported on identifying late payments related to an order-to-cash process, initially aiming to review the overall standardization of the process. However, they needed to restart the project because of low data quality for further analysis. The event log was re-extracted, and the project went through another round. However, this iterative process of data preparation and event log re-extraction was taking too long, and the project was eventually stopped. Similarly, Interviewee I16 reported on a procure-to-pay process where they learned that "*the ERP system users were not following the predefined steps in the system*". However, the "*organization lost interest in using process mining when they learned about the different levels of granularity in their data, and the work they would need to put to it to be able to extract an event log suitable for process mining*". Again, *laborious* data preparation of the data was considered as the terminator here. Interviewee I9 reported on a similar project, where the suggestion to track customers with bad payment records was not acted upon because it required data integration from different systems. This was perceived as too complex and resource-intensive. Interviewee I5 discussed a project where the automation rate needed to be added to an existing dashboard. However, because the required *data was not fully accessible*, and to make it accessible would be too *laborious*, the project was terminated.

Loss of interest. Interviewee I5 reported about the *lost of interest* in the process mining initiative being the cause for the initiative to be terminated. The interviewee mentioned that the stakeholders "*were not sure what to ask for. They had the idea that it would be good to use process mining, but they could not specify what they wanted*"; on top of that, there were more important projects to deal with. In the end, the stakeholders ended up *losing interest* in the initiative, and "*at some point, we just had to stop the project*".

5.4.3 Wait time

Process mining initiatives may also culminate in insights related to long wait times that lead to delays, triggering different recommended and performed actions [2, 146]. Below, we present the terminators we identified related to *wait time*.

Lack of expertise. Interviewee I12 referred to an eye surgery process where the department was "*eager about taking action on the acquired insights*", and "*they were keen on improving things*". One of the acquired insights was related to a high wait time before surgery. In this case, however, the project was halted before any improvements could be made because the process analyst of the project switched jobs. With the departure of this employee the organization *lost their expert* in process mining which hampered the continuity of the project.

Lack of incentive. Interviewee I17 referred to a femur surgery process where they found a high length of stay in the hospital before surgery. After discussing this with the stakeholders, no action was taken because of the way the "fee for services" is established. In some countries, doctors and hospitals can earn more money by asking for more exams or for more days in the hospital. As such, there was *no financial incentive* to implement the recommended action.

On the contrary, implementing the action would result in decreased income for the doctors and hospital.

Denial. Interviewee I8 described a process mining study about a financial support request procedure, uncovering excessive delays due to long-lasting checks. The proposed solution would streamline the process by automating several steps. Despite the clear benefits of the proposed solution, "*there were two managers that simply ignored the insights and recommendations, pushing their own solution, and nothing was done regarding the acquired process mining insights*".

5.4.4 Rework

Process mining insights related to rework can trigger different actions [31, 112]. We identified one process mining project terminator connected to *rework*.

Lack of incentive. Interviewee I17 reported on a case about a consultation with physicians, where they identified a high amount of unnecessary repeated activity. The interviewee learned that the repeated activities were related to unnecessary exams. After inquiring with the stakeholders about this issue, no action was taken because both the hospital and the physicians were financially benefiting from this behavior because of the funding scheme of hospitals. Again, there is *no incentive* for the stakeholders to implement the action, as it would affect them negatively. According to the interviewee, "*in some cases, discussing the issues with higher-level management staff in a hospital can help reduce bad practices, but this is rare, especially in public hospitals. In private hospitals, addressing bad practices is somewhat easier*".

5.4.5 Discovered process

The discovered process is also reported as an insight, as it can bring awareness and trigger different actions as, for example, identifying improvement opportunities or improving communication with stakeholders [50, 99]. Below, we present the terminators we identified for the insight *discovered process*.

Denial. Interviewee I13 reported on a case where they were working on a ticketing system handling outstanding payments of a big financial institution. The process mining analysis revealed that clients with outstanding payments were not properly charged. However, no action was taken because "*the* *stakeholder preferred to say that they do not believe process mining results rather than to admit they were wrong and taking the responsibility, disregarding the fact that they were losing a lot of money*". Another situation in which the insights were *denied* by those involved was mentioned by interviewee I12 that reported on a process mining initiative in a hospital related to a Gastroenterology process. The interviewee explained that the department was reluctant to take any actions based on the acquired process mining insights, that the staff were not enthusiastic either, and that they kept *denying* the acquired insights. As a result, the project was terminated.

Loss of interest. Interviewee I13 reported on an occasion related to a credit analysis process where the manager, after learning that the problem was not in their department, *lost interest* in the findings and decided to halt the project without sharing the insights.

Lack of expertise. Interviewee I4 referred to a meter-to-cash process of an energy provider, where the organization had a threshold to decide when to conduct a manual check on the energy consumption of a customer. With process mining, they learned this threshold should be readjusted because in 80% of the cases where a manual check was conducted, no deviation from the calculated consumption was detected. However, no action was taken because the organization *did not have the expertise* to make the required changes in the system and did not want to invest in a consultant.

5.4.6 Compliance

Literature reports on different process mining insights related to compliance, along with the corresponding triggered recommended and performed actions [61, 79]. Below, we present the terminator for performed actions we identified for the insight *compliance*.

Lack of incentive. Interviewee I14 reported on the handling of a service request process. The interviewee identified "*strangely equalized service delivery time*". In other words, service delivery time could take any amount of minutes when the service time took more than 10 minutes; otherwise, it was always set to 10 minutes. Despite the large potential impact on customer satisfaction, the project was shut down after the presentation of the findings.

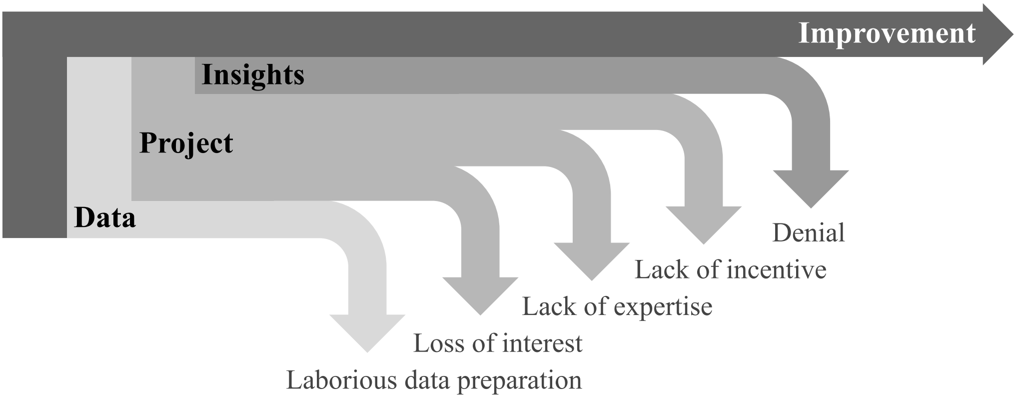


Figure 5.2: Causes for certain recommended actions not resulting in performed actions. These causes can stem from the data, the project itself, or the acquired process mining insights

5.5 Discussion

In the previous section, we reported on a total of five different *terminators*, i.e. causes why organizations were not progressing from recommended to performed actions. Below, we present our contributions, recommendations, and research limitations.

5.5.1 Contributions

The terminators identified in this study can be organized into three main dimensions: causes related to the analyzed process *data*, causes related to the process mining *project* itself, and causes related to the acquired *insights*. Figure 5.2 provides an overview of the relation between the different causes and the three dimensions. As indicated by the layout, the three main dimensions are in a temporal order. Below, we discuss each dimension and its related project termination causes in detail.

Data. Aspects related to data are also often responsible for a process mining initiative not moving forward. According to interviewees, process mining initiatives have been called off because of inaccessible data, low data quality, or complex data, making data preparation a laborious effort. As we know from the literature, many different event log generation [119] and event data preprocessing [83] tasks require proper expertise to be performed in order to generate an event log and improve the quality of the data for process mining purposes.

Project. A project can be terminated because there are other projects with higher priority in the backlog of the organization, diminishing the incentive for the project continuity, or because sponsors of the initiative may lose interest in the project for different reasons. Other causes for the termination of a process mining project may be the existence of an alternative solution that is pushed top-down into the organization, stagnation regarding technical knowledge from the staff, complacency with not following good practices, and the departure of key employees from the organization, which are related to lack of incentive, loss of interest, and lack of expertise.

Insights. There are also causes for the termination of a process mining initiative linked directly to the acquired insights. In some cases, there is a strong denial of the acquired insights. This can be a problem by itself or lead to the concealment of the acquired insights in an attempt to not undergo scrutiny of one's work or because the acquired insights are actually related to mediocre work from another department in the organization. Another aspect that can lead to the termination of a project is reluctance to take any actions based on the insights, which can be because of disbelief in the acquired insights.

5.5.2 Recommendations

Reflecting on the identified terminators, we derive two conjectures. First, data preparation for process mining remains (too) laborious for organizations. Although efforts have been made in the area of event data preprocessing and data quality improvement [83, 127], data quality remains a major challenge and a dominant factor in the termination of process mining initiatives. Organizations require the research community and technology vendors to pay more attention to the standardization of data and the provision of guidance for extraction and preprocessing.

Second, we need to rethink approaches to process improvement based on process mining insights. Our examples have shown that the employees to whom the process mining insights are presented are either (1) not the ones with the mandate to make changes to the process, or (2) not the ones who benefit from the changes. To follow up on the insights, it is insufficient to involve local, functional managers. Managers with decision-making power on end-to-end processes are required. Moreover, there needs to be a (financial) incentive for stakeholders to implement actions, or else projects are easily terminated. If process mining initiatives are to have a significant positive impact on an organization, the right management level needs to be involved.

5.5.3 Limitations

Our work is subject to limitations, such as generalizability, because it is a qualitative study in its essence. To mitigate this, we integrated and analyzed data from different sources (i.e., literature and interviews). We also worked on mitigating the researcher bias by jointly building the data collection and analysis protocols and jointly conducting, reviewing, and discussing the coding effort related to the data analysis phase, and conducting the synthesis phase of our research methodology. To mitigate participant bias, we made sure not to share the interview questions previously with the participants and not interview direct colleagues, reducing the chance for the participants to change their answers or behavior to favor the researcher.

Against the acknowledgement of and countermeasures to mitigate these limitations, and in the context of progressing process mining insights into process improvement, we are confident that our results appropriately reflect the realm of recommended and performed actions and, especially, the causes behind certain recommended actions not resulting in performed actions, leading to process mining project termination.

5.6 Conclusion

In this work, we investigated to what extent recommended actions are also performed, and the causes for certain recommended actions not resulting in performed actions. We identified five causes that can potentially lead to the termination of a process mining initiative because recommended actions were not performed, and these causes are related to three main dimensions: the data, the process mining project itself, and the acquired process mining insights. With this understanding, we contribute to the body of knowledge regarding progressing process mining insights into process improvement.

With the understanding of the causes hampering progress to process improvement from process mining insights presented in this work, we aim to provide a basis for augmenting existing process mining methodologies. Therefore, we plan to incorporate our findings into a comprehensive proposal for enhancing process mining methodologies. We highlight the fact that process mining can bring value to organizations, but not without any further effort. Awareness of the causes we unveil with this work can be critical to moving beyond process mining insights.

In future work, we aim to steer our focus towards the process improvement initiatives in more detail. Specifically, we intend to investigate the link between process improvement initiatives that are triggered by tools and techniques other than process mining and the relations between process improvement initiatives that are triggered by insights from process mining and by insights from other methods. Our goal is to develop a broader understanding of the factors that drive process improvement initiatives in order to enhance the effectiveness of process mining-driven process improvement initiatives.

Chapter 6

Challenges and Recommendations

Abstract

Many organizations have adopted process mining to analyze their business processes, gain insights into their performance, and identify improvement opportunities. Several academic case studies and reports from practice leave no doubt that process mining tools can deliver substantial value to organizations and help them to realize improvements. However, both organizations and academics have also realized that the path from obtaining insights via process mining to realizing the desired improvements is far from trivial. Existing process mining methodologies pay little to no attention to this matter and mainly focus on how to obtain insights through process mining. In this paper, we address this research gap by conducting a qualitative study based on 17 semi-structured interviews. We identify seven challenges pertaining to translating process mining insights into process improvements. Furthermore, we provide five specific recommendations for practitioners and stakeholders that should be considered before starting a new process mining initiative. By doing so, we aim to close the gap between insights and action and help organizations to effectively use process mining to realize process improvements.

Published in V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, and H. A. Reijers. Progressing from process mining insights to process improvement: Challenges and recommendations. In International Conference on Enterprise Design, Operations and Computing (EDOC), 2023 [121]

6.1 Introduction

Over the last years, many organizations have adopted process mining to analyze their business processes, gain insights into their performance, and identify improvement opportunities [51, 107]. Countless academic case studies [107] and reports from practice [51] leave no doubt that process mining tools can deliver substantial value to organizations and help them to realize improvements with respect to relevant performance indicators, such as throughput time [143], conformance [59], or customer satisfaction [31]. However, both organizations and academics have also realized that the path from obtaining insights via process mining to realizing the desired improvements is far from trivial [38]. In fact, moving beyond diagnostics has been identified as one of the current key challenges of process mining [136]. Existing process mining methodologies, such as Process Diagnostics [16], L* [133], or PM^2 [138], pay little attention to this matter and mainly focus on how to obtain insights through process mining. Some recent papers have contributed to the discourse by investigating how process mining insights can trigger automated actions [8, 96, 97]. They, however, take a rather technical perspective and do not consider organizational concerns or challenges.

In this paper, we address this research gap and set out to understand the challenges that arise on the path from translating process mining insights into process improvements. Specifically, we aim to answer the following research question: *"Which challenges do organizations face when translating process mining insights into process improvements?"*. To answer this question, we conducted a qualitative study based on semi-structured interviews with 17 process mining experts. In this way, we were able to detect seven challenges that organizations have to overcome in this context. Based on the identified challenges, we further derive five specific recommendations that can help organizations making a successful transition from process mining insights to process improvements. With the detected challenges and recommendations, we contribute to the stream of process mining literature that is concerned with process mining methodologies [16, 133, 138]. Specifically, we extend their scope by providing guidance for the final step in a process mining project.

The rest of the paper is structured as follows. Section 6.2 introduces the background and the research gap. Section 6.3 describes our research method.

Section 6.4 presents the identified challenges of translating process mining insights into process improvements. Section 6.5 reflects on our findings and provides the recommendations we derived. Finally, Section 6.6 concludes the paper.

6.2 Background

In this section, we discuss the background of our research. Our objective is to demonstrate to what extent existing research focuses on the translation of process mining insights into process improvements. To this end, we first review existing process mining methodologies. Then, we reflect on how process mining insights have been used across different studies.

6.2.1 Process Mining Methodologies

The effective use of process mining for process improvement is often a complex endeavor that goes way beyond the use of process mining software [38]. Process mining methodologies, therefore, aim to provide a reference structure for the application of process mining by defining a number of specific steps. Among others, those steps include scope definition, data collection, the application of process mining techniques such as discovery and conformance checking, result analysis, and process improvement [38]. Several such process mining methodologies have been defined in the past, the most prominent being the process diagnostics methodology (PDM) [16], L* [133], and PM^2 [138].

While these methodologies differ with respect to several details, they have two main things in common. First, they propose a similar high-level flow involving steps such as data collection, application of process mining techniques, and result analysis. Second, they only pay little attention to how process mining insights can be translated into process improvements. At the same time, however, they acknowledge that this step is important. The authors of PDM highlight that the interpretation of the insights identified through their methodology is critical but lies in the responsibility of the organization [16]. The authors of L* explain that their methodology can lead to four different improvement actions: redesigning, adjusting, intervening, and supporting. Yet, they only discuss a few examples of what each action entails and do not reflect on how those actions can be implemented [133]. Also, the authors of PM^2 explicitly acknowledge the importance of process improvement based on the obtained insights by including a step called *process improvement and support*. They, however, argue that the realization of such improvements is typically done in the context of a separate project [138].

This lack of attention with respect to the translation of insights into improvements is also discussed in a relatively recent meta study of process mining case studies by Emamjome et al. [38]. They point out that the last phase of process mining projects is only superficially considered in the analyzed studies and, hence, has a low degree of "thoroughness". They conclude that most case studies they analyzed fit somewhere between the following two categories: 1) "the studies provide insights without any recommendation", and 2) "the studies provide some recommendations on how to improve the process(es), but do not refer to any implementation".

To understand how process mining insights are actually used in real-life cases, we review respective literature in the next section.

6.2.2 Use of Process Mining Insights

The value an organization can realize through process mining highly depends on what the organization does with the obtained insights. Recognizing this, many researchers investigated how process mining insights are used or can be used. In general, we can distinguish three main categories for the use of process mining insights: 1) supporting process understanding and documentation, 2) improving the investigated process, and 3) improving information system(s) supporting the investigated process. Below, we briefly elaborate on each category.

Using process mining insights to support *process understanding and documentation* relates to the explorative use of process mining. Simply put, process mining can help organizations to understand what is going on inside their organization. Besides the discovery of the control flow [7], i.e., the order of activities, process mining can also help to uncover how resources interact [3] or to identify business rules [17, 59]. Some authors highlight the importance of writing [99] and presenting [103] reports based on the acquired process mining insights, yielding documentation creation, reviewing, or updating.

For a more comprehensive overview, we refer the interested reader to the literature study from Garcia et al. [35].

In line with the main objective of process mining, many authors aim to use process mining insights to *improve the investigated process* by generating respective recommendations. Such recommendations can be generic and refer to process change or, simply, redesign [50, 85]. Some, however, are more specific and include preventing a specific activity from happening [31], eliminating an activity [2], or increasing the frequency of a specific activity [17]. Works focusing on the resource perspective suggest actions such as adding resources [24, 26] or increasing resource involvement [124].

As the execution of many processes is supported by one or more information systems, process mining insights can also reveal how to *improve those information systems* in different ways. Some authors discuss rather general aspects such as improving the information system's usability [71, 112, 129]. Other studies report on redefining [126] and adjusting [76] specific feature settings to be more permissive or restrictive based on thresholds identified through process mining. There are also studies reporting on testing new information system features [56] or identifying opportunities for implementing automation [42, 106].

The brief review above illustrates that process mining insights can provide valuable input for both understanding and improving processes and the associated information systems. However, what is currently missing is a clear path towards implementation. As an example, consider a scenario where process mining insights are used to recommend the introduction of an additional quality check in a process. While this recommendation is useful, especially because it is based on a data-driven analysis of the underlying process, putting this recommendation into action is far from trivial. Among others, this requires commitment from both the process manager and the process participants, proper communication of the changes, an allocation of the required resources, additional training, etc. While several authors discuss the importance of communication [78, 137, 146] and also training in such contexts [143], these concerns are generally only superficially considered. As a result, it remains unclear which challenges need to be overcome to translate insights (or recommendations based on insights) into process improvements.

With this paper, we aim to close this gap by identifying and understanding the challenges that occur in this context. In the next section, we explain the methodology of our study.

6.3 Research Method

To identify and understand the challenges that need to be overcome to translate process mining insights into process improvements, we interviewed 17 experts with several years of industrial experience in process mining projects. Below, we describe our research method. Specifically, we elaborate on the definition of the target population and the interview protocol, the data collection, and the data analysis.

6.3.1 Definition of target population and interview protocol

Driven by our research question, our target population included process analysts, business analysts, and researchers with experience in process mining projects in industry. We defined our semi-structured interview protocol consisting of a set of predefined open-ended questions inspired by [15, 19]. The intention was to understand the interviewees' experiences and perspectives related to what happens with process mining insights after they have been acquired. We conducted a test run of our interview protocol with two participants that are not part of this research. With this test run, we verified that the predefined questions were well suited to obtain the desired insights.

6.3.2 Data collection

We sent personal invitations to potential participants via e-mail and LinkedIn. In total, we interviewed 17 process mining experts. Table 6.1 provides an overview of the interviewees. It shows the interviewees' job title (where PM stands for *Process Mining*), experience with process mining in years (cf. column *Exp.*), as well as the continent, size and domain (where IT stands for *Information Technology*) of the organization they work for. The interviewees have an average of seven years of industrial experience with process mining. Eleven of them also obtained a PhD in the process mining field and, therefore, also had additional exposure to the subject. The interviewees used a large variety of process mining tools including ARIS Process Mining, Celonis, Fluxicon Disco, Minit, PAFnow, ProM, UpFlux, UiPath Process Mining, and SAP Signavio.

The interviews were conducted as follows. First, we asked the participants a couple of questions about themselves such as *"What is your role with respect to process mining in your organization?"* and *"For how long have you been working with process mining?"*. Next, we asked general questions about the process mining projects, such as *"What usually triggers the use of process mining in your organization?"* and *"What are usually the expected insights from the stakeholders of a process mining initiative?"*. Then, we asked them to share some details about process mining projects they have been involved with and to talk about the process that was under investigation, the effort that was required to acquire the insights, and which main insights were obtained. Finally, we asked: *"What happened to the process mining insights after they have been acquired?"*. On average, the interviews lasted 54 minutes.

Table 6.1: Interviewees' demographics

Ref.	**Job title**	**Exp.**	**Organization**		
			Continent	**Size**	**Domain**
I1	Business Analyst	5-10	Asia	201-500	Oil and Gas
I2	PM Consultant	10-15	Europe	51-200	IT
I3	Transformation Consultant	5-10	Europe	1k-5k	IT
I4	PM Consultant	5-10	Europe	1k-5k	IT
I5	PM Product Owner	10-15	Europe	1k-5k	Finance
I6	Researcher / PM Consultant	10-15	Europe	1k-5k	Education
I7	Researcher / PM Consultant	15-20	Europe	1k-5k	Education
I8	PM Specialist	5-10	Europe	>10k	Public
I9	Senior Manager	10-15	Europe	>10k	Finance
I10	PM Specialist	10-15	Europe	>10k	Audit
I11	PM Specialist	10-15	Europe	>10k	Public
I12	PM Product Owner	15-20	Europe	>10k	Healthcare
I13	PM Product Owner	5-10	North America	1k-5k	IT
I14	PM Consultant	5-10	North America	>10k	IT
I15	PM Analyst	1-5	Oceania	1k-5k	Food
I16	Researcher / PM Consultant	10-15	Oceania	1k-5k	Education
I17	PM Product Owner	5-10	South America	51-200	IT

6.3.3 Data analysis

Each interview was audio-recorded and transcribed. Then, we anonymized the transcriptions by removing any information that could reveal the interviewees' identity or the organization they worked for. We conducted a qualitative coding using four main steps [19, 113]. First, we familiarized ourselves with the interviews by reading them and taking general notes. Second, we re-read the interviews and wrote memos. For example, when an interviewee talked about the customer expectations being much different from the process mining outcomes such that they decided to discontinue the project, we added memo notes such as "expectation" and "project ends". Third, we reviewed our codes to identify possible connections among the codes or the possibility of merging multiple codes into higher-level categories. Finally, we identified multiple categories concerning challenges relating to translating process mining insights into process improvements.

6.4 Findings

In this section, we present the findings of our study. In total, we identified seven specific challenges that can impair an organization's ability to translate process mining insights into process improvements. We classified these seven challenges into three main categories: 1) organizational commitment, 2) expertise, and 3) expectations. In the subsequent sections, we elaborate on each category in detail and illustrate the respective challenges by using quotes from our interviews. An overview of the three main categories and the seven challenges, as well as the number of supporting interviewees for each category, is depicted in Figure 6.1.

6.4.1 Organizational Commitment

The fact that change requires organizational commitment has been emphasized in BPM literature for a long time [11, 54, 130]. However, our interviews revealed that this awareness is often limited when it comes to the application of process mining. We identified two specific challenges in this context: lack of top-level management support and change resistance.

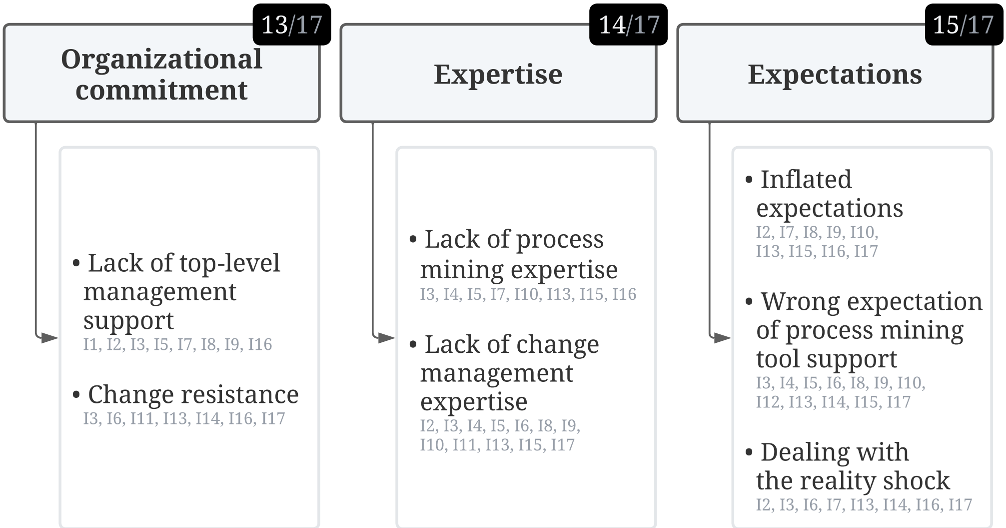


Figure 6.1: Overview of the identified challenges

Lack of top-level management support. Several interviewees highlighted that, without support from top-level management, process mining projects do not yield much besides acquiring insights about the analyzed processes. For instance, Interviewee I2 pointed out that "*in order to have changes, you do need some top-level support because you need a budget. Another reason why you need top-level support is for them to be able to say that now it's part of the vision and we should spend time with it, it is part of the initiatives to actually improve this*". This point is also supported by interviewee I16, who reported on a project where process mining insights were successfully translated into process improvements: "*(...) but you need to understand that this is the head of the organization. He wants to improve the process. If he wasn't the CEO, if he was a developer, for example, I do not think he would be able to make this change.*"

Interviewee I3 also emphasized the importance of involving a manager or director (i.e., non c-level managers) who can understand and deliver the insights to their team because "*then you get into the normal psychology of human change by having an ambassador, a leader who says we need to change, and supports the work required to change*". Similarly, I8 stated that "*the organization started to check and act upon process mining insights because there was a new program manager that really believed in process mining*".

Change resistance. Resistance to change is a well-known and well-studied phenomenon [109]. In our interviews, we encountered both individual as well as organizational resistance to change. As for individual resistance, the interviewees mentioned instances of resistance that can be related to *habit* and *fear of the unknown*. *Habit* describes the problem of people resisting change because they need to alter the way they work. Resistance due to a *fear of the unknown* is more abstract and can be attributed to the uncertainty that individuals experience when changes are introduced. As for organizational resistance, interviewees mentioned instances of *structural inertia* and *threat to established power relationships*. *Structural inertia* refers to changes that interfere with the organizations' mechanisms built to produce stability in the work processes. The *threat to established power relationships* occurs when these power relationships are at risk because of the redistribution of the responsibility for decision-making.

Interviewee I6 described a case that relates both to *habit* and *structural inertia*: "*... it is important to grow confidence on process mining with smaller suggestions for improvement first, and really think through which kind of recommendations of improvements to make, because asking someone to change the way that he or she works might not be the smartest way of going for it*". Interviewee I6 also shared an interesting reflection on how they handled the anticipated resistance to change: "*the stakeholders already know that I am not going to try to replace anyone or any decision, I will try only to support or try to provide information or ways for them to do their work just as they were doing before but with very small changes. Just then is when people start accepting the suggestions to change*". This example highlights how important it is to involve someone in which the affected people in the organization trust. In this particular case, the interviewee has been responsible for different process mining projects in the organization for almost four years.

Interviewees I13 and I14 also described examples of *fear of the unknown*. I13 mentioned that "*... some people get 'cold feet' about going forward with process mining projects because people will demand responses from them later and they are just afraid to take the responsibility*". Similarly, I14 pointed out that "*people are too scared of having to change*". They both also mentioned cases of people impeding the process mining project by hiding information. Interviewee I13 stated that "*sometimes we could show people up in the value chain, directors, the potential value of process mining, and they would sponsor our conversations with the operations team* *who, however, didn't want their directors to know everything that was going on within operations. They didn't want to be monitored. Then, they told their directors that they did not have all the data or did not find anything meaningful*". Interviewee I14 also shared a case where information was hidden by managers, stating that "*managers know they can do better, but they also know that their bosses do not know they can do better. So they can play 'life easy', and using process mining would take this advantage away from them*".

6.4.2 Expertise

Realizing process improvements through process mining requires the organization to have certain expertise at its disposal. On the one hand, it is critical that the process mining insights can be properly understood and interpreted, i.e., there is a need for process mining expertise. On the other hand, identified weaknesses must also lead to effective changes in the organization's processes, i.e., there is also an immanent need for change management expertise.

Lack of process mining expertise. Several interviewees pointed out that, according to their experience, the output of process mining tools can hardly be properly interpreted without an employee who is capable of understanding both process mining as well as the domain. Interviewee I4 mentioned that in one of the organizations they worked on, the organization had "*purchased the license, and they were supposed to use the tool themselves, but they weren't able to generate any findings or insights*". While the interviewee, having several years of process mining experience, was not specifically hired for that project, they had to step in to prevent the process mining project from being canceled.

Interviewees I5 referred to "*the need for a process mining expert working in the project, especially one that can also learn or previously know about the domain*". Similarly, interviewee I16 stated that "*process mining is a good tool for communication within the team if they are interested from the beginning, but process mining needs a good process analyst and involvement of a domain expert*". Also, interviewee I7 mentioned that "*you need a process mining expert who can translate the event log data into insights to the organization*", and interviewee I13 mentioned that the "*big blockers to buying and using process mining are that companies over and over again say that they do not have the people to analyze what process mining is showing them*".

Lack of change management expertise. Several interviewees highlighted that there are different cases that demand for an (impartial) change management expert. In some cases, there is a lack of technical expertise and no commitment from the stakeholders to work on the changes. For example, interviewee I4 mentioned that the proposed changes based on process mining insights were never implemented because *"it tends to be complicated making changes and running two configurations in the same live data. We did not have the expertise to make these changes, and the stakeholders didn't want to commit with their own resources to do it"*.

In other cases, as reported by interviewee I11, there is a lack of financial support and of a manager with company-wide access. Interviewee I11 mentioned that the impact they could make was *"initially small to nonexistent, because they needed a strong manager to bring widespread process mining initiatives in the organization, make these initiatives continuous and more effective, but this manager was not there"*. Interviewee I11 also mentioned that when there was a *"strong manager"*, related to the financial department of the company, he was capable of implementing a company-wise widespread process mining initiative. According to the interviewee, this manager had broad access to different departments and financial support.

Finally, as raised by interviewee I17, there should be an impartial change manager to deal with political-related aspects that are harming the company: *"... [we] should put someone capable to do the change management in the company to recover the lost money caused by inefficient employees"*. This, however, did not happen because, in this case, *"a very well related person, that does not follow good practices can be protected by their peers. We can detect such behaviors, but nothing happens, and the project ends"*.

6.4.3 Expectations

We found that the application of process mining is often associated with high expectations and partially also with misconceptions. We identified three main challenges in this context: inflated expectations, wrong expectation of process mining tool support, and dealing with the reality shock.

Inflated expectations. Several interviewees highlighted the importance of being aware of the effort required to translate process mining insights into improvements, and not expect that process mining will magically improve the process.

Interviewee I13 shared that "*people have been disappointed with process mining in the past, but mostly because either they had inflated expectations or they under-estimated the work that needs to go into turning insights into something useful*". According to interviewee I17, "*the most successful projects have a good alignment between expectations and insights*". The interviewee mentioned that they drive this alignment based on previous experience and previously defined templates building on expected and acquired insights.

Interviewee I2 suggested to handle inflated expectations by starting the process mining project small: "*oftentimes it is difficult to turn process mining insights into value, because if you find something, then you might not know, for example, the person whose responsibility this is to pick it up. It also might be a not known pain point, which would lead you to first needing to convince people that actually what you found is true. So, my approach is to typically start small, not with the biggest money maker process, to start gaining some trust in the solution and start with problems that people already know and about which they might already have some hypothesis*". According to interviewee I2, starting small also makes it easier for the company to acquire experience using process mining and understand how fast the company is in implementing, analyzing, and getting value out of the process mining project. Similarly, to narrow down stakeholder's expectations, interviewee I9 mentioned that "*before starting any project we always sit together with the client, ask them about their priorities, and also whether there is any specific challenge that they would like us to focus on, or that they would like more insights about or more recommendations about*".

Wrong expectation of process mining tool support. This challenge relates to the problem that stakeholders still see process mining as a "*full-fledged process improver setup*", which is not the case. Therefore, process mining methodologies and advocates should consider including change management initiatives as one of its stages, or at least provide initial guidance regarding the effort required to move process mining insights into action.

Interviewee I4 mentioned that "*process mining should not be the only tool or artifact for process improvement. It should be aligned with other tools and initiatives for that. Data itself is not enough to really understand the underlying problem.*

With data and process mining we can, most of the time, describe the problem well, but we can't really say how to improve it; there needs to be some sort of process understanding that then is used to finally improve the process". Interviewee I12 highlighted that "*process mining is a tool to support process redesign initiatives in the organization. So, an advice for making process mining more usable, the process should be analyzed and then there should be a second phase to work upon improvements based on what we saw*".

Also interviewee I15 highlighted that "*the limitation of the tool compared to the expectations of the stakeholder is a challenge. Process mining requires a few stages to actually bring value to the customers: we need to build the data model, then do the analysis, then, based on the insights, think of how to turn insights into action. And the action part is the challenging part for business. Turning insights into action is certainly a pain point for most businesses. Turning insights into action involves different departments; it involves how the business operated before and how they are going to operate in the future, and the most challenging part is that it involves multiple departments, and it really depends on how the senior managers are going to do. It really depends on how you manage your company*".

Interviewee I13 shared a situation that they went through when after they showed their process mining tool to a friend that was working at a big tech company, this friend asked them: "*are you telling me that I should pay you money for you to show me my problems and not solve them? Really?*". According to I13, "*there is a need for expectation management and, of course, this inflated-expectations is not a problem exclusive to process mining. And as long as process mining is not something that is well understood by the market, inflated expectations will always be there*".

Dealing with the reality shock. Process mining is a "*big mouth*" and it will uncover "*hard truths to swallow*". While some interviewees mention that "*it is easier for managers trying to use process mining to say that it doesn't work than to accept the insights it can deliver*" (I13) and "*process mining is too truthful*" (I14), interviewee I3 suggests a mean to deal with the reality shock: "*you need to involve your customer because then they evolve in the way of thinking at the same rate as you. If you don't do that and you simply take the data, go back to your cave and start analyzing it, you come back conceptually and mentally three steps ahead of them, and if you then just drop it on them, they could be very defensive because you* *basically tell them their process is a mess, and that's very often what it is. So, you need to take them along on the journey."*.

The reality shock can occur for the organization conducting a process mining project to understand and improve their own processes, for the process participants, and it can also be for service providers or process analysts. An example of a reality shock for the organization, interviewee I6 shared a case where the nurses of a hospital were highly stressed with their work. The managers of the hospital did not understand how that could be, considering how much idle time the nurses had, based on usual process discovery-related insights acquired. The interviewee decided to look more closely at the daily work of the nurses by conducting observation sessions. They learned that the idle time was just a reflection of limited data availability related to their daily work process. The managers learned the hard way that process mining can only show what is included in the event log. All the times, the nurses needed to hurry to a patient's room to attend to a patient's call had not been recorded in any information system.

An example of a reality shock for the process participant, also shared by interviewee I6, related to long waiting times for an emergency room. At first, the physicians were not enthusiastic about the process mining project that was started by the management team. The interviewee learned that one of the reasons for this long wait was that physicians switch context too often. In other words, the doctor has a certain amount of patients in the waiting room; one has an orthopedic problem, the other one has a cardiac problem, the other a neurological problem, etc. The physicians did not notice, but they were taking too long to think of the different special reasons related to different patient needs. The interviewee suggested them to group patients per type of complaint and analyze each group together. They applied the suggestion to one department and could see the waiting time of all patients reduced by 20%. Thereafter, the physicians started to accept the technique.

As an example of a reality shock for the process mining service provider, interviewee I13 shared an example related to a credit card sales process. In essence, the process was concerned with selling a credit card to clients in a physical store. Part of the selling process was a credit analysis to check the customer's credit status. The analysis revealed that every time a human was involved in the credit analysis, it took double the time to close the sale, and the likelihood of a successful sale decreased. After analyzing the credit analyst's actual work, the interviewee noticed they were very fast. They further investigated this inefficiency and learned that whenever a manual credit analysis was triggered, the client in the store was said to wait and would walk around the store and eventually simply leave. The interviewee suggested a very simple solution to this problem (e.g., offer coffee to the client or talk to them for a while), but *"once the manager of the credit checking group realized that it wasn't the credit analysis that was delaying the process, it was not his fault and he didn't care about making any changes anymore and they did not continue using process mining after that"*.

6.5 Recommendations

The findings from our interviews reveal that translating process mining insights into process improvements comes with substantial challenges. Our interviews also highlight that it is likely that the transition from insights to improvements is never made if these challenges remain unaddressed. It is not particularly surprising that several of the challenges we identified relate to phenomena that have been made in the context of change management, such as resistance to change [75]. Yet, process mining projects, and hence also the associated challenges, differ from traditional change management projects, digital transformation projects, and process redesign initiatives. Most importantly, in process mining projects, the insights that provide the starting point and argument for changes are acquired through software. Naturally, this does not only changes the nature of change resistance but also calls for specific expertise for interpreting results and implementing changes. As existing process mining methodologies have paid little to no attention to these aspects [38], we derived five recommendations that organizations should consider when starting a process mining initiative. The recommendations provide specific input on how process mining projects should be prepared, set up, conducted and who should be involved. Specifically, our derived recommendations for process mining projects in practice are the following:

R1 - Engage top-level management support: Top-level management support should be secured before the start of the process mining initiative. It is essential for getting appropriate financial support, conveying the importance of the initiative, and ensuring the ability to actually implement the required changes.

R2 - Be ready to face resistance to change: Resistance to change must be expected in every process mining initiative and should be handled appropriately. We found that it is particularly about communication. If people understand which changes will be implemented and why, they are much more likely to support their implementation. Handling fears and concerns, therefore, is a critical activity.

R3 - Have process mining and domain expertise at your disposal: One of the critical steps in every process mining initiative is the interpretation of the acquired results. This requires an individual who is familiar with both process mining and the respective domain of the organization. Such a person should be either hired or educated on time.

R4 - Have change management competence at your disposal: Translating process mining insights into process improvements requires change. Hence, it is essential to have change management expertise available in the organization. Such a change manager will follow up on the recommendations of the process analyst (see R3) and develop a strategy on how to successfully implement the desired changes.

R5 - Manage expectations: Expectations among several stakeholders of a process mining initiative are often unrealistic. Therefore, it is important to manage expectations with respect to the outcome and also the effort that will be required to realize process improvements through process mining. People need to be aware that process mining is a tool and will not magically improve processes without any effort.

The recommendations stem from the identified challenges (cf., Section 6.4). For convenience, Table 6.2 presents which recommendations address which challenges.

Table 6.2: Recommendations to challenges mapping

R1	R2	R3	R4	R5	Challenge
•					Lack of top-level management support
	•				Change resistance
		•			Lack of process mining expertise
			•		Lack of change management expertise
				•	Inflated expectations
				•	Wrong expectation of process mining tool support
				•	Dealing with the reality shock

6.6 Conclusion

In this paper, we investigated which challenges organizations face when translating process mining insights into process improvements. To this end, we conducted a qualitative study involving 17 interviews with process mining experts. Based on these interviews, we identified seven challenges, which we turned into five specific recommendations that organizations using process mining should consider. Among others, we highlighted the importance of top-level management support and the availability of expertise with respect to process mining, the domain, and change management. After all, turning process mining insights into improvement requires change and, therefore, also a respective commitment from several levels of the organization.

Naturally, our study is subject to limitations. Most importantly, our study is qualitative and, hence, limited in terms of generalizability. We, however, attempted to mitigate this concern by involving process mining experts that worked in different organizations and settings, have used different process mining tools and approaches, and faced different problems in their organizations. Other biases, e.g. with respect to the analysis, we mitigated by jointly building the data collection protocol, and jointly conducting, reviewing, and discussing the coding effort related to the data analysis. Therefore, we are confident that our results appropriately reflect the challenges organizations face, and provide valuable input about how process mining insights can be translated into process improvements.

In future work, we aim to validate our findings in the context of a large case study. Furthermore, we plan to incorporate our findings into a comprehensive proposal for a process mining methodology.

Chapter 7 Conclusion

In this final chapter, we outline the main contributions of this thesis, reflect on limitations and opportunities, and conclude the chapter with potential future work directions.

7.1 Contributions and Implications

In this thesis, we focused on the initial and final stages of a three-stage process mining pipeline. This pipeline starts with the extraction of event logs, progresses through process discovery and analysis, and concludes with the translation of process mining insights into process improvement. Next, we reflect on our contributions, which extend the knowledge base of the process mining field in two aspects:

- the event log extraction, in Section 7.1.1; and
- the translation of process mining insights into process improvement, in Section 7.1.2.

Figure 7.1 depicts how our contributions can be positioned as enhancement opportunities for process mining methodologies. We use PM^2 methodology [138] for exemplification.

7.1.1 Part α: From Data to Event Log

Towards Understanding the Role of the Human. In Chapter 2, we presented a taxonomy of manual tasks in event log extraction, answering RQ1 *"What are the specific manual tasks that humans perform in the context of event log extraction?"*.

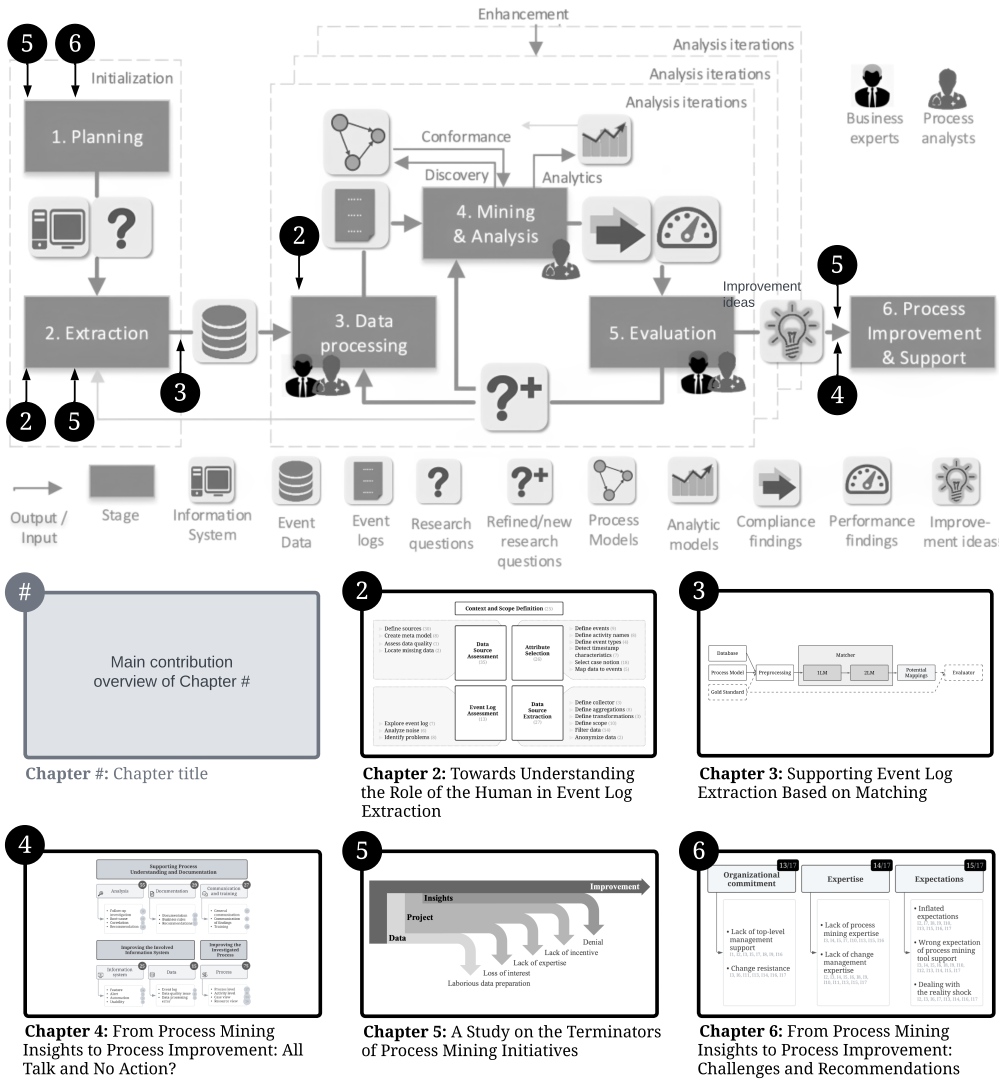


Figure 7.1: Overview of how our findings can be integrated into an existing process mining methodology. The numbers represent the Chapter where you can learn more about the contribution.

We used a mix of literature review and qualitative data coding to set structure and understanding to the diverse space of manual tasks.

With this work we showed how human involvement and expertise play an important role in event log extraction. This taxonomy enables a birds-eye view, which can provide a basis for future research concerning (semi-)automation of the different manual tasks that are still involved in the event log extraction. The provided taxonomy also serves as a basis for novice process analysts to learn about the different tasks they may need to perform while extracting an event log for process mining purposes.

Supporting Event Log Extraction Based on Matching. In Chapter 3, we presented an architecture for identifying potential mappings between a process model and a database. This answered RQ2 "*How to link a reference model and the underlying database to support the event log extraction?*". Based on our taxonomy of manual tasks in event log extraction, we zoomed in on each manual task and explored their potential for (semi-)automation. We decided to focus on the task related to mapping event data from relational databases to process model activities. Without such a mapping, the location of the relevant event data to compose the event log is unknown. Also, the automation of this particular mapping task is potentially valuable for practical process mining consultancy settings, as it is time-consuming to perform manually.

This contribution can inspire other research initiatives on (semi-) automation of each manual task composing our taxonomy of manual tasks in event log extraction. In addition, this work can serve as inspiration for the development of new tools and methodologies towards automated event log extraction. These can potentially reduce the time and expertise required for extracting an event log. As a result, they could make process mining more accessible to small and medium-sized organizations, reducing the burden for the application of process discovery and analysis.

7.1.2 Part ω: From Insights to Improvement

All Talk and No Action? In Chapter 4, we structured the intervention space of actions triggered by process mining insights for process improvement. Also, we offered an in-depth view into the variety of actions. This answered RQ3 "*What are the actions organizations can take towards process improvement?*". To accomplish this, we conducted a systematic literature review.

By understanding the diversity of actions triggered by process mining insights, we provide a basis for enhancing the effectiveness of the implementation phase of existing process mining methodologies. Also, our findings have implications for research on the development of actionable tools and frameworks. In addition, the derived overview of actions triggered by process mining insights can serve as a catalog for practitioners that aim to translate process mining insights into actual process improvement. Such a catalog can be particularly useful for novice process mining consultants, and managers supporting them, in outlining their initial plans of action for their process mining projects.

A Study on the Terminators of Process Mining Initiatives. In Chapter 5, we presented a view on recommended and performed actions, as well as a first set of causes why process mining initiatives do not materialize into process improvements. We refer to these causes as terminators. These answered RQ4 "*To what extent are recommended actions also performed?*". We leveraged a mix of systematic literature review, semi-structured interviews, and content analysis to derive an understanding of the extent to which recommended actions are actually performed, and what the specific causes are of process mining initiatives being halted.

With this understanding of terminators of process mining initiatives, we provide a basis for further research into augmenting existing process mining methodologies. Techniques for detecting and anticipating specific terminators can be studied and developed. Also, case studies can be conducted to explore whether specific terminators occur in specific organizational settings, allowing for targeted solutions toward detecting and anticipating the terminators. In addition, practitioners can benefit from previous knowledge of the existence of specific causes why recommended actions are not also performed towards process improvement. This awareness may serve as a basis for practitioners to prepare risk management plans for their process mining initiatives, aiming toward more effective process mining support for their organizations.

Challenges and Recommendations. In Chapter 6, we unveiled a set of challenges faced by practitioners when progressing process mining insights into process improvements. This answered RQ5 "*Which challenges do organizations face when translating process mining insights into process improvements?*". Based on these challenges, we proposed a set of recommendations that researchers and practitioners should keep in mind when starting a process mining initiative. To identify these challenges, we leveraged a set of semi-structured interviews with experienced process mining practitioners, which the data we analyzed by performing content analysis.

With the identified challenges and recommendations, we provide a basis for future research to develop an in-depth understanding of each challenge and provide prescriptive recommendations to tackle them. In addition, with the awareness about these challenges, practitioners can pro-actively work toward preventing their occurrence during a process mining initiative. They can, for example, work on managing expectations from the beginning and making sure to have the support of top-level management.

7.2 Limitations and Opportunities

In this section, we discuss the limitations of the work described in this thesis. In Chapter 2, the generated taxonomy was derived from literature. Although we focused on reported case studies, observation sessions and further exploration of how practitioners conduct event log extraction in real-life scenarios could benefit the taxonomy.

In Chapter 3, we proposed an architecture to support the (semi-) automated mapping of database tables and process model activity labels. Although we used real-life databases and process models from the Business Process Management Academic Initiative repository [145], these databases and process models had meaningful labelling of their entities, which may not be the case in all organizational contexts, potentially limiting the practical applicability of this work.

In Chapter 4, we provided an understanding of the realm of actions triggered by process mining insights. Our findings are derived from an extensive literature study focusing on case studies. Nonetheless, our work could benefit from case studies considering organizations from different domains to have a broader view of the actions space, also allowing us to understand which are the differences and commonalities across different domains in terms of the actions triggered by process mining insights.

In Chapter 5, we identified different causes for process mining initiatives not progressing from process mining insights into process improvement. Although we relied both on an extensive literature review and on semi-structured interviews with experienced process mining practitioners, case studies within organizations from different domains could benefit our findings.

In Chapter 6, we unveiled challenges hampering the translation of process mining insights into process improvements. To do so, we leveraged a set of semi-structured interviews with experienced process mining practitioners. This work could have benefited from interviews with top-level management and ground-floor workers of the organizations using process mining. Regarding the derived recommendations, there is an opportunity for research on more prescriptive recommendations to progress process mining insights into process improvement.

7.3 Future Work

In each chapter of this thesis we presented future work ideas. In this section, we reflect on further potential future works within the context of this thesis.

A clear direction is to devise a detailed, enhanced process mining methodology to be tested in a real-life project. It would be interesting to do so in an organization that has previously conducted an unsuccessful process mining initiative, if we could attempt to conduct a similar setup of such an initiative (with the same questions and data) with the new methodology. Other potential future research directions are related to complementing the empirical findings with practical data by conducting case studies in organizations from different domains.

In this work we unveiled a variety of challenges hampering the translation of process mining insights into process improvements. These challenges are still primarily related to human factors, e.g., resistance to change, unrealistic expectations, lack of engagement, among others. Although research fields such as change management and work psychology study human behavior in organizations [75, 100], it is clear that there is still space for research on integrating this knowledge into organizational structure and staff. Ultimately, this integration can enable organizations to leverage process mining effectively, i.e., acquiring insights and returning them to the organization through actual process improvement. Thus, a potential future research direction is to study the interplay between human work and psychology by exploring personality traits [12, 55] of workers that are directly or indirectly related to process mining initiatives to identify potential "trouble maker" workers, anticipate their potentially risky behavior, and "treat before the need for medicine", i.e., provide training, reflection moments, instigate open-communication, and so on.

Altogether, in this thesis, we identified a variety of manual tasks still involved in event log extraction and unveiled challenges and specific causes that still hamper process mining's effectiveness in supporting organizations. Considering this and acknowledging that research is a never-ending process, we are confident that this work can serve as a basis for various future research endeavours.

Bibliography

[1] S. Agostinelli, F. Covino, G. D'Agnese, C. D. Crea, F. Leotta, and A. Marrella. Supporting governance in healthcare through process mining: a case study. *IEEE Access*, 8:186012–186025, 2020.

[2] Ü. Aksu and H. A. Reijers. How business process benchmarks enable organizations to improve performance. In *IEEE International Enterprise Distributed Object Computing Conference*, pages 197–208. IEEE, 2020.

[3] C. Alvarez, E. Rojas, M. Arias, J. Munoz-Gama, M. Sepúlveda, V. Herskovic, and D. Capurro. Discovering role interaction models in the emergency room using process mining. *Journal of Biomedical Informatics*, 78:60–77, 2018.

[4] R. Andrews, M. T. Wynn, K. Vallmuur, A. Hofstede, E. Bosley, M. Elcock, and S. Rashford. Pre-hospital retrieval and transport of road trauma patients in queensland: A process mining analysis. In *Business Process Management Workshops*, pages 199–213. Springer, 2019.

[5] R. Andrews, M. T. Wynn, K. Vallmuur, A. Hofstede, E. Bosley, M. Elcock, and S. Rashford. Leveraging data quality to better prepare for process mining: An approach illustrated through analysing road trauma prehospital retrieval and transport processes in Queensland. *Int. Journal of Environmental Research and Public Health*, 16(7), 2019.

[6] R. Andrews, C. G. J. van Dun, M. T. Wynn, W. Kratsch, M. K. E. Röglinger, and A. Hofstede. Quality-informed semi-automated event log generation for process mining. *Decision Support Systems*, 132(4):113265, 2020.

[7] A. Augusto, R. Conforti, M. Dumas, M. La Rosa, F. M. Maggi, A. Marrella, M. Mecella, and A. Soo. Automated discovery of process models

from event logs: review and benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 31(4):686–705, 2018.

[8] P. Badakhshan, G. Bernhart, J. Geyer-Klingeberg, J. Nakladal, S. Schenk, and T. Vogelgesang. The action engine - turning process insights into action. In *Demo Track at International Conference on Process Mining (ICPM)*, 2019.

[9] P. Badakhshan, B. Wurm, T. Grisold, J. Geyer-Klingeberg, J. Mendling, and J. vom Brocke. Creating business value with process mining. *The Journal of Strategic Information Systems*, 31(4):101745, 2022.

[10] R. B. Bahaweres, H. Amna, and D. Nurnaningsih. Improving purchase to pay process efficiency with RPA using fuzzy miner algorithm in process mining. In *Decision Aid Sciences and Applications (DASA)*. IEEE, 2022.

[11] W. Bandara, M. Indulska, S. Chong, and S. Sadiq. Major issues in business process management: an expert perspective. In *European Conference on Information Systems (ECIS)*, pages 1240–1251, 2007.

[12] M. Barrick and A. Ryan. *Personality and Work: Reconsidering the Role of Personality in Organizations*. Wiley, 2004.

[13] E. Benevento, P. M. Dixit, M. F. Sani, D. Aloini, and W. M. P. van der Aalst. Evaluating the effectiveness of interactive process discovery in healthcare: A case study. In *Business Process Management Workshops*, pages 508–519. Springer, 2019.

[14] M. L. Bernardi, M. Cimitile, C. Di Francescomarino, and F. M. Maggi. Do activity lifecycles affect the validity of a business rule in a business process? *Information Systems*, 62(12):42–59, 2016.

[15] C. Boyce, P. Neale, and P. International. *Conducting in-depth interviews: a guide for designing and conducting in-depth interviews for evaluation input*. Pathfinder International, 2006.

[16] M. Bozkaya, J. Gabriels, and J. M. E. M. van der Werf. Process diagnostics: A method based on process mining. In *Information, Process, and Knowledge Management (eKNOW)*, pages 22–27, 2009.

[17] Z. D. Bozorgi, I. Teinemaa, M. Dumas, M. La Rosa, and A. Polyvyanyy. Process mining meets causal machine learning: discovering causal rules from event logs. In *International Conference on Process Mining (ICPM)*, pages 129–136. IEEE, 2020.

[18] Z. D. Bozorgi, M. Dumas, M. L. Rosa, A. Polyvyanyy, M. Shoush, and I. Teinemaa. Learning when to treat business processes: Prescriptive process monitoring with causal inference and reinforcement learning. In *Advanced Information Systems Engineering (CAiSE)*, pages 364–380. Springer, 2023.

[19] A. Bryman. *Social research methods*. Oxford University Press, 2016.

[20] D. Calvanese, M. Montali, A. Syamsiyah, and W. M. P. van der Aalst. Ontology-driven extraction of event logs from relational databases. In *Business Process Management Workshops*, pages 140–153. Springer, 2016.

[21] D. Calvanese, T. E. Kalayci, M. Montali, and A. Santoso. OBDA for log extraction in process mining. In *Reasoning Web*, LNCS, pages 292–345. Springer, 2017.

[22] D. Calvanese, T. E. Kalayci, M. Montali, and S. Tinella. Ontology-based data access for extracting event logs from legacy data: The onprom tool and methodology. In *Business Information Systems*, LNBIP, pages 220–236. Springer, 2017.

[23] I. P. Cardenas, M. Espinoza, J. Armas-Aguirre, and H. Aguirre-Mayorga. Security of the information model on process mining: case study of the surgery block. In *Congreso Internacional de Innovación y Tendencias en Ingeniería*, pages 1–5. IEEE, 2021.

[24] O. Cela, A. Front, and D. Rieu. CEFOP: A method for the continual evolution of organisational processes. In *Research Challenges in Information Science (RCIS)*, pages 33–43. IEEE, 2017.

[25] M. Cho, M. Song, and S. Yoo. A systematic methodology for outpatient process analysis based on process mining. In *Asia Pacific Business Process Management (AP-BPM)*, volume 181 of *LNBIP*, pages 31–42. Springer, 2014.

[26] M. Cho, M. Song, M. Comuzzi, and S. Yoo. Evaluating the effect of best practices for business process redesign: An evidence-based approach based on process mining techniques. *Decision Support Systems*, 104(12): 92–103, 2017.

[27] J. V. de Camargo, N. M. de Moreira Bohnenberger, V. Stein Dani, J. P. M. de Oliveira, E. Sosa-Sánchez, G. Polančič, and L. H. Thom. A complementary analysis of the behavior of BPMN tools regarding process modeling problems. In *Enterprise, Business-Process and Information Systems Modeling*, pages 43–59. Springer, 2022.

[28] M. de Leoni and W. M. P. van der Aalst. Data-aware process mining: Discovering decisions in processes using alignments. In *ACM SAC*, page 1454–1461, 2013.

[29] E. G. L. de Murillas, W. M. P. van der Aalst, and H. A. Reijers. Process mining on databases: Unearthing historical data from redo logs. In *Business Process Management (BPM)*, volume 9253 of *LNCS*, pages 367–385. Springer, 2016.

[30] E. G. L. de Murillas, H. A. Reijers, and W. M. P. van der Aalst. Connecting databases with process mining: A meta model and toolset. *Software and Systems Modeling*, 18(2):1209–1247, 2019.

[31] M. Dees, M. de Leoni, W. M. P. van der Aalst, and H. A. Reijers. What if process predictions are not followed by good recommendations? In *Business Process Management Industry Forum*, pages 61–72, 2019.

[32] G. Deeva, J. De Smedt, P. De Koninck, and J. De Weerdt. Dropout prediction in moocs: A comparison between process and sequence mining. In *Business Process Management Workshops*, pages 243–255. Springer, 2018.

[33] P. Delias and G.-T. Nguyen. Prototyping a business process improvement plan. an evidence-based approach. *Information Systems*, 101(11):101812, 2021.

[34] K. Diba, K. Batoulis, M. Weidlich, and M. Weske. Extraction, correlation, and abstraction of event data for process mining. *WIREs Data Mining and Knowledge Discovery*, 10(3), 2020.

[35] C. dos Santos Garcia, A. Meincheim, E. R. Faria Junior, M. R. Dallagassa, D. M. V. Sato, D. R. Carvalho, E. A. P. Santos, and E. E. Scalabrin. Process mining techniques and applications – a systematic mapping study. *Expert Systems with Applications*, 133:260–295, 2019.

[36] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers. *Fundamentals of Business Process Management*. Springer, 2018.

[37] J. Eggers, A. Hein, M. Böhm, and H. Krcmar. No longer out of sight, no longer out of mind? How organizations engage with process mining-induced transparency to achieve increased process awareness. *Business & Information Systems Engineering*, 63:491–510, 2021.

[38] F. Emamjome, R. Andrews, and A. H. M. ter Hofstede. A case study lens on process mining in practice. In *On the Move to Meaningful Internet Systems (OTM)*, pages 127–145. Springer, 2019.

[39] E. V. Epure, C. Hug, R. Deneckére, and S. Brinkkemper. What shall I do next? Intention mining for flexible process enactment. In *Advanced Information Systems Engineering (CAiSE)*, volume 8484 of *LNCS*, pages 473–487. Springer, 2014.

[40] M. ER, N. Arsad, H. M. Astuti, R. P. Kusumawardani, and R. A. Utami. Analysis of production planning in a global manufacturing company with process mining. *Journal of Enterprise Information Management*, 31(2): 317–337, 2018.

[41] T. G. Erdogan and A. Tarhan. A goal-driven evaluation method based on process mining for healthcare processes. *Applied Sciences*, page 894, 2018.

[42] B. M. Esiefarienrhe and I. D. Omolewa. Application of process mining to medical billing using L∗ life cycle model. In *International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2021.

[43] K. Figl, J. Mendling, and M. Strembeck. The influence of notational deficiencies on process model comprehension. *Journal of the Association for Information Systems*, 14(6):312–338, 2013.

[44] C. M. Flath and N. Stein. Towards a data science toolbox for industrial analytics applications. *Computers in Industry*, 94:16–25, 2018.

[45] C. Fleig, D. Augenstein, and A. Mädche. Process mining for business process standardization in erp implementation projects – an SAP S/4 HANA case study from manufacturing. In *Business Process Management Workshops*, pages 149–155. Springer, 2018.

[46] A. Gal. *Uncertain Schema Matching*, volume 3. Morgan & Claypool, 2011.

[47] C. d. S. Garcia, A. Meincheim, F. C. G. Filho, E. A. P. Santos, and E. E. Scalabrin. Getting insights to improve business processes with agility: A case study using process mining. In *Systems, Man and Cybernetics (SMC)*, pages 1336–1343. IEEE, 2019.

[48] K. Gerke, K. Petruch, and G. Tamm. Optimization of service delivery through continual process improvement: A case study. In *INFORMATIK – Business Process and Service Science*, pages 94–107. Gesellschaft für Informatik e.V., 2010.

[49] J. Geyer-Klingeberg, J. Nakladal, F. Baldauf, and F. Veit. Process mining and robotic process automation: A perfect match. In *Business Process Management (BPM)*, pages 124–131, 2018.

[50] K. Goel, S. J. J. Leemans, M. T. Wynn, A. H. M. ter Hofstede, and J. Barnes. Improving PhD student journeys: Insights from an australian higher education institution. In *Business Process Management Industry Forum*, pages 27–38. CEUR-WS.org, 2021.

[51] T. Grisold, J. Mendling, M. Otto, and J. vom Brocke. Adoption, use and management of process mining in practice. *Business Process Management Journal*, 27(2):69–387, 2020.

[52] B. R. Gunnarsson, S. K. L. M. vanden Broucke, and J. De Weerdt. Predictive process monitoring in operational logistics: A case study in aviation. In *Business Process Management Workshops*, pages 250–262. Springer, 2019.

[53] M. Gupta, A. Serebrenik, and P. Jalote. Improving software maintenance using process mining and predictive analytics. In *International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2017.

[54] M. Hammer and J. Champy. *Reengineering the corporation: a manifesto for business revolution*. Collins Business Essentials. HarperCollins, 2009.

[55] R. Hogan. *Personality and the Fate of Organizations*. Erlbaum, 2007.

[56] C. Huang, H. Cai, Y. Li, J. Du, F. Bu, and L. Jiang. A process mining based service composition approach for mobile information systems. *Mobile Information Systems*, 2017:1–13, 2017.

[57] J. E. Ingvaldsen and J. A. Gulla. Industrial application of semantic process mining. *Enterprise Information Systems*, 6(2):139–163, 2012.

[58] E. Jagroep, J. M. E. M. van der Werf, J. Broekman, L. Blom, R. van Vliet, and S. Brinkkemper. A resource utilization score for software energy consumption. In *ICT for Sustainability*, pages 19–28, 2016.

[59] M. Jans and M. Hosseinpour. How active learning and process mining can act as continuous auditing catalyst. *International Journal of Accounting Information Systems*, 32:44–58, 2019.

[60] M. Jans, J. M. van der Werf, N. Lybaert, and K. Vanhoof. A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications*, pages 13351–13359, 2011.

[61] M. Jans, M. Alles, and M. Vasarhelyi. The case for process mining in auditing: Sources of value added and areas of application. *IJAIS*, pages 1–20, 2013.

[62] W. Jareevongpiboon and P. Janecek. Ontological approach to enhance results of business process mining and analysis. *Business Process Management Journal*, 19(3):459–476, 2013.

[63] Y. Jean-Mary, E. Shironoshita, and M. Kabuka. Ontology matching with semantic verification. *Web Semantics*, 7(3):235–251, 2009.

[64] O. A. Johnson, T. Ba Dhafari, A. Kurniati, F. Fox, and E. Rojas. The ClearPath Method for Care Pathway Process Mining and Simulation. In *Business Process Management Workshops*, pages 239–250. Springer, 2019.

[65] B. Jokonowo, J. Claes, R. Sarno, and S. Rochimah. Process mining in supply chains: a systematic literature review. *International Journal of Electrical and Computer Engineering*, 8(6):4626–4636, 2018.

[66] S. Kedem-Yemini, N. S. Mamon, and G. Mashiah. An analysis of cargo release services with process mining: a case study in a logistics company. In *Industrial Engineering and Operations Management (IEOM)*, pages 726–736, 2018.

[67] G. Kipping, D. Djurica, S. Franzoi, T. Grisold, L. Marcus, S. Schmid, J. vom Brocke, J. Mendling, and M. Röglinger. How to leverage process mining in organizations - towards process mining capabilities. In *Business Process Management (BPM)*, pages 40–46. Springer, 2022.

[68] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, 2017.

[69] D. Knoll, G. Reinhart, and M. Prüglmeier. Enabling value stream mapping for internal logistics using multidimensional process mining. *Expert Systems with Applications*, 124:130–142, 2019.

[70] J. J. Koorn, I. Beerepoot, V. Stein Dani, X. Lu, I. van de Weerd, H. Leopold, and H. A. Reijers. Bringing rigor to the qualitative evaluation of process mining findings: An analysis and a proposal. In *2021 3rd International Conference on Process Mining (ICPM)*, pages 120–127, 2021.

[71] M. Kudo, T. Nogayama, A. Ishida, and M. Abe. Business process analysis and real-world application scenarios. In *Signal-Image Technology and Internet-Based Systems (SITIS)*. IEEE, 2013.

[72] P. Lambrix and H. Tan. Sambo - a system for aligning and merging biomedical ontologies. *Journal of Web Semantics*, 4(3):196–206, 2006.

[73] K. Lashkevich, F. Milani, D. Chapela-Campa, I. Suvorau, and M. Dumas. Why am i waiting? data-driven analysis of waiting times in business processes. In *Advanced Information Systems Engineering (CAiSE)*, pages 174–190. Springer, 2023.

[74] K. Lashkevich, F. Milani, and N. Danylyshyn. Analysis templates for identifying improvement opportunities with process mining. In *European Conference on Information Systems (ECIS)*, 2023.

[75] T. Lauer. *Success factor person: right leadership in change*, pages 83–106. Springer, 2021.

[76] C. K. H. Lee, G. T. S. Ho, K. L. Choy, and G. K. H. Pang. A RFID-based recursive process mining system for quality assurance in the garment industry. *International Journal of Production Research*, 52(14):4216–4238, 2013.

[77] C. K. H. Lee, K. L. Choy, G. T. S. Ho, and C. H. Y. Lam. A slippery genetic algorithm-based process mining system for achieving better quality assurance in the garment industry. *Expert Systems with Applications*, pages 236–248, 2016.

[78] M. Leemans, W. M. P. van der Aalst, M. G. J. van den Brand, R. R. H. Schiffelers, and L. Lensink. Software process analysis methodology – a methodology based on lessons learned in embracing legacy software. In *International Conference on Software Maintenance and Evolution (ICSME)*, pages 665–674. IEEE, 2018.

[79] S. J. J. Leemans, E. Poppe, and M. T. Wynn. Directly follows-based process mining: exploration and a case study. In *International Conference on Process Mining (ICPM)*, pages 25–32. IEEE, 2019.

[80] A. M. Lemos, C. C. Sabino, R. M. F. Lima, and C. A. L. Oliveira. Using process mining in software development process management: A case study. In *Systems, Man and Cybernetics (SMC)*, pages 1181–1186. IEEE, 2011.

[81] H. Leopold, M. Niepert, M. Weidlich, J. Mendling, R. Dijkman, and H. Stuckenschmidt. Probabilistic optimization of semantic process model matching. In *Business Process Management (BPM)*, pages 319–334. Springer, 2012.

[82] G. Li, E. G. L. de Murillas, R. M. de Carvalho, and W. M. P. van der Aalst. Extracting object-centric event logs to support process mining on databases. In *Information Systems in the Big Data Era*, pages 182–199. Springer, 2018.

[83] Y. Liu, V. S. Dani, I. Beerepoot, and X. Lu. Turning logs into lumber: Preprocessing tasks in process mining, 2024.

[84] J. Madhavan, P. Bernstein, and E. Rahm. Generic schema matching with cupid. *Proc 27th VLDB Conference*, 2001.

[85] E. Mahendrawathi, S. O. Zayin, and F. J. Pamungkas. ERP post implementation review with process mining: a case of procurement process. *Procedia Computer Science*, 124:216–223, 2017.

[86] A. Mamudu, W. Bandara, M. Wynn, and S. Leemans. A process mining success factors model. In *Business Process Management (BPM)*, pages 143–160. Springer, 2022.

[87] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker. Application of process mining in healthcare - A case study in a Dutch Hospital. In *Biomedical Engineering Systems and Technologies*, pages 425–438. Springer, 2009.

[88] F. Marazza, F. A. Bukhsh, O. Vijlbrief, J. Geerdink, S. Pathak, M. Keulen, and C. Seifert. Comparing process models for patient populations: Application in breast cancer care. In *Business Process Management Workshops*, pages 496–507. Springer, 2019.

[89] N. Martin, D. Fischer, G. Kerpedzhiev, K. Goel, S. Leemans, M. Roeglinger, W. M. P. van der Aalst, M. Dumas, M. La Rosa, and C. Stoiber. Opportunities and challenges for process mining in organisations – results of a delphi study. *Business & Information Systems Engineering*, 63(5):511–527, 2022.

[90] L. Măruşter and N. R. T. P. van Beest. Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, 21(3):267–297, 2009.

[91] C. Meilicke, H. Leopold, E. Kuss, H. Stuckenschmidt, and H. A. Reijers. Overcoming individual process model matcher weaknesses using ensemble matching. *Decision Support Systems*, 100:15–26, 2017.

[92] A. Meincheim, C. dos Santos Garcia, J. C. Nievola, and E. E. Scalabrin. Combining process mining with trace clustering: manufacturing shop floor process - an applied case. In *International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2017.

[93] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana,

D. Aloini, et al. Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127:103994, 2022.

[94] D. Nikovski, A. Esenther, X. Ye, M. Shiba, and S. Takayama. Matcher composition methods for automatic schema matching. In *Enterprise Information Systems*, pages 108–123. Springer, 2013.

[95] E. H. J. Nooijen, B. Dongen, and D. Fahland. Automatic discovery of data-centric and artifact-centric processes. In *Business Process Management Workshops*, pages 316–327. Springer, 2012.

[96] G. Park and W. Aalst. Action-oriented process mining: bridging the gap between insights and actions. *Progress in Artificial Intelligence*, pages 1–22, 2022.

[97] G. Park and W. M. P. van der Aalst. A general framework for action-oriented process mining. In *Business Process Management Workshops*, pages 206–218. Springer, 2020.

[98] M. Park, M. Song, T. H. Baek, S. Y. Son, S. J. Ha, and S. W. Cho. Workload and delay analysis in manufacturing process using process mining. In *Asia Pacific Business Process Management (AP-BPM)*, volume 219 of *LNBIP*. Springer, 2015.

[99] A. Partington, M. T. Wynn, S. Suriadi, C. Ouyang, and J. Karnon. Process mining for clinical processes: A comparative analysis of four australian hospitals. *ACM TMIS*, 5(4):1–18, 2015.

[100] M. Peeters, J. Jonge, de, and T. Taris. *An introduction to contemporary work psychology*. Wiley-Blackwell, 2014.

[101] E. M. L. Peters, G. Dedene, and J. Poelmans. Understanding service quality and customer churn by process discovery for a multi-national banking contact center. In *International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2013.

[102] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson. Systematic mapping studies in software engineering. In *Evaluation and Assessment in Software Engineering (EASE)*, 2008.

[103] A. Polyvyanyy, A. Pika, M. T. Wynn, and A. H. ter Hofstede. A systematic approach for discovering causal dependencies between observations and incidents in the health and safety domain. *Safety Science*, 118:345–354, 2019.

[104] R. Post, I. Beerepoot, X. Lu, S. Kas, S. Wiewel, A. Koopman, and H. A. Reijers. Active anomaly detection for key item selection in process auditing. In *Process Mining Workshops*, pages 167–179. Springer, 2022.

[105] J. Potavin, N. Jongswat, and W. Premchaiswadi. Applying Fuzzy-Genetic mining in conformance and dependency relations. In *ICT & KE*, pages 228–235. IEEE, 2012.

[106] F. Ramires and P. Sampaio. Process mining and lean six sigma: a novel approach to analyze the supply chain quality of a hospital. *International Journal of Lean Six Sigma*, pages 594–621, 2021.

[107] L. Reinkemeyer. Process mining in action. *Process mining in action: principles, use Cases and outlook*, 2020.

[108] F. Rismanchian, S. H. Kassani, S. M. Shavarani, and Y. H. Lee. A data-driven approach to support the understanding and improvement of patients' journeys: a case study using electronic health records of an emergency department. *Value in Health*, pages 18–27, 2023.

[109] S. P. Robbins and T. A. Judge. *Organizational behavior*. Pearson Education, 2016.

[110] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro. Process mining in healthcare: A literature review. *Journal of biomedical informatics*, 61: 224–236, 2016.

[111] E. Rojas, A. Cifuentes, A. Burattin, J. Munoz-Gama, M. Sepúlveda, and D. Capurro. Performance analysis of emergency room episodes through process mining. *International Journal of Environmental Research and Public Health*, 16(7):251–263, 2019.

[112] V. A. Rubin, A. A. Mitsyuk, I. A. Lomazova, and W. M. P. van der Aalst. Process mining can be applied to software too! In *International Symposium on Empirical Software Engineering and Measurement (ESEM)*. ACM, 2014.

[113] J. Saldaña. *The coding manual for qualitative researchers*. SAGE, 2015.

[114] J. Samalikova, R. Kusters, J. Trienekens, T. Weijters, and P. Siemons. Toward objective software process information: experiences from a case study. *Software Quality Journal*, pages 101–120, 2010.

[115] K. Smit, , and J. M. and. Process mining in the rail industry: a qualitative analysis of success factors and remaining challenges. In *Humanizing Technology for a Sustainable Society (HTSS)*. University of Maribor Press, 2019.

[116] V. Stein Dani. Event Log Extraction: How to Minimize the Effort of the Human-in-the-Loop? In *International Conference on Process Mining (ICPM) Doctoral Consortium*. CEUR-WS.org, 2020.

[117] V. Stein Dani, M. ER, J. J. Koorn, J. M. E. M. van der Werf, H. Leopold, and H. A. Reijers. Pair modeling: Does one plus one add up? In *Business Process Management Workshops*, pages 261–272. Springer, 2022.

[118] V. Stein Dani, C. Freitas, and L. Thom. Recommendations for visual feedback about problems within BPMN process models. *Software and Systems Modeling*, 21:2039–2065, 2022.

[119] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, X. Lu, I. Beerepoot, J. J. Koorn, and H. A. Reijers. Towards understanding the role of the human in event log extraction. In *Business Process Management Workshops*, pages 86–98. Springer, 2022.

[120] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, I. Beerepoot, and H. A. Reijers. From process mining insights to process improvement: All talk and no action? In *Cooperative Information Systems (CoopIS)*, pages 275–292. Springer, 2023.

[121] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, and H. A. Reijers. Progressing from process mining insights to process improvement: Challenges and recommendations. In *Enterprise Design, Operations, and Computing (EDOC)*, pages 152–168. Springer, 2023.

[122] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, and H. A. Reijers. Supporting event log extraction based on matching. In *Business Process Management Workshops*, pages 322–333. Springer, 2023.

[123] V. Stein Dani, H. Leopold, J. M. E. M. van der Werf, I. Beerepoot, and H. A. Reijers. From loss of interest to denial: A study on the terminators of process mining initiatives. In *Advanced Information Systems Engineering (CAiSE)*, 2024. Accepted for publication.

[124] M. Stuit and H. Wortmann. Discovery and analysis of e-mail-driven business processes. *Information Systems*, 37(2):142–168, 2012.

[125] S. Suriadi, M. T. Wynn, C. Ouyang, A. Hofstede, and N. Dijk. Understanding process behaviours in a large insurance company in Australia: A case study. In *Advanced Information Systems Engineering (CAiSE)*, volume 7908 of *LNCS*, pages 449–464. Springer, 2013.

[126] I. Tawakkal, A. P. Kurniati, and G. A. A. Wisudiawan. Implementing heuristic miner for information system audit based on DSS01 COBIT5. In *International Conference on Computer, Control, Informatics and its Applications*. IEEE, 2016.

[127] A. H. M. Ter Hofstede, A. Koschmider, A. Marrella, R. Andrews, D. A. Fischer, S. Sadeghianasl, M. T. Wynn, M. Comuzzi, J. De Weerdt, K. Goel, N. Martin, and P. Soffer. Process-data quality: The true frontier of process mining. *Journal of Data and Information Quality*, 15(3):1–21, 2023.

[128] K. Toth, K. Machalik, G. Fogarassy, and A. Vathy-Fogarassy. Applicability of process mining in the exploration of healthcare sequences. In *IEEE Neumann Colloquium (NC)*, pages 151–156. IEEE, 2017.

[129] B. Trinkenreich, G. Santos, V. Confort, and F. Santoro. Toward using business process intelligence to support incident management metrics selection and service improvement. In *International Conferences on Software Engineering and Knowledge Engineering (SEKE)*. KSI, 2015.

[130] P. Trkman. The critical success factors of business process management. *International Journal of Information Management*, 30(2):125–134, 2010.

[131] N. van Beest and L. Maruster. A process mining approach to redesign business processes - a case study in gas industry. In *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. IEEE, 2007.

[132] H. van der Aa, H. Leopold, and H. A. Reijers. Comparing textual descriptions to process models – the automatic detection of inconsistencies. *Information Systems*, 64:447–460, 2017.

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

[134] W. M. P. van der Aalst. Extracting event data from databases to unleash process mining. In *BPM-Driving innovation in a digital world*, pages 105–128. Springer, 2015.

[135] W. M. P. van der Aalst. *Process Mining: Data Science in Action*. Springer, 2016.

[136] W. M. P. van der Aalst and J. Carmona. *Scaling process mining to turn insights into actions*, pages 495–502. Springer, 2022.

[137] W. M. P. van der Aalst, H. A. Reijers, A. Weijters, B. F. Dongen, A. Medeiros, M. Song, and H. M. W. Verbeek. Business process mining: An industrial application. *Information Systems*, pages 713–732, 2007.

[138] M. L. van Eck, X. Lu, S. J. J. Leemans, and W. M. P. van der Aalst. PM^2: a process mining project methodology. In *Advanced Information Systems Engineering*, pages 297–313. Springer, 2015.

[139] G. van Hulzen, N. Martin, B. Depaire, and G. Souverijns. Supporting capacity management decisions in healthcare using data-driven process simulation. *Journal of Biomedical Informatics*, page 104060, 2022.

[140] H. M. W. Verbeek, J. C. A. M. Buijs, B. F. van Dongen, and W. M. P. van der Aalst. XES, XESame, and ProM 6. In P. Soffer and E. Proper, editors, *Information Systems Evolution*, pages 60–75. Springer, 2011.

[141] M. Walicki and D. R. Ferreira. Sequence partitioning for process mining with unlabeled event logs. *Data and Knowledge Engineering*, 70(10):821–841, 2011.

[142] Y. Wang, F. Caron, J. Vanthienen, L. Huang, and Y. Guo. Acquiring logistics process intelligence: methodology and an application for a chinese bulk port. *Expert Systems with Applications*, 41(1):195–209, 2014.

[143] J. D. Weerdt, A. Schupp, A. Vanderloock, and B. Baesens. Process mining for the multi-faceted analysis of business processes - a case study in a financial services organization. *Computers in Industry*, 64(1):57–67, 2013.

[144] M. Weidlich, R. Dijkman, and J. Mendling. The icop framework: Identification of correspondences between process models. In *Advanced Information Systems Engineering*, volume 6051, pages 483–498, 2010.

[145] M. Weske, G. Decker, M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers. Model collection of the BPM academic initiative, 2020. URL https://doi.org/10.5281/zenodo.3758704.

[146] P. Zerbino, D. Aloini, R. Dulmin, and V. Mininno. Towards analytics-enabled efficiency improvements in maritime transportation: a case study in a mediterranean port. *Sustainability*, 11(16):4473, 2019.

[147] P. Zerbino, A. Stefanini, and D. Aloini. Process science in action: A literature review on process mining in business management. *Technological Forecasting and Social Change*, 172(9):121021, 2021.

[148] Z. Zhou, Y. Wang, and L. Li. Process mining based modeling and analysis of workflows in clinical care - a case study in a chicago outpatient clinic. In *International Conference on Networking, Sensing and Control (ICNSC)*, pages 590–595. IEEE, 2014.

[149] L. Zimmermann, F. Zerbato, and B. Weber. What makes life for process mining analysts difficult? a reflection of challenges. *Software and Systems Modeling*, 2023.

Curriculum Vitae

As of August 29, 2024, Vinicius Stein Dani is currently a PhD Candidate and Researcher in the Department of Information and Computing Sciences at Utrecht University in the Netherlands. Since February 2020, he has been part of the Utrecht Process Science group, where he has focused on understanding and improving business processes using social, data, and process science techniques. He has (co-)lectured courses such as Business Process Management in the Master of Business Informatics program from 2020 to 2024 and *Procesmodelleren* in the Bachelor of Information Science program during 2021 and 2022. On top of teaching, he has (co-)supervised many Bachelor's and Master's students in their theses and research projects in the context of process mining. From 2021 to 2023, he co-organized the Master of Business Informatics colloquium, encouraging knowledge exchange among Master students working on different research topics. Vinicius' contributions to research have been recognized with a Best Paper Runner-up Award at the 2023 edition of the International Conference on Cooperative Information Systems (CoopIS), for the paper he co-authored with Henrik Leopold, Jan Martijn E. M. van der Werf, Iris Beerepoot and Hajo A. Reijers titled "From Process Mining Insights to Process Improvement: All Talk and No Action?" [120].

Before joining Utrecht University, Vinicius gained valuable industrial and academic experience. In 2019, he worked as a Software Developer and Analyst at Compass.UOL in Erechim, RS, Brazil. There, he analyzed, documented, and implemented client requirements for projects using Oracle web commerce. A significant achievement was related to improving how clients' e-commerce platforms integrated with their enterprise resource planning systems to support multiple distribution centers. This role sharpened his technical skills and deepened his understanding of real-world business process challenges. From 2017 to 2019, Vinicius pursued a Master's Degree in Computer Science with a CAPES scholarship at the Federal University of Rio Grande do Sul, where he also gained experience (co-)lecturing in a Software Engineering course and (co-)supervising students, both on Bachelor level. His research focused on Business Process Management and Information Visualization, especially on business process modeling. He explored new ways to provide visual feedback about problems in process models, adding valuable insights to the field. Before that, he pursued a Master of Business Administration in Information Technology Management at the Federal University of Rio Grande do Sul, while concurrently, he was a Systems Analyst at the University of Passo Fundo from 2015 to 2017. In this role as a Systems Analyst, he analyzed, designed, and implemented systems to meet the needs of internal clients. Notably, he helped develop a new registration system for the institution's entrance exams and redesign the "Pedagogical Course Design" system.

In 2010, Vinicius co-founded Otimalog, a company based in the Santa Maria Technological Incubator in Santa Maria, RS, Brazil. For nearly five years, he and his team developed software solutions aimed at planning and optimizing important educational business processes. Their main products were a software for automatically creating and optimizing timetables and an innovative solution for handling multiple-choice tests. Prior to this, starting in 2005, Vinicius worked independently in several software-related projects, providing custom-made software as a service solutions to clients in Brazil and Uruguay. These ventures provided him with valuable expertise in software analysis and development, project and service management, customer relationship and people management, teamwork, and overall business operations.

SIKS Dissertation Series

2016 01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines
02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow
03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support
04 Laurens Rietveld (VUA), Publishing and Consuming Linked Data
05 Evgeny Sherkhonov (UvA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
06 Michel Wilson (TUD), Robust scheduling in an uncertain environment
07 Jeroen de Man (VUA), Measuring and modeling negative emotions for virtual training
08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data
09 Archana Nottamkandath (VUA), Trusting Crowdsourced Information on Cultural Artefacts
10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms
11 Anne Schuth (UvA), Search Engines that Learn from Their Users
12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems
13 Nana Baah Gyan (VUA), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization
15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
16 Guangliang Li (UvA), Socially Intelligent Autonomous Agents that Learn from Human Reward
17 Berend Weel (VUA), Towards Embodied Evolution of Robot Organisms
18 Albert Meroño Peñuela (VUA), Refining Statistical Data on the Web
19 Julia Efremova (TU/e), Mining Social Structures from Genealogical Data
20 Daan Odijk (UvA), Context & Semantics in News & Web Search
21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
22 Grace Lewis (VUA), Software Architecture Strategies for Cyber-Foraging Systems
23 Fei Cai (UvA), Query Auto Completion in Information Retrieval
24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior
26 Dilhan Thilakarathne (VUA), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media
28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning
30 Ruud Mattheij (TiU), The Eyes Have It
31 Mohammad Khelghati (UT), Deep web content monitoring
32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations
33 Peter Bloem (UvA), Single Sample Statistics, exercises in learning from just one example
34 Dennis Schunselaar (TU/e), Configurable Process Trees: Elicitation, Analysis, and Enactment
35 Zhaochun Ren (UvA), Monitoring Social Media: Summarization, Classification and Recommendation
36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry
38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design
39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
40 Christian Detweiler (TUD), Accounting for Values in Design
41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
42 Spyros Martzoukos (UvA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora

43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice
44 Thibault Sellam (UvA), Automatic Assistants for Database Exploration
45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control
46 Jorge Gallego Perez (UT), Robots to Make you Happy
47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks
48 Tanja Buttler (TUD), Collecting Lessons Learned
49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis
50 Yan Wang (TiU), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

---

2017 01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime
02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation
03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines
04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store
05 Mahdieh Shadi (UvA), Collaboration Behavior
06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search
07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly
08 Rob Konijn (VUA), Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery
09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text
10 Robby van Delden (UT), (Steering) Interactive Play Behavior
11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment
12 Sander Leemans (TU/e), Robust Process Mining with Guarantees
13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology
14 Shoshannah Tekofsky (TiU), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
15 Peter Berck (RUN), Memory-Based Text Correction
16 Aleksandr Chuklin (UvA), Understanding and Modeling Users of Modern Search Engines
17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution
18 Ridho Reinanda (UvA), Entity Associations for Search
19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)
22 Sara Magliacane (VUA), Logics for causal inference under uncertainty
23 David Graus (UvA), Entities of Interest — Discovery in Digital Traces
24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning
25 Veruska Zamborlini (VUA), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search
26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch
27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors
28 John Klein (VUA), Architecture Practices for Complex Contexts
29 Adel Alhuraibi (TiU), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"
30 Wilma Latuny (TiU), The Power of Facial Expressions
31 Ben Ruijl (UL), Advances in computational methods for QFT calculations
32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives
33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity
34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics
35 Martine de Vos (VUA), Interpreting natural science spreadsheets
36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging
37 Alejandro Montes Garcia (TU/e), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
38 Alex Kayal (TUD), Normative Social Applications
39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR
40 Altaf Hussain Abro (VUA), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems
41 Adnan Manzoor (VUA), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets
43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval
44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering
45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement

46 Jan Schneider (OU), Sensor-based Learning Support
47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration
48 Angel Suarez (OU), Collaborative inquiry-based learning

2018 01 Han van der Aa (VUA), Comparing and Aligning Process Representations
02 Felix Mannhardt (TU/e), Multi-perspective Process Mining
03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction
04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
05 Hugo Huurdeman (UvA), Supporting the Complex Dynamics of the Information Seeking Process
06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems
07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems
08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems
09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations
10 Julienka Mollee (VUA), Moving forward: supporting physical activity behavior change through intelligent technology
11 Mahdi Sargolzaei (UvA), Enabling Framework for Service-oriented Collaborative Networks
12 Xixi Lu (TU/e), Using behavioral context in process mining
13 Seyed Amin Tabatabaei (VUA), Computing a Sustainable Future
14 Bart Joosten (TiU), Detecting Social Signals with Spatiotemporal Gabor Filters
15 Naser Davarzani (UM), Biomarker discovery in heart failure
16 Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children
17 Jianpeng Zhang (TU/e), On Graph Sample Clustering
18 Henriette Nakad (UL), De Notaris en Private Rechtspraak
19 Minh Duc Pham (VUA), Emergent relational schemas for RDF
20 Manxia Liu (RUN), Time and Bayesian Networks
21 Aad Slootmaker (OU), EMERGO: a generic platform for authoring and playing scenario-based serious games
22 Eric Fernandes de Mello Araújo (VUA), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
23 Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis
24 Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
25 Riste Gligorov (VUA), Serious Games in Audio-Visual Collections
26 Roelof Anne Jelle de Vries (UT),Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology
27 Maikel Leemans (TU/e), Hierarchical Process Mining for Scalable Software Analysis
28 Christian Willemse (UT), Social Touch Technologies: How they feel and how they make you feel
29 Yu Gu (TiU), Emotion Recognition from Mandarin Speech
30 Wouter Beek (VUA), The "K" in "semantic web" stands for "knowledge": scaling semantics to the web

---

2019 01 Rob van Eijk (UL),Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty
03 Eduardo Gonzalez Lopez de Murillas (TU/e), Process Mining on Databases: Extracting Event Data from Real Life Data Sources
04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data
05 Sebastiaan van Zelst (TU/e), Process Mining with Streaming Data
06 Chris Dijkshoorn (VUA), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets
07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms
08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes
09 Fahimeh Alizadeh Moghaddam (UvA), Self-adaptation for energy efficiency in software systems
10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction
11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs
12 Jacqueline Heinerman (VUA), Better Together
13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation
14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses
15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments
16 Guangming Li (TU/e), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models
17 Ali Hurriyetoglu (RUN),Extracting actionable information from microtexts
18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication
19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents
20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence
21 Cong Liu (TU/e), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection
22 Martin van den Berg (VUA),Improving IT Decisions with Enterprise Architecture
23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification
24 Anca Dumitrache (VUA), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing
25 Emiel van Miltenburg (VUA), Pragmatic factors in (automatic) image description

26 Prince Singh (UT), An Integration Platform for Synchromodal Transport
27 Alessandra Antonaci (OU), The Gamification Design Process applied to (Massive) Open Online Courses
28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
29 Daniel Formolo (VUA), Using virtual agents for simulation and training of social skills in safety-critical circumstances
30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems
31 Milan Jelisavcic (VUA), Alive and Kicking: Baby Steps in Robotics
32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games
33 Anil Yaman (TU/e), Evolution of Biologically Inspired Learning in Artificial Neural Networks
34 Negar Ahmadi (TU/e), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES
35 Lisa Facey-Shaw (OU), Gamification with digital badges in learning programming
36 Kevin Ackermans (OU), Designing Video-Enhanced Rubrics to Master Complex Skills
37 Jian Fang (TUD), Database Acceleration on FPGAs
38 Akos Kadar (OU), Learning visually grounded and multilingual representations

2020 01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour
02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models
03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding
04 Maarten van Gompel (RUN), Context as Linguistic Bridges
05 Yulong Pei (TU/e), On local and global structure mining
06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
07 Wim van der Vegt (OU), Towards a software architecture for reusable game components
08 Ali Mirsoleimani (UL),Structured Parallel Programming for Monte Carlo Tree Search
09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
10 Alifah Syamsiyah (TU/e), In-database Preprocessing for Process Mining
11 Sepideh Mesbah (TUD), Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models
12 Ward van Breda (VUA), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases
15 Konstantinos Georgiadis (OU), Smart CAT: Machine Learning for Configurable Assessments in Serious Games
16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling
17 Daniele Di Mitri (OU), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences
18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems
20 Albert Hankel (VUA), Embedding Green ICT Maturity in Organisations
21 Karine da Silva Miras de Araujo (VUA), Where is the robot?: Life as it could be
22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
24 Lenin da Nóbrega Medeiros (VUA/RUN), How are you feeling, human? Towards emotionally supportive chatbots
25 Xin Du (TU/e), The Uncertainty in Exceptional Model Mining
26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context
28 Bibeg Limbu (TUD), Multimodal interaction for deliberate practice: Training complex skills with augmented reality
29 Ioan Gabriel Bucur (RUN), Being Bayesian about Causal Inference
30 Bob Zadok Blok (UL), Creatief, Creatiever, Creatiefst
31 Gongjin Lan (VUA), Learning better – From Baby to Better
32 Jason Rhuggenaath (TU/e), Revenue management in online markets: pricing and online advertising
33 Rick Gilsing (TU/e), Supporting service-dominant business model evaluation in the context of business model innovation
34 Anna Bon (UM), Intervention or Collaboration? Redesigning Information and Communication Technologies for Development
35 Siamak Farshidi (UU), Multi-Criteria Decision-Making in Software Production

2021 01 Francisco Xavier Dos Santos Fonseca (TUD),Location-based Games for Social Interaction in Public Space
02 Rijk Mercuur (TUD), Simulating Human Routines: Integrating Social Practice Theory in Agent-Based Models
03 Seyyed Hadi Hashemi (UvA), Modeling Users Interacting with Smart Devices
04 Ioana Jivet (OU), The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning
05 Davide Dell'Anna (UU), Data-Driven Supervision of Autonomous Systems

06 Daniel Davison (UT), "Hey robot, what do you think?" How children learn with a social robot
07 Armel Lefebvre (UU), Research data management for open science
08 Nardie Fanchamps (OU), The Influence of Sense-Reason-Act Programming on Computational Thinking
09 Cristina Zaga (UT), The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play
10 Quinten Meertens (UvA), Misclassification Bias in Statistical Learning
11 Anne van Rossum (UL), Nonparametric Bayesian Methods in Robotic Vision
12 Lei Pi (UL), External Knowledge Absorption in Chinese SMEs
13 Bob R. Schadenberg (UT), Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning
14 Negin Samaeemofrad (UL), Business Incubators: The Impact of Their Support
15 Onat Ege Adali (TU/e), Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm
16 Esam A. H. Ghaleb (UM), Bimodal emotion recognition from audio-visual cues
17 Dario Dotti (UM), Human Behavior Understanding from motion and bodily cues using deep neural networks
18 Remi Wieten (UU), Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks
19 Roberto Verdecchia (VUA), Architectural Technical Debt: Identification and Management
20 Masoud Mansoury (TU/e), Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems
21 Pedro Thiago Timbó Holanda (CWI), Progressive Indexes
22 Sihang Qiu (TUD), Conversational Crowdsourcing
23 Hugo Manuel Proença (UL), Robust rules for prediction and description
24 Kaijie Zhu (TU/e), On Efficient Temporal Subgraph Query Processing
25 Eoin Martino Grua (VUA), The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications
26 Benno Kruit (CWI/VUA), Reading the Grid: Extending Knowledge Bases from Human-readable Tables
27 Jelte van Waterschoot (UT), Personalized and Personal Conversations: Designing Agents Who Want to Connect With You
28 Christoph Selig (UL), Understanding the Heterogeneity of Corporate Entrepreneurship Programs

---

2022 01 Judith van Stegeren (UT), Flavor text generation for role-playing video games
02 Paulo da Costa (TU/e), Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey
03 Ali el Hassouni (VUA), A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare
04 Ünal Aksu (UU), A Cross-Organizational Process Mining Framework
05 Shiwei Liu (TU/e), Sparse Neural Network Training with In-Time Over-Parameterization
06 Reza Refaei Afshar (TU/e), Machine Learning for Ad Publishers in Real Time Bidding
07 Sambit Praharaj (OU), Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics
08 Maikel L. van Eck (TU/e), Process Mining for Smart Product Design
09 Oana Andreea Inel (VUA), Understanding Events: A Diversity-driven Human-Machine Approach
10 Felipe Moraes Gomes (TUD), Examining the Effectiveness of Collaborative Search Engines
11 Mirjam de Haas (UT), Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers' engagement with robots and tasks during second-language tutoring
12 Guanyi Chen (UU), Computational Generation of Chinese Noun Phrases
13 Xander Wilcke (VUA), Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge
14 Michiel Overeem (UU), Evolution of Low-Code Platforms
15 Jelmer Jan Koorn (UU), Work in Process: Unearthing Meaning using Process Mining
16 Pieter Gijsbers (TU/e), Systems for AutoML Research
17 Laura van der Lubbe (VUA), Empowering vulnerable people with serious games and gamification
18 Paris Mavromoustakos Blom (TiU), Player Affect Modelling and Video Game Personalisation
19 Bilge Yigit Ozkan (UU), Cybersecurity Maturity Assessment and Standardisation
20 Fakhra Jabeen (VUA), Dark Side of the Digital Media - Computational Analysis of Negative Human Behaviors on Social Media
21 Seethu Mariyam Christopher (UM), Intelligent Toys for Physical and Cognitive Assessments
22 Alexandra Sierra Rativa (TiU), Virtual Character Design and its potential to foster Empathy, Immersion, and Collaboration Skills in Video Games and Virtual Reality Simulations
23 Ilir Kola (TUD), Enabling Social Situation Awareness in Support Agents
24 Samaneh Heidari (UU), Agents with Social Norms and Values - A framework for agent based social simulations with social norms and personal values
25 Anna L.D. Latour (UL), Optimal decision-making under constraints and uncertainty
26 Anne Dirkson (UL), Knowledge Discovery from Patient Forums: Gaining novel medical insights from patient experiences
27 Christos Athanasiadis (UM), Emotion-aware cross-modal domain adaptation in video sequences
28 Onuralp Ulusoy (UU), Privacy in Collaborative Systems
29 Jan Kolkmeier (UT), From Head Transform to Mind Transplant: Social Interactions in Mixed Reality

30 Dean De Leo (CWI), Analysis of Dynamic Graphs on Sparse Arrays
31 Konstantinos Traganos (TU/e), Tackling Complexity in Smart Manufacturing with Advanced Manufacturing Process Management
32 Cezara Pastrav (UU), Social simulation for socio-ecological systems
33 Brinn Hekkelman (CWI/TUD), Fair Mechanisms for Smart Grid Congestion Management
34 Nimat Ullah (VUA), Mind Your Behaviour: Computational Modelling of Emotion & Desire Regulation for Behaviour Change
35 Mike E.U. Ligthart (VUA), Shaping the Child-Robot Relationship: Interaction Design Patterns for a Sustainable Interaction

---

2023 01 Bojan Simoski (VUA), Untangling the Puzzle of Digital Health Interventions
02 Mariana Rachel Dias da Silva (TiU), Grounded or in flight? What our bodies can tell us about the whereabouts of our thoughts
03 Shabnam Najafian (TUD), User Modeling for Privacy-preserving Explanations in Group Recommendations
04 Gineke Wiggers (UL), The Relevance of Impact: bibliometric-enhanced legal information retrieval
05 Anton Bouter (CWI), Optimal Mixing Evolutionary Algorithms for Large-Scale Real-Valued Optimization, Including Real-World Medical Applications
06 António Pereira Barata (UL), Reliable and Fair Machine Learning for Risk Assessment
07 Tianjin Huang (TU/e), The Roles of Adversarial Examples on Trustworthiness of Deep Learning
08 Lu Yin (TU/e), Knowledge Elicitation using Psychometric Learning
09 Xu Wang (VUA), Scientific Dataset Recommendation with Semantic Techniques
10 Dennis J.N.J. Soemers (UM), Learning State-Action Features for General Game Playing
11 Fawad Taj (VUA), Towards Motivating Machines: Computational Modeling of the Mechanism of Actions for Effective Digital Health Behavior Change Applications
12 Tessel Bogaard (VUA), Using Metadata to Understand Search Behavior in Digital Libraries
13 Injy Sarhan (UU), Open Information Extraction for Knowledge Representation
14 Selma Čaušević (TUD), Energy resilience through self-organization
15 Alvaro Henrique Chaim Correia (TU/e), Insights on Learning Tractable Probabilistic Graphical Models
16 Peter Blomsma (TiU), Building Embodied Conversational Agents: Observations on human nonverbal behaviour as a resource for the development of artificial characters
17 Meike Nauta (UT), Explainable AI and Interpretable Computer Vision – From Oversight to Insight
18 Gustavo Penha (TUD), Designing and Diagnosing Models for Conversational Search and Recommendation
19 George Aalbers (TiU), Digital Traces of the Mind: Using Smartphones to Capture Signals of Well-Being in Individuals
20 Arkadiy Dushatskiy (TUD), Expensive Optimization with Model-Based Evolutionary Algorithms applied to Medical Image Segmentation using Deep Learning
21 Gerrit Jan de Bruin (UL), Network Analysis Methods for Smart Inspection in the Transport Domain
22 Alireza Shojaifar (UU), Volitional Cybersecurity
23 Theo Theunissen (UU), Documentation in Continuous Software Development
24 Agathe Balayn (TUD), Practices Towards Hazardous Failure Diagnosis in Machine Learning
25 Jurian Baas (UU), Entity Resolution on Historical Knowledge Graphs
26 Loek Tonnaer (TU/e), Linearly Symmetry-Based Disentangled Representations and their Out-of-Distribution Behaviour
27 Ghada Sokar (TU/e), Learning Continually Under Changing Data Distributions
28 Floris den Hengst (VUA), Learning to Behave: Reinforcement Learning in Human Contexts
29 Tim Draws (TUD), Understanding Viewpoint Biases in Web Search Results

---

2024 01 Daphne Miedema (TU/e), On Learning SQL: Disentangling concepts in data systems education
02 Emile van Krieken (VUA), Optimisation in Neurosymbolic Learning Systems
03 Feri Wijayanto (RUN), Automated Model Selection for Rasch and Mediation Analysis
04 Mike Huisman (UL), Understanding Deep Meta-Learning
05 Yiyong Gou (UM), Aerial Robotic Operations: Multi-environment Cooperative Inspection & Construction Crack Autonomous Repair
06 Azqa Nadeem (TUD), Understanding Adversary Behavior via XAI: Leveraging Sequence Clustering to Extract Threat Intelligence
07 Parisa Shayan (TiU), Modeling User Behavior in Learning Management Systems
08 Xin Zhou (UvA), From Empowering to Motivating: Enhancing Policy Enforcement through Process Design and Incentive Implementation
09 Giso Dal (UT), Probabilistic Inference Using Partitioned Bayesian Networks
10 Cristina-Iulia Bucur (VUA), Linkflows: Towards Genuine Semantic Publishing in Science
11 withdrawn
12 Peide Zhu (TUD), Towards Robust Automatic Question Generation For Learning
13 Enrico Liscio (TUD), Context-Specific Value Inference via Hybrid Intelligence
14 Larissa Capobianco Shimomura (TU/e), On Graph Generating Dependencies and their Applications in Data Profiling
15 Ting Liu (VUA), A Gut Feeling: Biomedical Knowledge Graphs for Interrelating the Gut Microbiome and Mental Health

16 Arthur Barbosa Câmara (TUD), Designing Search-as-Learning Systems
17 Razieh Alidoosti (VUA), Ethics-aware Software Architecture Design
18 Laurens Stoop (UU), Data Driven Understanding of Energy-Meteorological Variability and its Impact on Energy System Operations
19 Azadeh Mozafari Mehr (TU/e), Multi-perspective Conformance Checking: Identifying and Understanding Patterns of Anomalous Behavior
20 Ritsart Anne Plantenga (UL), Omgang met Regels
21 Federica Vinella (UU), Crowdsourcing User-Centered Teams
22 Zeynep Ozturk Yurt (TU/e), Beyond Routine: Extending BPM for Knowledge-Intensive Processes with Controllable Dynamic Contexts
23 Jie Luo (VUA), Lamarck's Revenge: Inheritance of Learned Traits Improves Robot Evolution
24 Nirmal Roy (TUD), Exploring the effects of interactive interfaces on user search behaviour
25 Alisa Rieger (TUD), Striving for Responsible Opinion Formation in Web Search on Debated Topics
26 Tim Gubner (CWI), Adaptively Generating Heterogeneous Execution Strategies using the VOILA Framework
27 Lincen Yang (UL), Information-theoretic Partition-based Models for Interpretable Machine Learning
28 Leon Helwerda (UL), Grip on Software: Understanding development progress of Scrum sprints and backlogs
29 David Wilson Romero Guzman (VUA), The Good, the Efficient and the Inductive Biases: Exploring Efficiency in Deep Learning Through the Use of Inductive Biases
30 Vijanti Ramautar (UU), Model-Driven Sustainability Accounting
31 Ziyu Li (TUD), On the Utility of Metadata to Optimize Machine Learning Workflows