Understanding and Supporting Process Mining Analyses at the Individual Level

Lisa Zimmermann[1]

[1]University of St. Gallen, St Gallen, Switzerland

1. Motivation and Background

Process mining (PM) bridges data- and process science, thus combining disciplines like business process management, process automation, databases, machine learning, and visual analytics [1]. It enables the analysis of processes through data records, mainly in business contexts where processes are “what companies do whenever they deliver a service or a product to customers” [2]. Notably, also internal actions, or more broadly all “coherent series of changes, both man-made and naturally occurring” can be understood and analyzed from a process perspective [3].

While the technology advanced significantly in recent years and PM tools are transitioning towards comprehensive platforms, aiming to establish PM as integral in business process management initiatives [4], practical PM projects might still fall short of expectations [5]. Mamudu et al. [6] identified factors affecting project success, including among others, data quality, tool choice, team composition, analyst expertise, and project planning. In addition to data aspects, techniques, and organizational structures, these findings emphasize the importance of human factors for the success of PM projects. In line with the research framework from vom Brocke et al. [7], I will refer to these human factors as the *Individual Level* of PM. It involves the use, adoption, and conduct of PM tasks by individual actors, such as process analysts [7].

Process analysts play a central role in PM projects, applying techniques and deriving insights for their organizations [8, 9, 10]. However, their tasks are often emergent, ad-hoc, and manual, which consequently poses significant challenges, especially for novice analysts [11, 9, 12]. Therefore, the focus of my doctoral project is the individual level of process mining, with the objective of conducting detailed analyses and developing tools and methodologies to support less experienced analysts in successfully implementing PM projects.

In the context of an individual perspective, I want to emphasize those steps in PM projects that require human intervention and reasoning. PM projects commonly involve six phases: scoping, data collection, data preparation, mining and analysis, evaluation, and implementation [13]. Analysts are especially involved in the scoping, analysis, and evaluation phases, where they formulate questions, employ techniques, explore data, and assess results to provide answers [8, 9]). My doctoral project will therefore focus on these three project phases.

Proceedings of the Best BPM Dissertation Award, Doctoral Consortium, and Demonstrations & Resources Forum co-located with 22nd International Conference on Business Process Management (BPM 2024), Krakow, Poland, September 1st to 6th, 2024.

lisa.zimmermann@unisg.ch (L. Zimmermann)
0000-0002-6149-7060 (L. Zimmermann)

2. Research Proposition

As part of my doctoral project, I aim to facilitate the successful implementation of PM projects by analysts —also in the case of less experienced, novice analysts. To this end, I pursue two main research goals (RGs):

- **RG1: Contribute to a better understanding of the individual level of PM**, i.e., how and why analysts act during a process analysis.
- **RG2: Develop methodological guidance and software-based support for novice analysts** to assist analysts during selected tasks of process analyses.

Design science research principles suggest that a detailed understanding of a problem should be established prior to the development of new artifacts, such as support systems or methodologies [14]. Therefore, the two RGs are interdependent, with findings from the first goal informing the development of support mechanisms in the second.

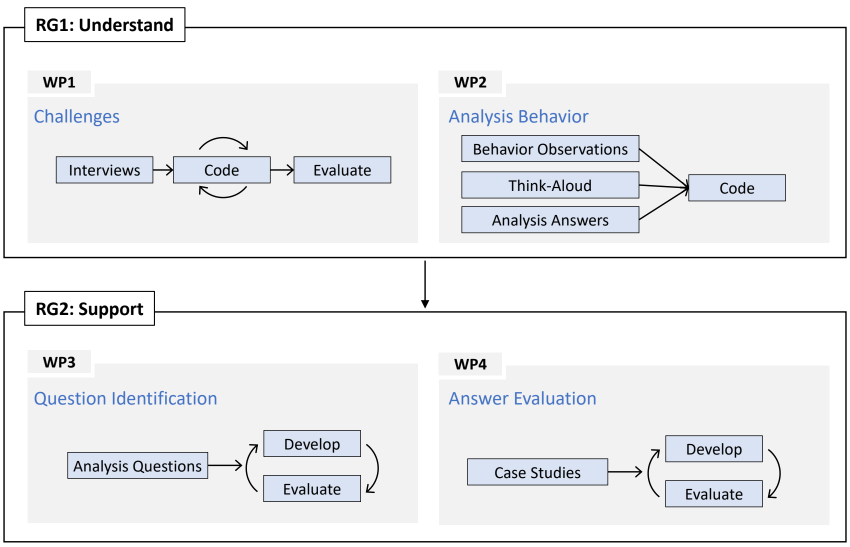

Figure 1: Proposed work packages

Figure 1 provides an overview of the proposed work packages (WPs) and their methodical implementation. To achieve RG1, I will mainly utilize the grounded theory research method [15] and analyze empirical data (e.g., interviews, behavior observations) to gain insights into analysts' challenges (WP1) and work practices (WP2). Building on the findings from WP1 and WP2, I aim to develop tools and methodological guidance (RG2). Based on the findings from WP1, which is already completed, the considerations outlined in Sect. 1, and a review of related work, I identified two areas in which support for novice analysts is lacking. First, WP3 focuses on question identification to support analysts to scope their analysis more effectively and generate relevant insights. Second, WP4 focuses on answer evaluation. Both WP3 and WP4 will follow design science research guidelines [14].

In the remainder, each WP is outlined in more detail, including its motivation, relevant related work, its proposed methodological implementation, and its status.

3. Challenges for Process Mining Analysts (WP1)

Design science research advocates that tools should be designed based on well-understood problems that address real challenges [14]. Therefore, WP1 contributed to defining the problem space by identifying the actual challenges faced by PM analysts. Three research questions (RQs) were raised and answered:

1. What are the challenges perceived by individual process analysts during a PM project?
2. Do the discovered PM challenges differ in their relevancy and in the extent to which experienced process analysts are able to solve them in practice?
3. What are mitigation strategies applied in practice to overcome these challenges?

Previously, challenges in PM have been analyzed from an organizational viewpoint [16], or reported rather fragmented in case studies [17, 18, 19]. More detailed studies focusing on the challenges of analysts based on empirical data have only been conducted in related fields, as by Wongsuphasawat et al. [20], focusing on exploratory data analysis.

Therefore, we studied the challenges that PM analysts experience during their work practice. In [21], we answered RQ 1 by following grounded theory [15]. The iterative coding approach of interviews with analysts resulted in the identification of 23 challenges. Subsequently, we verified the existence and relevancy of the challenges in an online survey and captured strategies applied by more experienced analysts to mitigate or overcome them (RQs 2-3). Results are published in [22]. Our findings revealed that especially the formulation of questions (cf. WP3) and the identification of answers and conclusions based on analysis results (cf. WP4) represent significant challenges for individual analysts and remain largely unresolved in many PM projects in practice [22].

4. Analysis Behavior (WP2)

Analysts perform sequences of operations and reasoning steps during process analysis, such as data manipulations, creation and interpretation of representations, formulation and testing of hypotheses, etc. [23, 24, 25]. Existing studies suggest that the *analysis behavior*[1] is rather analyst-specific and depends on factors like domain knowledge and prior experience [26]. It can be observed that the analysis behavior for the same task differs across analysts [24, 25]. However, based on my knowledge, a holistic understanding of which analysis operations are applied in which situations and what constitutes an effective process analysis remains unknown. A better understanding of these factors could make a decisive contribution and inform the development of support tools for novice analysts [27]. Therefore WP2 focuses on the analysis phase of PM projects and aims to answer the following RQs:

[1] With the term *Analysis Behavior*, I am referring to the observable sequence of analysis operations, including leveraged tool functions, that occur when an analyst conducts a PM analysis.

1. What is the behavior of analysts during the mining and analysis phase?
2. Why do analysts adopt a particular behavior and do they encounter difficulties with it?
3. How can effective and less effective analysis behavior be distinguished?

To answer the RQs, I will analyze screen recordings of the analysis process, including behavior observations, concurrent think-aloud data, and answers. This data has been captured during a large study with 40 analysts, as part of the ProMiSE project [28]. Initial insights into the data suggest that its holistic analysis is not trivial. To answer the RQs, I will therefore focus on the usage and effectiveness of visual elements during the analysis process, such as directly-follows-graphs, charts, variant visualizations, etc. It is undeniable that visualizations play a central role in a majority of the analyses of event data [8, 29] and answers to the formulated RQs can provide important input for tool development in this regard.

The work on WP2 has been started with the review, annotation, and coding [15] of the multi-modal data. The final round of coding and the interpretation of results are outstanding.

5. Question Formulation (WP3)

Multiple factors are influential to the analysis phase in PM projects [26]. One aspect is the question (task) that the analyst aims to answer. However, support for designing analysis questions is rare. Previously, Zerbato et al. [30] raised awareness of the problem by providing an overview of the diverse ways analysis questions may be derived. Additionally, Ullrich and Lata [31] confirmed that initial questions can aid non-technical users in approaching a PM analysis. Based on their experience, they formulated a set of *standard* questions process analysts might be interested in and provided analyses and visualizations that guide them in answering these questions. When questions are provided, Barbieri et al. [32] proposed how they could be classified according to their required analysis technique and translated into SQL statements that can be applied to a data set to automatically answer a question. Additionally, our research revealed that question design is perceived as one of the most significant challenges analysts are least able to overcome [22]. Consequently, rather than focusing on answering questions, WP3 is concerned with the fundamental work of guiding question clarification and formulation. In particular, the following RQs will be addressed:

1. What are the characteristics of questions posed during PM analyses?
2. How can the characteristics of PM analysis questions be arranged in a taxonomy?
3. How can a taxonomy and question bank be utilized to support question identification and formulation in PM projects?
4. Is the developed support usable and effective for deriving PM analysis questions?

I first aim to understand and categorize analysis questions following established taxonomy development guidelines [33, 34] (RQs 1-2). To this end, I already collected questions from literature and practice and developed the taxonomy together with my co-authors. The set of questions and the taxonomy provide an overview of the formulation of PM analysis questions, allow for the identification of blind spots, and provide a baseline to clarify and refine questions based on their underlying concepts and information needs. On top of the taxonomy and the set of questions, I aim to propose and evaluate an application that guides the structured use of both artifacts for the design of questions (RQs 3-4).

6. Answer Evaluation (WP4)

Similar to WP3, WP4 addresses one of the challenges identified in [22]. We found that analysts struggle with identifying clear answers to the initially raised questions and deriving conclusions from them. Koorn et al. [35] confirmed that there indeed exists a lack of rigorous evaluation measures for PM results, hindering analysts in identifying whether their answers are sufficient. They proposed qualitative validation strategies to verify the correctness and relevancy of results. Beerepoot et al. [36] provided a methodological solution to the problem, validated in the medical sector. However, both approaches involve project stakeholders and lack guidance for analysts to objectively assess their findings with respect to the analysis question. Thus, WP4 will focus on evaluating analysis outcomes, examining the relationship between analysis steps, their impact on the results, and investigating objective evaluation methods and their applicability to PM. Specifically, I raise the following RQs:

1. What are the core elements and their relationships that lead to an answer during the analysis process?
2. What are measures to objectively evaluate the quality of PM answers?
3. How can analysts be supported in the assessment of the quality of their PM answers?
4. Is the developed support usable and effective for assessing the quality of PM answers?

Several papers suggest that case studies offer valuable insights into work practices [23, 35]. Therefore, it seems promising to rely on a literature review of PM case studies to inform a conceptual model of the *analysis space*[2]. This model will hold important information about the elements that influence answers in PM projects (RQ 1). I will further explore whether quality indicators from related fields, as for example summarized in [37], can be adapted to PM (RQ 2). Based on the results of RQs 1-2, I will identify blind spots and develop support for assessing answer quality following design science principles [14] (RQ 3). While the exact form of this support is yet uncertain, the development of novel quality metrics and corresponding application guidelines are likely. I intend to evaluate the effectiveness of the developed artifacts in one or several case studies (RQ4), following the suggestions of Kitchenham et al. [38].

Initial work on WP4 has been started to verify its feasibility. The literature review is ongoing and confirms the feasibility of defining a conceptual model of the analysis space. The identification and assessment of quality metrics for the elements of the conceptual model (i.e., its components and links of components) remains open.

Acknowledgments

This doctoral project is funded by the Swiss National Science Foundation as part of the ProMiSE project under Grant No.: 200021_197032.

[2] I am referring to a conceptual domain encompassing all elements and activities relevant for the analysis phase of PM projects.

References

[1] W. Van Der Aalst, W. van der Aalst, Process Mining: Data Science in Action, Springer, 2016.

[2] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, Fundamentals of business process management, Springer, 2013.

[3] J. vom Brocke, W. van der Aalst, T. Grisold, W. Kremser, J. Mendling, B. Pentland, J. Recker, M. Roeglinger, M. Rosemann, B. Weber, Process science: the interdisciplinary study of continuous change, Available at SSRN 3916817 (2021).

[4] M. Kerremans, D. Sugden, N. Duffy, Magic Quadrant for Process Mining Platforms, Technical Report G00790664, Gartner, Inc., 2024. URL: https://www.gartner.com/doc/reprints?id=1-2HGG0P7J&ct=240502&st=sb.

[5] L. Reinkemeyer, Process mining in action, Springer, 2020.

[6] A. Mamudu, W. Bandara, M. T. Wynn, S. J. Leemans, A process mining success factors model, in: Int. Conf. on Business Process Management (BPM), Springer, 2022.

[7] J. vom Brocke, M. Jans, J. Mendling, H. A. Reijers, A five-level framework for research on process mining, Business & Information Systems Engineering (2021) 1–8.

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

[9] T. Grisold, J. Mendling, M. Otto, J. vom Brocke, Adoption, use and management of process mining in practice, Business Process Managagement Journal 27 (2021) 369–387.

[10] G. Kipping, D. Djurica, S. Franzoi, T. Grisold, L. Marcus, S. Schmid, J. v. Brocke, J. Mendling, M. Röglinger, How to leverage process mining in organizations-towards process mining capabilities, in: Int. Conf. on Business Process Management (BPM), Springer, 2022, pp. 40–46.

[11] M. L. v. 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), Springer, 2015, pp. 297–313.

[12] F. Zerbato, A. Burattin, H. Völzer, P. N. Becker, E. Boscaini, B. Weber, Supporting provenance and data awareness in exploratory process mining, in: Int. Conf. on Advanced Information Systems Engineering (CAiSE), Springer, 2023, pp. 454–470.

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

[14] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of Management Information Systems 24 (2007) 45–77.

[15] A. Strauss, J. M. Corbin, Grounded theory in practice, Sage, 1997.

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

[17] M. Eggert, J. Dyong, Applying process mining in small and medium sized it enterprises–

challenges and guidelines, in: Int. Conf. on Business Process Management (BPM), Springer, 2022, pp. 125–142.
[18] K. Smit, J. Mens, Process mining in the rail industry: A qualitative analysis of success factors and remaining challenges, in: BLED 2019 Proceedings, volume 25, 2019.
[19] R. Syed, S. J. Leemans, R. Eden, J. A. Buijs, Process mining adoption, in: Int. Conf. on Business Process Management (BPM), Springer, 2020, pp. 229–245.
[20] K. Wongsuphasawat, Y. Liu, J. Heer, Goals, process, and challenges of exploratory data analysis: An interview study, arXiv:1911.00568 (2019).
[21] L. Zimmermann, F. Zerbato, B. Weber, Process mining challenges perceived by analysts: An interview study, in: Int. Conf. on Business Process Modeling, Development and Support (BPMDS), Springer, 2022, pp. 3–17.
[22] L. Zimmermann, F. Zerbato, B. Weber, What makes life for process mining analysts difficult? a reflection of challenges, Software and Systems Modeling (2023) 1–29.
[23] 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.
[24] E. Sorokina, P. Soffer, I. Hadar, U. Leron, F. Zerbato, B. Weber, Pem4ppm: A cognitive perspective on the process of process mining, in: Int. Conf. on Business Process Management (BPM), Springer, 2023, pp. 465–481.
[25] F. Zerbato, P. Soffer, B. Weber, Initial insights into exploratory process mining practices, in: Int. Conf. on Business Process Management (BPM), Springer, 2021, pp. 145–161.
[26] F. Zerbato, P. Soffer, B. Weber, Process mining practices: evidence from interviews, in: Int. Conf. on Business Process Management (BPM), Springer, 2022, pp. 268–285.
[27] W. Maalej, M. Nayebi, T. Johann, G. Ruhe, Toward data-driven requirements engineering, IEEE software 33 (2015) 48–54.
[28] F. Zerbato, L. Zimmermann, H. Völzer, B. Weber, Promise: Process mining support for end-users, Proceedings of Int. Conf. on Advanced Information Systems Engineering (RPE@CAiSE) (2023).
[29] A. Yeshchenko, J. Mendling, A survey of approaches for event sequence analysis and visualization, Information Systems (2023) 102283.
[30] 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.
[31] C. Ulrich, T. Lata, Business miner: Process mining insights for business users, in: Proceedings of the ICPM Doctoral Consortium and Demo Track 2023, CEUR Workshop Proceedings; accepted for publication, CEUR-WS.org, 2023.
[32] L. Barbieri, E. Madeira, K. Stroeh, W. van der Aalst, A natural language querying interface for process mining, Journal of Intelligent Information Systems (2022) 1–30.
[33] D. Kundisch, J. Muntermann, A. M. Oberländer, D. Rau, M. Röglinger, T. Schoormann, D. Szopinski, An update for taxonomy designers: methodological guidance from information systems research, Business & Information Systems Engineering (2021) 1–19.
[34] R. C. Nickerson, U. Varshney, J. Muntermann, A method for taxonomy development and its application in information systems, European Journal of Information Systems 22 (2013) 336–359.

[35] J. J. Koorn, I. Beerepoot, V. S. Dani, X. Lu, I. Van de Weerd, H. Leopold, H. A. Reijers, Bringing rigor to the qualitative evaluation of process mining findings: an analysis and a proposal, in: Int. Conf. on Process Mining (ICPM), 2021, pp. 120–127.
[36] I. Beerepoot, N. Martin, J. Koorn, From insights to intel: evaluating process mining insights with healthcare professionals, in: Hawaii Int. Conf. on System Sciences 2023 (HICSS-56), volume 3, 2023.
[37] M. Behrisch, M. Blumenschein, N. W. Kim, L. Shao, M. El-Assady, J. Fuchs, D. Seebacher, A. Diehl, U. Brandes, H. Pfister, et al., Quality metrics for information visualization, in: Computer Graphics Forum, volume 37, Wiley Online Library, 2018, pp. 625–662.
[38] B. Kitchenham, L. Pickard, S. L. Pfleeger, Case studies for method and tool evaluation, IEEE software 12 (1995) 52–62.