ProMiSE: Process Mining Support for End-Users

Francesca Zerbato[1,*], Lisa Zimmermann[1,*], Hagen Völzer[1,*] and Barbara Weber[1,*]

[1]Institute of Computer Science, University of St.Gallen, St.Gallen, Switzerland

Abstract

In the past decade, process mining has gained momentum in academia and the industry, as it supports organizations in deriving insights from event data recorded from process executions. The increasing adoption of process mining in practice entails supporting process analysts in their work. Indeed, their analysis includes many exploratory tasks that require them to rely on their experience to interpret the data and steer the analysis. This knowledge-intensive nature of process mining can be challenging for less experienced analysts and calls for methodological and operational guidance tailored to their needs.

In this paper, we present *ProMiSE*, a project funded by the Swiss National Science Foundation that embraces this novel direction in process mining research. The first goal of the project is to improve our understanding of how analysts work in practice, i.e., the *process of process mining*. Then, methodological guidance and software-based support are developed to assist novice analysts during their analysis.

The results obtained in the first two years of ProMiSE have helped to build a solid empirical basis on process mining, laying the foundation for the development of user-centered support, which we will realize in the coming years with the help of our project partners and international collaborators.

Keywords

Process of Process Mining, User Behavior Analysis, Process Mining Guidance, Software Support

Project Information *Project Name*: Process Mining Support for End-Users (ProMiSE); *Funding Agency*: Swiss National Science Foundation (SNSF); *URL*: https://data.snf.ch/grants/grant/197032; *Running Period*: 01.11.2020 – 31.10.2024; *Applicant*: Barbara Weber (University of St. Gallen); *Participants*: Francesca Zerbato, Lisa Zimmermann, Hagen Völzer; *Project Partners*: Andrea Burattin (Technical University of Denmark), Tijs Slaats (University of Copenhagen), Pnina Soffer (University of Haifa)

1. Introduction

Process mining is a flourishing discipline rooted in workflow mining [1] and in the discovery of software development and maintenance processes [2]. It blends concepts from machine learning, data mining, and business process management (BPM) [3]. Process mining enables the automated analysis of event data recorded by an IT system that supports the execution of a (business) process. The goal of the analysis is to discover, visualize, improve, or automate the

RPE@CAiSE'23: Research Projects Exhibition at the International Conference on Advanced Information Systems Engineering, June 12–16, 2023, Zaragoza, Spain

*Corresponding author.
francesca.zerbato@unisg.ch (F. Zerbato); lisa.zimmermann@unisg.ch (L. Zimmermann); hagen.voelzer@unisg.ch (H. Völzer); barbara.weber@unisg.ch (B. Weber)
0000-0001-7797-4602 (F. Zerbato); 0000-0002-6149-7060 (L. Zimmermann); 0000-0003-3547-3847 (H. Völzer); 0000-0002-6004-4860 (B. Weber)

process. Over the years, process mining has grown in many directions, leading to the emergence of many open-source and commercial tools and increasing adoption by organizations.

However, despite its success in academia and industry, research on process mining has mainly advanced from a technical perspective, favoring the development of algorithms and tools over guidance for individual users [4], particularly process analysts. As a result, many analytical tasks still lack methodological and operational support [5].

While this lack of support manifests itself at various stages of a process mining project [6], it becomes particularly critical for the analysis phase and, above all, for *exploratory analysis* tasks. Indeed, the knowledge-intensive character of process mining analysis requires analysts to explore the data and find their way through the many possible techniques that can be applied to it [7]. Due to its emergent nature, such an exploratory process cannot be fully specified before it has happened [8], making it hard to guide process analysts and, in particular, non-expert users such as novice analysts.

With our project, *Process Mining Support for End-Users (ProMiSE)* [1], we aim to close this gap by contributing to a better understanding of how process analysts do their work in practice and by providing methodological guidance and software-based support to assist novice analysts.

This broad project goal can be split into two **research objectives** (ROs).

- The first objective is to **understand how analysts do process mining in practice**, i.e., how they act and think in the *process of process mining* (RO1). To achieve this, we analyze the behavior of individual process mining users in a comprehensive manner. In detail, we use a rich source of multi-modal data, including interaction and verbal data, to investigate the work practices of analysts with different levels of expertise. This approach strengthens a new line of research in process mining [5, 9], borrowing from neighboring fields such as visual analytics [7] and data-driven requirements engineering [10], which already focus on the analysis of user behavior.
- The second objective is to **develop and evaluate methodological guidance and software-based support** to assist novice analysts during the analysis (RO2). The developed guidance and support will build upon the results of RO1. First, we aim to develop guidelines to support novice analysts. The envisioned guidelines aim to complement the high-level support provided by existing methodologies [6, 11] with practical advice for specific aspects of the analysis phase, such as analysis tactics or mitigation strategies for everyday challenges. In addition, we plan to support the user during the analysis, bearing in mind that the unstructured and emergent nature of process mining analysis might not be suitable for approaches that prescribe or enforce behavior. Thus, we will develop and evaluate software tools and artifacts to support process analysts in specific tasks.

Relevance to Information Systems Engineering Research ProMiSE is relevant to information systems engineering in several ways. Firstly, the research in this project follows the principles of design science research [12] and user-centered design [13], where the design of the support is informed by an in-depth analysis of user behavior data and then evaluated with end-users. Secondly, process mining can be seen as a strand of *data-driven requirements engineering* [10, 14]. Traditionally, process mining identifies improvement and automation op-

[1]Grant information about ProMiSE: https://data.snf.ch/grants/grant/197032

portunities for business processes supported by information systems. However, process mining is not restricted to business processes but can also be applied to manufacturing, healthcare, and other types of processes. Thirdly, the project itself can be seen as a case study in data-driven engineering. The status-quo in process mining is comparable to other knowledge-intensive work: A variety of automated tool functions exist but those are used in an exploratory and unstructured way. How can then empirical evidence about the use of tool functions help to engineer the next generation of tool support? In this regard, we hope to transfer learnings from the project to engineer the support for knowledge-intensive work in other areas.

2. Work Packages and Outputs

The work of ProMiSE is organized into four main work packages (WPs) which relate to the two research objectives of the project as sketched in Fig. 1.

RO1 focuses on understanding user behavior based on data collected from users engaging in process mining analysis (WP1), which we analyze based on specific research questions (WP2).

Study Design and Data Collection (WP1). The data collection is organized into a series of observational studies aimed to collect different kinds of behavioral data from process mining users with varying levels of experience and expertise. On the one hand, we aim to collect data about the analysis process in the tools, such as screen recordings and interaction logs, to learn from the usage of widely used process mining tools and artifacts (implicit feedback). On the other hand, we aim to collect verbal data from think-aloud and interviews to learn about the thinking processes of analysts, including their reflections on the analysis process and the perceived challenges (explicit feedback). By combining different data modalities, we aim to create a unique dataset that includes both interaction and verbal data. The latter serves to

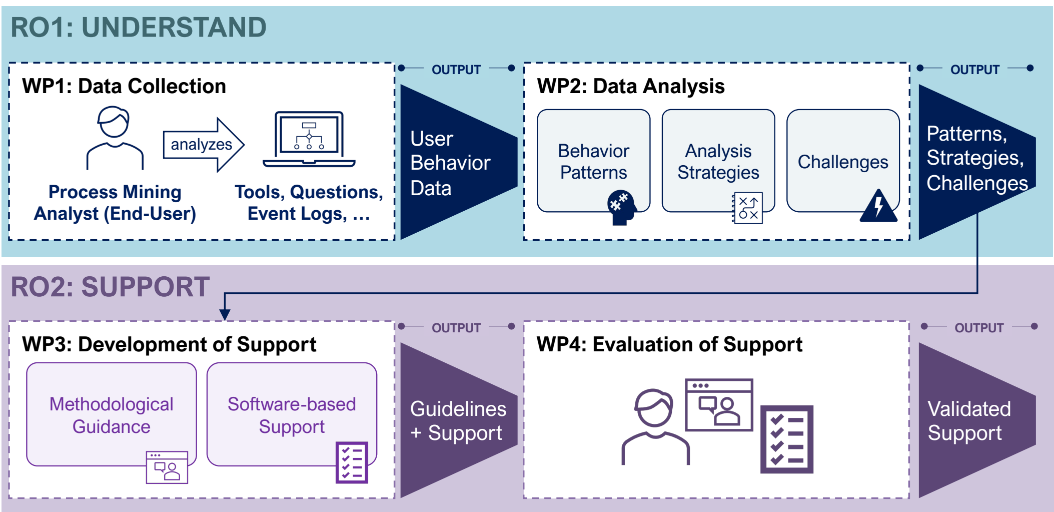

Figure 1: Overview of the main research objectives (ROs) and work packages (WPs) of ProMiSE.

contextualize the user interactions and ease the explanation of the observed behavior.

The intended output of WP1 is a collection of multi-modal data capturing the process of process mining both in terms of low-level interactions with process mining software and higher-level thinking processes of the analysts.

Data Analysis (WP2). In WP2, we study the collected data from various viewpoints to understand user behavior guided by the following research questions.

- "What are recurring patterns of behavior in process mining?" We analyze interaction traces at varying levels of abstraction and from different perspectives. Examples of such perspectives can be the expertise of the analyst and the tool used, but also specific phases of the analysis, such as data exploration. Analysts tend to reapply the same work practices to different datasets [5], suggesting that common analysis patterns as well as different working modes exist, which are dictated by the analyst's expertise or the context.
- "Can we identify common approaches to process mining tasks?" We analyze verbal and interaction data to find out whether there are common strategies that analysts employ. This includes how analysts typically start, how they use different analysis artifacts, and which factors they take into account to steer their analysis.
- "What are common challenges arising during the analysis?" We identify challenges faced by analysts during the analysis by looking into verbal data. This direction is motivated by work on technical and organizational challenges [3, 15], which we aim to integrate.

To achieve RO1, ProMiSE will combine the knowledge gained on behavioral patterns, analysis strategies, and challenges. The insights gained with RO1 will inform the development of guidance and support (WP3), which we will then evaluate with novice analysts (WP4) (cf. Fig. 1).

Development (WP3) and Evaluation (WP4) of Support. WP3 focuses on developing methodological guidance and software-based support for novice analysts in a similar vein to work done on the process of process modeling [16]. The envisioned support encompasses guidelines inspired by "effective" analysis strategies but also novel visualizations for process mining results, as well as mechanisms to support analysts in keeping track of their analysis for reproducibility purposes. Moreover, we aim to support analysts with recommendations based on the operational knowledge of expert users, providing hints on how to execute analysis tasks including where to start from, how to develop questions, and how to overcome common challenges. Then, in WP4, we aim to iteratively evaluate the support developed in WP3 with novice analysts, i.e., our end-users.

3. Results and Ongoing Work

As of spring 2023, ProMiSE has reached the halfway point of its duration, and crucial milestones have been achieved. Below we summarize the results obtained and report on ongoing work.

3.1. Understanding how Analysts do Process Mining in Practice

Study Design and Data Collection. The first efforts of the project focused on the design and execution of a study targeting process mining analysts (cf. WP 1). For the study, participants were tasked with answering an analysis question by analyzing a provided event log using a process mining tool of choice. During the task, we collected different sources of behavioral data, including (i) the interaction traces of the analysts working in the process mining tools captured via screen recordings, (ii) their reasoning processes captured via think-aloud, and (iii) the application logs of the process mining tools, if available. Also, we tracked (iv) the (final) answers to the question posed in the task and complemented all the gathered information with (v) retrospective *interviews*. While (i)-(iv) concern mainly the task planned for the study, (v) enriches the collected data with insights coming from the general work practices of the analysts. The design of the study was first tested in a pilot study with 12 participants. Based on the lessons learned from the pilot, we refined our design, and in the summer of 2021, we conducted a large-scale data collection involving 41 analysts with varying levels of experience and expertise. This rich multi-modal data collection formed a solid empirical basis for our analysis.

Exploration Patterns, Analysis Strategies, and Challenges. The analysis of the data collected from both studies allowed us to contribute to RO1 in several ways.

From the interaction traces collected in the pilot study, we derived insights into analysis patterns and strategies of exploratory process mining [17]. Supported by qualitative and visual analysis, we discovered different behavior patterns based on time spent by analysts on high-level activities executed in process mining tools and their order. The patterns were complemented by exploration goals and strategies derived from the interviews. Our findings revealed that although process analysts share some goals with data scientists [7], their exploration strategies depend on the characteristics of the studied process, making process mining analysis a promising area to explore further.

Based on the analysis of interview data collected in the second study, we addressed the research questions about strategies and challenges. In [18], we characterized common strategies of the analysis phase [6] and examined factors affecting their use in practice. We grouped the discovered strategies into four groups. The first group includes strategies often used at the beginning of the analysis to understand the problem, the data, and the domain. A second group of strategies supports the analysis planning, e.g., by prioritizing analysis directions based on the foreseen value of the findings. Then, in the third group, we identified strategies to execute the analysis by applying process mining techniques within tools, e.g., testing hypotheses. The last group of strategies covered the verification and validation of the results. Our work also shed light on factors that influence the application of the strategies. For example, we discovered that the presence or absence of an analysis question, the availability of stakeholders, and the analyst's role within their team influence whether and how specific strategies are applied.

Besides strategies, we discovered common challenges experienced by process analysts while they engage in process mining projects [19]. In total, we identified 23 challenges, including issues such as keeping the focus of the analysis, formulating questions to guide the analysis, not having sufficient analysis experience, or struggling to draw conclusions from analysis results. These findings helped us identify the actual support needs of analysts and provide detailed insights, such as situations in which the challenges occur and how they affect an analysis. The obtained insights already hint towards potential solutions. As a first step towards support, we will develop guidance to circumvent common challenges. By learning from experts who encountered challenges in their past experience and successfully overcame them, we plan to derive *mitigation strategies*, which novice analysts can also apply.

Last but not least, towards achieving RO1, we are currently analyzing the interaction data. Following a qualitative coding approach [20], we have coded all the data from multiple perspectives and have started applying quantitative methods to mine recurring patterns of behavior.

3.2. Towards Developing Support for Process Analysts

The data collection and analysis conducted in the context of RO1 laid the groundwork for the development of support, allowing us not only to derive preliminary results in this direction but also to identify areas where support is still lacking (cf. RO2).

Question Development in Process Mining. One area is the development of questions to guide process mining analysis. Inspired by the discovered challenges [19] and the fact that questions are a critical factor for applying analysis strategies [18], we focused on providing support for question development. In [21], we derived recommendations from practice aimed at supporting process analysts in developing analysis questions or dealing with the lack thereof. To further support question development, we are currently working on gathering process mining questions from literature and user surveys. We envision that the categorization of questions and their link to analysis techniques could support analysts in identifying and formulating relevant questions and devising a plan to answer them with the help of tools, such as done in [22].

Mining of Multi-Granular Activities. Another challenge we identified concerns how to chose a proper granularity level for representing activities in logs [23, 24]. This fundamental problem emerged from the exploratory analysis of the interaction traces. In this setting, it is hard to foresee what is the most appropriate granularity level of process activities that can yield meaningful patterns. Analysts would benefit from exploring activities at different levels of granularity in the same analysis without having to fix one during preprocessing, as often done for event logs. Further advancements in this directions are planned, as a successful solution to the granularity problem in process mining would not only have far-reaching implications for the process of process mining but also for the integration of process mining with task mining as well as the analysis of human behavioral data and process data in the context of IoT.

Analytic Provenance and Data Awareness. A third problem area identified from our studies is that process analysts often lose awareness of their current data selection and struggle to track how the current analysis step relates to previous steps, previous results, and the goals of the analysis. We have analyzed this problem and proposed tool support to address it [25]. The tool design that we have proposed aims at increasing the transparency and rigor of exploratory process mining as a basis for its stepwise maturation. In this work, we have also provided an initial evaluation concerning the feasibility of some aspects of the design.

Interactive Modeling. From our study, we have learned that process analysis is often about generating and testing hypotheses based on patterns from visualizations. But patterns can also be found through easy-to-read textual abstractions such as *rules*. We are currently developing and evaluating tool support for generating novel visualizations and rules for *process outcome analysis*. Process outcome analysis is a critical use case in process mining that allows one to explain the distribution of final or intermediate process outcomes. Such explanations are a basis for process improvements, e.g., increasing the share of "good" outcomes of the process. In this work, we leverage and extend Sankey diagrams and interpretable machine learning techniques.

4. Conclusion and Outlook

With this paper, we have provided an overview of the main research activities of ProMiSE. While the project is planned to run for approximately two more years, significant milestones have been achieved and disseminated to the research community. So far, the project has focused on understanding the process of process mining, putting a particular emphasis on the analysis phase. Through this effort, the team in St. Gallen has worked closely with project partners and collaborators to identify typical analysis challenges and strategies as well as problem areas that might benefit from the development of support in the short term.

In the coming months, we will complement the qualitative insights obtained with the quantitative insights into behavior patterns. While doing so, we will look deeper at the granularity of process activities and work towards supporting multi-granular analysis. Moreover, we will focus on translating the findings into support for the areas identified above. For software-based support, we will follow at least two lines of work: (i) increase provenance and data awareness for the process analyst to reach higher maturity in exploratory process mining [25], and (ii) create support for process outcome analysis as described above. We also plan to extend tool support for *variant analysis*, one of the most frequently used functions in existing tools. Upon project completion, we will to share the results and anonymized data with the research community.

Funding. This work is funded by the Swiss National Science Foundation under grant no. 200021_197032.

References

[1] W. M. Van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, A. J. Weijters, Workflow mining: A survey of issues and approaches, Data Knowl Eng 47 (2003) 237–267.

[2] J. E. Cook, A. L. Wolf, Discovering models of software processes from event-based data, ACM Trans. Softw. Eng. Methodol. 7 (1998) 215–249. doi:10.1145/287000.287001.

[3] W. M. Van Der Aalst, et al., Process mining manifesto, in: Business Process Management Workshops, Springer, 2011, pp. 169–194. doi:10.1007/978-3-642-28108-2_19.

[4] C. Cho, H. R'bigui, The state-of-the-art of business process mining challenges, Int. J. Bus. Process. Integration Manag. 8 (2017) 285 – 303. doi:10.1504/IJBPIM.2017.10009731.

[5] C. Klinkmüller, R. Müller, I. Weber, Mining process mining practices: An exploratory characterization of information needs in process analytics, in: Int. Conf. on Business Process Management (BPM), Springer, Cham, 2019, pp. 322–337. doi:10.1007/978-3-030-26619-6_21.

[6] F. Emamjome, R. Andrews, A. H. ter Hofstede, A case study lens on process mining in practice, in: On the Move to Meaningful Internet Systems: OTM 2019 Conferences, Springer, 2019, pp. 127–145.

[7] K. Wongsuphasawat, Y. Liu, J. Heer, Goals, process, and challenges of exploratory data analysis: An interview study, arXiv preprint arXiv:1911.00568 (2019).
[8] M. Reichert, B. Weber, Enabling Flexibility in Process-Aware Information Systems: Challenges, Methods, Technologies, Springer, Berlin, Heidelberg, 2012.
[9] C. Capitán-Agudo, M. Salas-Urbano, C. Cabanillas, M. Resinas, Analyzing how process mining reports answer time performance questions, in: Int. Conf. on Business Process Management (BPM), Springer, 2022, pp. 234–250. doi:10.1007/978-3-031-16103-2_17.
[10] W. Maalej, M. Nayebi, T. Johann, G. Ruhe, Toward data-driven requirements engineering, IEEE Software 33 (2015) 48–54. doi:10.1109/MS.2015.153.
[11] M. L. Van Eck, X. Lu, S. J. Leemans, W. M. Van der Aalst, PM^2: A process mining project methodology, in: Int. Conf. on Advanced Information Systems Engineering (CAiSE), volume 9097 of *LNCS*, 2015, pp. 297–313. doi:10.1007/978-3-319-19069-3_19.
[12] R. Wieringa, Design science methodology for information systems and software engineering, Springer, 2014. doi:10.1007/978-3-662-43839-8.
[13] W. J. Hansen, User engineering principles for interactive systems, in: Fall Joint Computer Conference (AFIPS), 1972, pp. 523–532. doi:10.1145/1479064.1479159.
[14] X. Franch, Data-driven requirements engineering: A guided tour, in: Int. Conf. on Evaluation of Novel Approaches to Software Engineering (ENASE), volume 1375 of *CCIS*, Springer, 2020, pp. 83–105. doi:10.1007/978-3-030-70006-5_4.
[15] N. Martin, et al., Opportunities and challenges for process mining in organisations: Results of a Delphi study, Bus. Inf. Syst. Eng. 63 (2021). doi:10.1007/s12599-021-00720-0.
[16] J. Pinggera, et al., Styles in business process modeling: an exploration and a model, Softw Syst Model 14 (2015) 1055–1080. doi:10.1007/s10270-013-0349-1.
[17] F. Zerbato, P. Soffer, B. Weber, Initial insights into exploratory process mining practices, in: Int. Conf. on Business Process Management (BPM Forum), volume 427 of *LNBIP*, Springer, 2021, pp. 145–161. doi:10.1007/978-3-030-85440-9_9.
[18] F. Zerbato, P. Soffer, B. Weber, Process mining practices: Evidence from interviews, in: Int. Conf. on Business Process Management (BPM), LNCS, Springer, 2022, pp. 268–285. doi:10.1007/978-3-031-16103-2_19.
[19] L. Zimmermann, F. Zerbato, B. Weber, Process mining challenges perceived by analysts: An interview study, in: Enterprise, Business-Process and Information Systems Modeling (BPMDS), Springer, 2022, pp. 3–17. doi:10.1007/978-3-031-07475-2_1.
[20] J. Saldaña, The coding manual for qualitative researchers, Sage, 2015.
[21] F. Zerbato, J. J. Koorn, I. Beerepoot, B. Weber, H. A. Reijers, On the origin of questions in process mining projects, in: Int. Conf. on Enterprise Design, Operations, and Computing (EDOC), Springer, 2022, pp. 165–181. doi:10.1007/978-3-031-17604-3_10.
[22] L. Barbieri, E. Madeira, K. Stroeh, W. van der Aalst, A natural language querying interface for process mining, J Intell Inf Syst (2022) 1–30. doi:10.1007/s10844-022-00759-9.
[23] F. Zerbato, R. Seiger, G. D. Federico, A. Burattin, B. Weber, Granularity in process mining: Can we fix it?, in: BPM Problems to Solve Before We Die, volume 2938, CEUR-WS, 2021, pp. 40–44.
[24] I. Beerepoot, et al., The biggest business process management problems to solve before we die, Computers in Industry 146 (2023) 103837. doi:10.1016/j.compind.2022.103837.
[25] F. Zerbato, A. Burattin, H. Völzer, E. Boscaini, P. Nelson Becker, B. Weber, Supporting provenance and data awareness in exploratory process mining, in: Int. Conf. on Advanced Information Systems Engineering (CAiSE), 2023. Accepted for publication.