Process Mining and Visual Analytics

1st Han van der Aa
University of Vienna
Vienna, Austria
han.van.der.aa@univie.ac.at

2nd Claudio Di Ciccio
Utrecht University
Utrecht, Netherlands
c.diciccio@uu.nl

3rd Djordje Djurica
Wirtschaftsuniversität Wien
Vienna, Austria
djordje.djurica@wu.ac.at

4th Kathrin Figl
University of Innsbruck
Innsbruck, Austria
kathrin.figl@uibk.ac.at

5th Jan Mendling
Humboldt-Universität zu Berlin
Berlin, Germany
jan.mendling@hu-berlin.de

6th Luise Pufahl
Technical University of Munich
Heilbronn, Germany
luise.pufahl@tum.de

7th Jana-Rebecca Rehse
University of Mannheim
Mannheim, Germany
rehse@uni-mannheim.de

8th Maxim Vidgof
Wirtschaftsuniversität Wien
Vienna, Austria
maxim.vidgof@wu.ac.at

***Abstract*—Process mining and visual analytics have developed independent from each other over a longer period of time, even though they focus on similar research questions and they work with event sequence data. The authors took part in a Mini-Dagstuhl Seminar held at the International Conference on Business Informatics 2024 in Vienna. This paper summarizes position statements of the authors on how process mining research and visual analytics complement each other.**

***Index Terms*—Process Mining; Visual Analytics; Event Logs; Event Sequence Data; Visualization**

I. Position Statement of Han van der Aa

Process mining focuses on the analysis of data recorded from business processes, striving to provide helpful insights to users (e.g., analysts, workers, or managers) with respect to how well the process is currently running, where its bottlenecks are, and where improvement potential lies. Most research in process mining focuses on the development of algorithms that aim to provide such insights, whereas little work has been conducted on actually assessing how useful the results provided by these algorithms are for the users that need to interpret and make decisions based on them.

Given that process mining thus aims to present results that should be beneficial to users, there is a clear need to consider how these results can best be presented to maximize their usefulness. This, thus, connects to the field of visual analytics, which provides means to depict algorithmic results through advanced visualizations. Although these fields have interacted little so far, important strides are being made in recent years to bridge this gap, including this symposium.

In this context, I see particular potential to consider the role that natural language can play to enhance the effectiveness of process mining results. From a representational point of view, this includes the use of text to complement visual representations (e.g., process models) [1] and using text to replace less-intuitive parts of visualizations [2]. Furthermore, given the importance that text plays within visualizations of process mining results (e.g., for labels of process elements) [3], it is highly promising to consider how this can be accounted for in interactive visualizations, such as those involving aggregations of model elements.

II. Position Statement of Claudio Di Ciccio

Process mining is an inherently explorative task [4]. Particularly during the analysis phase, the user is engaged with elaboration, comparison and subsequent refinement of findings. As a consequence, process mining endeavors require multiple interactive iterations.

A number of requirements thus emerge. Let us focus on three of those, among others. First, the capability of singling out specific fragments of process behavior (like short- and long-term dependencies) is necessary to foster explainability of the results. Second, the inspection of process logs from multiple perspectives beyond control flow is called for, to capture the interrelations between data determining the evidenced executions. Third, the user should be able to navigate from the data to extracted artifacts and back, to review, filter, and calibrate the results. The usual process mining feedback loop of feeding algorithms with event logs and get an end-to-end activity-centric process model back may fail to aid in all of the three above aspects. Rule-based declarative specifications can help with the first requirement, but they bear an inherent cognitive complexity making them hard to be fully captured by the involved users [5]. Also, effectively representing multi-perspective aspects of a process (i.e., dealing with more data attributes than the sole task name labeling events) remains a still pending problem [6]. Linking the mined process representations back to the data they originated from is an open challenge on its own as well [7].

Visual analytics can significantly contribute to the evolution of process mining thanks to the body of knowledge brought to support analytical reasoning through interactive visual interfaces [8]. The special attention paid to the user perspective in its methodologies, the mantra of manipulating data and visualizations to alter one another accordingly [9], the extensive studies on the representation of complex phenomena [8] are all factors that tackle the aforementioned challenges. In a cross-fertilizing setting, process mining can endow visual analytics with a stream of techniques for the automated generation of process-oriented representations of system dynamics, thus leading to a potential impact for both disciplines [10].

Pre-print copy of the manuscript published by IEEE (available at `ieeexplore.ieee.org`)
identified by DOI: `10.1109/CBI62504.2024.00039`

III. Position Statement of Djordje Djurica

Process mining research has traditionally focused on developing and refining algorithms for automatic discovery and conformance checking. These algorithms are typically evaluated based on their effectiveness, often using metrics such as precision and recall against a gold standard [11]. However, recently, the field has expanded to include user studies on the utilization of process mining tools and their generated visual representations [12].

Despite this progress, there remains a significant gap. The conceptual modeling community has already recognized that practitioners frequently combine different visual representations to achieve a holistic view of complex real-world domains [13]. This approach could also be highly relevant for the process mining community, as process analysts are using both Directly-Follow Graphs (DFGs) and other representations available on the dashboards of process mining tools during the analysis tasks. However, across different visual representations there might be some overlapping information. Prior research demonstrates that too little overlap can make integration of multiple representations difficult for readers, but too much overlap leaves too little cognitive capacity available for identifying relevant information effectively, in turn decreasing the capacity of readers to integrate and reason about the information presented [14].

Therefore, it is important to explore how much overlapping information is sufficient for the optimal performance of the users and how can this information be effectively presented across visualizations of process mining tools. Specifically, combining standard process mining visual representations, such as DFGs, with representations adapted from information visualization community conveying relevant information about an observed process, as outlined in [12], can significantly advance this field of research. This approach can potentially help process analysts to better integrate and process information, leading to a more comprehensive understanding of the observed process and their improved performance during process analysis tasks.

IV. Position Statement of Kathrin Figl

Visualizations play a crucial role in enhancing different levels of process mining: descriptive, predictive, and prescriptive process mining. Their impact on our perception of process-specific phenomena varies depending on the application area. In descriptive process mining, visualizations can help in better understanding the underlying data and how the process operates (see, e.g. [11]). For instance, visualizations can illustrate the differences between various process variants or, in the context of conformance checking (see, e.g. [15]), highlight discrepancies between planned and actual processes. Open research questions focus, for example, on how best to represent time or resource consumption to make these critical dimensions easier to understand (see, e.g. [12]). Research on visualizations in the context of process mining should not only aim to improve process understanding but also extend our knowledge of how visualizations influence process-related decision-making processes and potential biases that might occur. For instance, when deviations from an ideal process model are visualized, the focus of the reader can be directed either towards the current status quo or the desired state, influencing process managers' openness to changes and improvements. Visualizations can shape how we perceive reality and may induce an anchoring bias.

In predictive process mining, visualizations can help to understand forecasts of future process behaviors and outcomes and may enable stakeholders to see probable future scenarios, such as potential problems or bottlenecks. Prescriptive process mining goes a step further by not only predicting future outcomes but also recommending actions to achieve desired results, such as certain interventions to prevent undesirable outcomes. Visualizations in this context are vital for illustrating the potential impact of different interventions, helping users to understand the trade-offs and benefits of each possible action. In this context, visualizations can also play a significant role in improving the dynamics of human-algorithm interaction (see, e.g. [16]). They may help to overcome algorithm aversion and human biases that impede trust in algorithmic process recommendations by making these recommendations more transparent and trustworthy.

In conclusion, by leveraging the power of visualizations, organizations can gain deeper insights into their processes and make more informed, effective decisions. Future research should focus on how visualizations can be optimized to enhance understanding, support decision-making, and address potential biases in process mining.

V. Position Statement of Jan Mendling and Maxim Vidgof

Process mining and visual analytics have some common ancestors in the 1970s. Shneiderman, one of the authors of the Nassi-Shneiderman diagrams, started his career with a contribution on flowcharts [17]. Later his interest developed in a direction of what eventually became visual analytics. His work on flowcharts also found fertile reflection in the research field of business process management, where process models played an important role since the 1990s [18], a.o. for workflow automation [19], [20]. The first process mining techniques were then defined based on workflow concepts [21], and eventually developed into a field at the intersection of business process management and data science [22].

Process mining and visual analytics largely developed with little mutual awareness and interaction. Partially, the two fields developed their own key concepts and terminology. While process mining takes so-called *event logs* as a starting point [22], visual analytics typically refers to the same data as *event sequence data* or *time-oriented data* more generally [23]. There have been several calls for a better integration of process mining and visual analytics, e.g. [24], [25]. Until recently, there have been few contributions that build on ideas from both fields in an integrated manner, a.o. [26], [27].

Some initiatives have contributed to an increased awareness and knowledge exchange between process mining and visual analytics researchers. The keynote by Miksch at the International Conference on Process Mining has been a milestone in this direction [28] as much as the Dagstuhl seminar on Human in the (Process) Mines organized by Di Ciccio, Miksch, Soffer, and Weber [10]. These are promising developments which will help to further advance process mining techniques and stimulate visual analytics with input from the huge application field of business process management.

VI. Position Statement of Luise Pufahl and Jana-Rebecca Rehse

For intricate data analyses like process mining, the availability of high-quality visualizations is fundamental. They allow humans to quickly access new information, efficiently draw conclusions, and eventually make better decisions for their organization. Hence, visualizations make complex analysis results approachable and actionable, also for non-expert users. Therefore, high-quality visualizations are of paramount importance for any process mining application: In communicating the results of a process mining analysis to the intended audience, they support the users in gaining insights on their process and hence recognizing the true value of process mining.

Although multiple authors have stressed the need for developing novel visualization techniques for process mining [25], [29], [30], this topic has not so far not been widely considered in research. In fact, many commercial process mining vendors have developed their own visualizations for process mining analyses, but these visualizations differ widely, both in the information shown and the visualization idioms used [15]. Due to the absence of empirical studies, we know very little about the effectiveness of those visualizations or the preferences of the users. However, judging from our preliminary research [15] and our conversations with visualization experts, there is much room for improvement.

For those reasons, process mining can only benefit from a close integration with visual analytics research. At the same time, visual analytics can also benefit from exploring the domain of process mining because it offers some interesting (research) challenges, such as the inclusion of a time perspective, multidimensional data, and the consideration of both external knowledge and data. Furthermore, visual analytics research could apply process mining techniques to observe dependencies between visualization tasks recorded in experimental settings or derived from literature or discover typical paths for such tasks. Therefore, collaborations between the two fields can benefit both research communities—an excellent starting point for a better connection.

References

[1] H. van der Aa, H. Leopold, and H. A. Reijers, "Comparing textual descriptions to process models–the automatic detection of inconsistencies," *Inf. Syst.*, vol. 64, pp. 447–460, 2017.

[2] K. Figl and J. Recker, "Exploring cognitive style and task-specific preferences for process representations," *Requir. Eng.*, vol. 21, pp. 63–85, 2016.

[3] F. Pittke, H. Leopold, and J. Mendling, "Automatic detection and resolution of lexical ambiguity in process models," *IEEE Trans. Software Eng.*, vol. 41, no. 6, pp. 526–544, 2015.

[4] F. Zerbato, P. Soffer, and B. Weber, "Process mining practices: Evidence from interviews," in *BPM*. Springer, 2022, pp. 268–285.

[5] K. Figl, C. Di Ciccio, and H. A. Reijers, "Do declarative process models help to reduce cognitive biases related to business rules?" in *ER*. Springer, 2020, pp. 119–133.

[6] D. Fahland, "Multi-dimensional process analysis," in *BPM*. Springer, 2022, pp. 27–33.

[7] F. Zerbato, A. Burattin, H. Völzer, P. N. Becker, E. Boscaini, and B. Weber, "Supporting provenance and data awareness in exploratory process mining," in *CAiSE*. Springer, 2023, pp. 454–470.

[8] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," in *Information Visualization*. Springer, 2008, pp. 154–175.

[9] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *VL*. IEEE Computer Society, 1996, pp. 336–343.

[10] C. Di Ciccio, S. Miksch, P. Soffer, B. Weber, and G. Meroni, "Human in the (process) mines (dagstuhl seminar 23271)," 2024.

[11] J. Mendling, D. Djurica, and M. Malinova, "Cognitive effectiveness of representations for process mining," in *BPM*. Springer, 2021, pp. 17–22.

[12] A. Yeshchenko and J. Mendling, "A survey of approaches for event sequence analysis and visualization," *Inf. Syst.*, p. 102283, 2023.

[13] B. Dobing and J. Parsons, "Dimensions of uml diagram use: a survey of practitioners," *J. Database Manag.*, vol. 19, no. 1, pp. 1–18, 2008.

[14] M. Jabbari, J. Recker, P. Green, and K. Werder, "How do individuals understand multiple conceptual modeling scripts?" *J. Assoc. Inf. Syst.*, vol. 23, no. 4, pp. 1037–1070, 2022.

[15] J. Rehse, L. Pufahl, M. Grohs, and L. Klein, "Process mining meets visual analytics: The case of conformance checking," in *HICSS*. ScholarSpace, 2023, pp. 5452–5461.

[16] J.-L. Leemets, K. Kubrak, F. Milani, and A. Nolte, "Persuasive visual presentation of prescriptive business processes," in *Research Challenges in Information Science: Information Science and the Connected World*. Springer Nature Switzerland, 2023, Conference Proceedings, pp. 398–414.

[17] I. Nassi and B. Shneiderman, "Flowchart techniques for structured programming," *ACM Sigplan Notices*, vol. 8, no. 8, pp. 12–26, 1973.

[18] A.-W. Scheer and A. Hars, "Extending data modeling to cover the whole enterprise," *Communications of the ACM*, vol. 35, no. 9, pp. 166–ff, 1992.

[19] S. Jablonski, "On the complementarity of workflow management and business process modeling," *ACM SIGOIS Bulletin*, vol. 16, no. 1, pp. 33–38, 1995.

[20] V. Gruhn, "Business process modeling and workflow management," *Int. J. Cooperative Inf. Syst.*, vol. 4, no. 02n03, pp. 145–164, 1995.

[21] A. J. M. M. Weijters and W. M. P. van der Aalst, "Process mining: discovering workflow models from event-based data," in *Belgium-Netherlands Conf. on Artificial Intelligence*. Citeseer, 2001.

[22] W. M. P. van der Aalst, *Data science in action*, 2nd ed. Springer, 2016.

[23] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, "Visualization of time-oriented data," 2023.

[24] S. Kriglstein, M. Pohl, S. Rinderle-Ma, and M. Stallinger, "Visual analytics in process mining: Classification of process mining techniques." in *EuroVA@ EuroVis*, 2016, pp. 43–47.

[25] T. Gschwandtner, "Visual analytics meets process mining: challenges and opportunities," in *SIMPDA*. Springer, 2017, pp. 142–154.

[26] M. De Leoni, M. Adams, W. M. P. van der Aalst, and A. H. ter Hofstede, "Visual support for work assignment in process-aware information systems: Framework formalisation and implementation," *Decis. Support Syst.*, vol. 54, no. 1, pp. 345–361, 2012.

[27] A. Yeshchenko, C. Di Ciccio, J. Mendling, and A. Polyvyanyy, "Visual drift detection for event sequence data of business processes," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 8, pp. 3050–3068, 2021.

[28] S. Miksch, "Visual analytics meets process mining: challenges and opportunities," in *ICPM*. IEEE, 2021, pp. xiv–xiv.

[29] L. Garcia-Banuelos, N. Van Beest, M. Dumas, M. La Rosa, and W. Mertens, "Complete and interpretable conformance checking of business processes," *IEEE Trans. Software Eng.*, vol. 44, no. 3, pp. 262–290, 2017.

[30] C. Klinkmüller, R. Müller, and I. Weber, "Mining process mining practices: An exploratory characterization of information needs in process analytics," in *BPM*. Springer, 2019, pp. 322–337.

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org)
identified by DOI: 10.1109/CBI62504.2024.00039

This document is a pre-print copy of the manuscript
(Aa et al. 2024)
published by IEEE (available at ieeexplore.ieee.org).

The final version of the paper is identified by DOI: 10.1109/CBI62504.2024.00039

References

Aa, Han van der, Claudio Di Ciccio, Djordje Djurica, Kathrin Figl, Jan Mendling, Luise Pufahl, Jana-Rebecca Rehse, and Maxim Vidgof (2024). "Process Mining and Visual Analytics". In: *CBI*. Ed. by Peter Fettke, Hans Weigand, Ulrich Frank, and Christian Huemer. IEEE. ISBN: 979-8-3315-2909-3. DOI: 10.1109/CBI62504.2024.00039.

BibTeX

```
@InProceedings{   Aa.etal/CBI2024:ProcessMiningVisualAnalytics,
  author        = {van der Aa, Han and Di Ciccio, Claudio and Djurica,
                  Djordje and Figl, Kathrin and Mendling, Jan and Pufahl,
                  Luise and Rehse, Jana-Rebecca and Vidgof, Maxim},
  booktitle     = {CBI},
  title         = {Process Mining and Visual Analytics},
  year          = {2024},
  crossref      = {CBI2024},
  doi           = {10.1109/CBI62504.2024.00039},
  keywords      = {Process Mining; Visual Analytics; Event Logs; Event
                  Sequence Data; Visualization}
}
@Proceedings{     CBI2024,
  title         = {26th International Conference on Business Informatics,
                  {CBI} 2024, Vienna, Austria, 9--13 September 2024},
  year          = {2024},
  editor        = {Fettke, Peter and Weigand, Hans, and Frank, Ulrich and
                  Huemer, Christian},
  isbn          = {979-8-3315-2909-3},
  publisher     = {{IEEE}}
}
```