Process of Process Mining: A Multi-perspective Event Log

Francesca Zerbato[1,*], Lisa Zimmermann[2] and Barbara Weber[2]

[1]Eindhoven University of Technology, Eindhoven, The Netherlands
[2]University of St. Gallen, St. Gallen, Switzerland

Abstract
Process mining is a data-driven discipline for understanding and improving business processes. However, despite its uptake in practice, there is still limited understanding of how analysts interact with software tools and artifacts to discover insights. This resource paper introduces the ProMiSE event log, a log derived from video-recorded sessions with process mining analysts. Each trace captures the *process of process mining* of a single analyst from three complementary perspectives, each describing the analysis process from a particular angle and several levels of abstraction. By making this event log publicly available, we aim to support research on analyst behavior and on methods for handling multi-granular and multi-perspective event data.

Keywords
Process mining, User behavior, Event log, Multi-granular data, Multi-perspective data

1. Introduction

Over the past two decades, process mining (PM) has established itself as a data-driven discipline for understanding and improving business processes [1]. While technical research has focused on designing algorithms and methods for process discovery, conformance checking, and predictive PM, empirical studies have investigated how process mining is actually conducted in practice, i.e., the *process of process mining* [2]. Understanding how analysts explore data using existing software tools and visualizations is essential for improving analysis guidance, tool design, and methodological support.

Despite being a complex, cognitive process, the process of process mining can be seen as a series of interactions between an analyst, the process mining tools they use, and the artifacts they explore, such as visualizations or documents. To analyze the behavior of process analysts through these interactions, we have recorded video sessions of 40 PM analysts performing an analysis task. From such sessions, we have derived an event log through several rounds of qualitative coding [3], which allowed us to segment the video recordings into meaningful events based on what we observed in the video. The events in these traces describe the analysis process along three main perspectives that capture (i) what the analyst was focusing on, (ii) the visualizations they looked at, and (iii) the tool functions they used.

The **ProMiSE event log** in this resource, named after the related research project, includes 5724 events organized into 40 partially ordered traces, each one corresponding to an analysis session. This log provides a structured view of how process analysts interact with tools and artifacts during a PM analysis task. In addition to capturing the analysis from multiple perspectives, a unique trait of this event log is its multi-granular nature: each event name embeds multiple levels of abstraction, allowing one to choose what abstraction level to use for activities during the analysis. Furthermore, since each trace captures the behavior of one analyst, the log exhibits high variability, similar to event logs derived from user interactions, e.g., through active window tracking [4], or sensor recordings [5].

By making this resource publicly available, we aim to encourage further research on (1) exploring analyst behavior through the study of interaction patterns (some analyses we have conducted are presented in Sect. 3); (2) evaluating techniques that handle unstructured and multi-granular event data

Proceedings of the BPM 2026 Best Dissertation Award, Doctoral Consortium, and Demonstration & Resources Forum (BPM 2026), Toronto, Canada, September 27th to October 2nd, 2026.
*Corresponding author.
f.zerbato@tue.nl (F. Zerbato); lisa.zimmermann@unisg.ch (L. Zimmermann); barbara.weber@unisg.ch (B. Weber)
0000-0001-7797-4602 (F. Zerbato); 0000-0002-6149-7060 (L. Zimmermann); 0000-0002-6004-4860 (B. Weber)

that may not meet the assumptions of conventional event logs; (3) developing techniques that discover multi-perspective processes that combine both totally and partially ordered event sequences.

2. Description of the Resource

The ProMiSE event log comprises 40 traces and 5724 events derived from 26.6 hours of video recordings. The full event log and all the files belonging to this resource are available on Zenodo https://zenodo.org/records/20995525 under the Creative Commons Attribution 4.0 International license (CC BY 4.0)[1].

Data Collection and Coding. The data was collected in 2021 during video-recorded meetings with process analysts from industry and academia, who were asked to perform an analysis of the Road Traffic Fine Management event log [6] while doing concurrent think-aloud. The analysis focused on finding scenarios in which traffic fines are not paid and, if possible, identify root causes of non-payment. The analysts could use and combine any tools among bupaR, Celonis, Disco, PM4Py, and ProM and no constraints were placed on the session duration. More details on the data collection are described in [7].

On average, each session lasted 40.1 minutes (median = 37 min; sd = 11.6 minutes). To derive the event sequences that make up the traces in our log, we used a qualitative coding approach to segment the video recordings into discrete events, each defined by a start time, an end time, and a name. In our data, events are time-stamped intervals where the name describes the analyst's behavior from a particular perspective at several levels of granularity. Figure 1 illustrates the coding process followed to generate the traces. We worked in several rounds as a team of four researchers, segmenting the video recordings with MAXQDA[2] and using think-aloud protocols to disambiguate the behavior we observed in the video, which was essential to identify the focus of the analysis. To ensure consistency, we iteratively refined the event boundaries and names, resolving any ambiguous cases through collaborative discussions. During the coding process, we also ensured to have at least 3 ms of gap between consecutive events to avoid them being treated as concurrent and did not code idle moments.

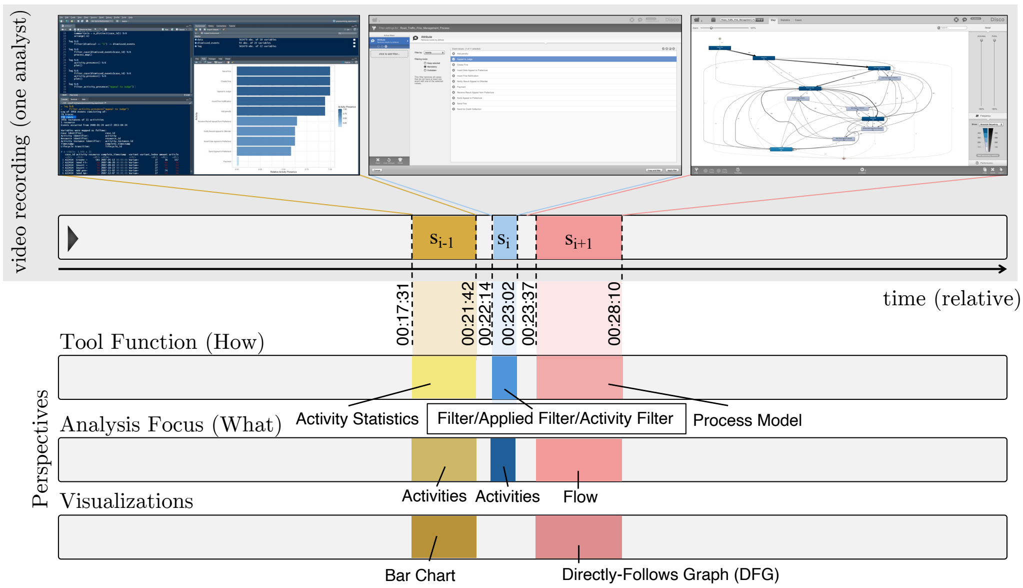

Figure 1: Event sequence (trace) generation from video recordings through qualitative coding. Video recordings were segmented using MAXQDA. For each video segment s_i, i.e., the same time-stamped interval, we derived one or more events belonging to different perspectives.

[1]https://creativecommons.org/licenses/by/4.0/deed.en
[2]https://www.maxqda.com

Data Structure and Characteristics. Table 1 provides an overview of the event log files included in this resource. The complete event log, available in *ProMiSE_EventLog_complete.csv*, contains 5724 events organized into 40 traces, one per analysis session, as we described above.

Table 1
Main event log files included in the resource and related event count statistics.

File name	Type	#Events	Trace Length		
			Mean	Median	SD
ProMiSE_EventLog_ToolFunction.csv	Tool Function (How)	2596	64.9	61.0	24.0
ProMiSE_EventLog_AnalysisFocus.csv	Analysis Focus (What)	2356	58.9	55.5	25.5
ProMiSE_EventLog_Visualizations.csv	Visualizations	772	19.3	19.5	9.72
ProMiSE_EventLog_complete.csv	All	5724	143.0	138.0	54.4

The events in the complete event log represent a particular perspective of the analysis process among: (1) *Tool Function (How)* that describes the function used in the process mining tool in a tool-agnostic manner; (2) *Analysis Focus (What)* that indicates what aspect of the process or artifact the analyst was looking at; and (3) *Visualizations* that tracks which visualizations was used.

These three perspectives may correspond to different segmentations of the video recordings. In some cases, the same video segment corresponds to multiple events, each belonging to a different perspective. For example, in Figure 1, from video segment s_{i-1} we derived the events "Activity Statistics" for *Tool Function (How)*, "Activities" for *Analysis Focus (What)*, and "Bar Chart" for *Visualizations*, all having the same start and end timestamps. In other cases, from one video segment we derived only events belonging to one or two of these perspectives, e.g., there is no *Visualizations* event matching s_i. Because of this, the resulting multi-perspective sequence is partially ordered and allows researchers to study the analysis process of one analyst from multiple angles simultaneously. The resource also contains the event logs obtained by filtering one perspective (Type) at a time (cf. Table 1), which corresponds to taking each one of these perspectives individually. These traces can be used by researchers to focus on specific aspects, e.g., the use of visualizations. The traces in 'ProMiSE_EventLog_ToolFunction.csv' and 'ProMiSE_EventLog_Visualizations.csv' are totally ordered, while those in 'ProMiSE_EventLog_AnalysisFocus.csv' are partially ordered, since analysts could focus on multiple aspects at once.

Figure 2 shows a screenshot of the Variant Explorer in Cortado [8], where the high-level variant visualization shows overlapping events between perspectives having the same start and end timestamps.

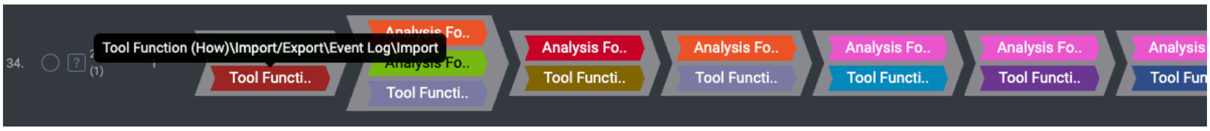

Figure 2: Excerpt of one trace in the Variant Explorer of Cortado, showing overlaps between events in the *Tool Function* and *Analysis Focus* perspectives.

The ProMiSE event log preserves multiple granularity levels in the event names. Figure 1 shows an example of multi-granular name in perspective *Tool Function* covering three levels of abstraction for the filter applied in s_i: "Filter/Applied Filter/Activity Filter." We identified up to four levels of granularity (L1–L4) in our event names. All 5724 events in the complete log have a name defined at L1, 5144 have up to L2, 1492 have up to L3, and 461 have up to L4. In addition to being part of the event name itself (as shown in Figures 1 and 2), they are included as separate event attributes to ease their adjustment during the analysis. The complete event log includes 90 different event names, which can be adjusted to 20 if L1 is chosen as the 'Activity' column. The file 'EventNames.pdf' provides a description of all the event names used in the complete event log.

An excerpt of the complete event log is shown in Table 2. The **ID** identifies the analyst session. **Event_name** is the multi-granular event name, divided into four granularity levels **L1** to **L4**. Each of these columns can be chosen as the 'Activity' column for PM analysis. The **Type** indicates the perspective, attributes **Start** and **End** are relative timestamps from the start of the session, and **UsedTool**

Table 2
Excerpt of the file 'ProMiSE_EventLog_complete.csv'.

ID	Event_name	Type	L1	L2	L3	L4	Start	End	UsedTool
1	Tool Function (How)\Visualization\Process Model\DFG\Interactive DFG	Tool Function (How)	Visualization	Process Model	DFG	Interactive DFG	00:08:14	00:08:49	Celonis
1	Analysis Focus (What)\Analysis Focus\Flow	Analysis Focus (What)	Analysis Focus	Flow			00:08:14	00:08:49	
1	Visualizations\Process Model\DFG	Visualizations	Process Model	DFG			00:08:14	00:08:49	
1	Tool Function (How)\PDF Reader	Tool Function (How)	PDF Reader				00:08:50	00:09:14	

reports for *ToolFunction* events the PM tool used for the session, if any. This column is not present in the event logs capturing the *Analysis Focus* and *Visualization* perspectives. In the repository, we also provide the .xes file of the complete event log, with 'ID' as case identifier and 'Event_name' as activity.

3. Preliminary Analysis

We have used the event logs presented in Section 2 in two different studies, where we combined them with think-aloud protocols to explore specific aspects of analysts' behavior.

Use of Tool Functions. In our recent work [9], we have investigated how analysts discover findings, i.e., conclusive insights. We selected 22 traces from 'ProMiSE_EventLog_ToolFunction.csv' and combined them with information about when a finding was discovered, which we derived from the think-aloud. For our analysis, we adjusted the granularity level to focus on 15 tool functions of interest, selecting event names from L1, L2, and L3. We then explored which *Tool Functions* are most commonly used during a PM analysis (based on our task) and which ones lead to findings, using statistical analysis and sequential pattern mining [10]. We found that process mining analysts spent more than half their time using interactive models like Directly-Follows Graphs (DFGs) or inspecting variants and reviewing documentation. However, customizable tool functions, such as plots, charts, and custom OLAP tables, led to several high-quality findings, even if their overall use was limited. Our sequential pattern mining analysis revealed that analysts often discovered insights by alternating between process documentation and visualizations, but used filters and custom data transformations to validate them.

Use of Visualizations. In [11], we focused on analyzing how PM analysts use visualizations. For our analysis, we selected 33 traces from 'ProMiSE_EventLog_complete.csv' and combined them with think-aloud protocols. We slightly adapted the event names in the visualization perspective and reorganized the events of L1, L2, and L3. We then analyzed which visualization types were used and when, and for which analytical purposes (analysis focus) they were used. We found that visualizations played a central role in PM analysis, accounting for more than one third of the total analysis time. By statistically analyzing the overlaps between the *Visualizations* and *Analysis Focus* perspectives, we found that process models, such as process maps, were mainly used for understanding the process flow, activities, and KPIs. Sequence charts were used less frequently, but supported case- and variant-level reasoning on variants, single attributes, and case-level statistics. Chart-based visualizations, instead, complemented the analysis by supporting performance and statistics inspection. Over time, analysts started with less visualization use while familiarizing themselves with the task and documentation, then relied more on process maps, and only later they used sequence charts and other visualizations. We provide a figure combining the visualization and analysis focus perspectives (choosing L2 for both) in the repository.

4. Possible Uses and Innovation for the BPM Community

The ProMiSE event log can support research in the PM community in several ways. First, the event log provides a source of empirical evidence to validate assumptions regarding the behavior of process analysts within our data collection settings. Although our log was derived from a single analysis task, as explained in Section 2, it can be used to investigate interaction patterns involving specific tools, artifacts, and visualizations. Indeed, with the analyses presented in Section 3, we aimed to gain a broad understanding of analysts' behavior and, thus, we did not examine the use of specific tool functions (e.g., DFGs) nor focus on other "local" patterns. Using our data to further investigate these aspects could lead to identifying gaps in visualization and tool design. Second, the event log is highly unstructured (one trace corresponds to one trace variant) and, resembles the characteristics of UI logs where traces are long and may show repetitive patterns over time. This makes the event log suitable for validating techniques that can abstract such unstructured data. In particular, the multi-granular event names provide predefined, semantic abstraction levels that can be used for event abstraction during the analysis, e.g., in process discovery tasks. Finally, the multi-perspective nature of this log opens new challenges for PM analysis, as the traces contain a high number of overlapping events, something that PM techniques often struggle to handle. In our work, we have used interactive sequential pattern mining techniques [10] to explore this log from multiple perspectives, which can inspire researchers to develop new PM techniques capable of discovering multi-perspective processes.

Acknowledgments

This work was made possible thanks to the ProMiSE project, funded by the Swiss National Science Foundation under Grant No. 200021_197032.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] J. Vom Brocke, M. Jans, J. Mendling, H. A. Reijers, A five-level framework for research on process mining, Business & Information Systems Engineering 63 (2021) 483–490.
[2] E. Sorokina, P. Soffer, I. Hadar, U. Leron, F. Zerbato, B. Weber, PEM4PPM: A cognitive perspective on the process of process mining, in: Intl Conf on Business Process Management, Springer, 2023, pp. 465–481.
[3] M. B. Miles, A. M. Huberman, J. Saldana, Qualitative data analysis, sage, 2014.
[4] I. Beerepoot, A month's worth of labelled active window tracking data, in: CEUR Workshop Proceedings, volume 3758, CEUR WS, 2024, pp. 61–65.
[5] D. J. Cook, A. S. Crandall, B. L. Thomas, N. C. Krishnan, Casas: A smart home in a box, Computer 46 (2012) 62–69.
[6] M. de Leoni, F. Mannhardt, Road traffic fine management process, Eindhoven University of Technology, Dataset 284 (2015).
[7] L. Zimmermann, F. Zerbato, K. Vrotsou, B. Weber, A process-oriented approach to analyze analysts' use of visualizations: Revealing insights into the what, when, and how, in: Computer Graphics Forum, Wiley Online Library, 2025, p. e70104.
[8] D. Schuster, S. J. van Zelst, W. M. van der Aalst, Cortado—an interactive tool for data-driven process discovery and modeling, in: International Conference on Applications and Theory of Petri Nets and Concurrency, Springer, 2021, pp. 465–475.
[9] F. Zerbato, L. Zimmermann, K. Vrotsou, B. Weber, From analysis to findings: How do process mining analysts discover results?, Information Systems (2025) 102596.
[10] K. Vrotsou, A. Nordman, Exploratory visual sequence mining based on pattern-growth, IEEE transactions on visualization and computer graphics 25 (2018) 2597–2610.
[11] L. Zimmermann, F. Zerbato, B. Weber, Visualization use in process mining analysis, in: BPMDS, Springer, Cham, 2026, pp. 55–72.