Visual Analytics Meets Process Mining: Challenges and Opportunities

Silvia Miksch, *TU Wien, A-1040, Vienna, Austria*
Claudio Di Ciccio, *Utrecht University, 3584 CC Utrecht, The Netherlands*
Pnina Soffer, *University of Haifa, 31905 Haifa, Israel*
Barbara Weber, *University of St. Gallen, 9000 St. Gallen, Switzerland*

Abstract—Visual Analytics (VA) integrates the outstanding capabilities of humans in terms of visual information exploration with the enormous processing power of computers to form a powerful knowledge discovery environment. In other words, VA is the science of analytical reasoning facilitated by interactive interfaces, capturing the information discovery process while keeping humans in the loop. Process Mining (PM) is a data-driven and process centric approach that aims to extract information and knowledge from event logs to discover, monitor, and improve processes in various application domains. The combination of interactive visual data analysis and exploration with PM algorithms can make complex information structures more comprehensible and facilitate new insights. Yet, this combination remains largely unexplored. In this article, we illustrate the concepts of VA and PM, how their combination can support the extraction of more insights from complex event data, and elaborate on the challenges and opportunities for analyzing process data with VA methods and enhancing VA methods using PM techniques.

Visual Analytics (VA) has been defined as the science of analytical reasoning supported by interactive visual interfaces [1]. To this end, VA's goal is to assist human users in comprehending complex and multi-variate data. In order to achieve that, VA intertwines the capabilities of computers and humans by combining automated analysis with interactive visualization methods. Process mining (PM) sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand [2]. The objective of PM is to discover, monitor, and improve real processes by extracting information and knowledge from event logs (also known as "event data" and used interchangeably with "event sequences" in VA), readily available in today's information systems, and answering business and domain-related questions to support domain-specific goals. However, the combination of VA and PM is a largely uncharted territory. In light of these circumstances, Silvia Miksch presented an invited talk at the 3rd International Conference on Process Mining (ICPM) in 2021 and we organized a Dagstuhl Seminar entitled "Human in the (Process) Mines" [3] in 2023.

During that Dagstuhl seminar, we discussed the following questions: *How to incorporate knowledge into an interactive analysis process of event data? How to make sense of multi-faceted process data? What comprehension metrics can be used to evaluate visualizations for process mining tasks? How to select appropriate visualizations for object-centric and multi-dimensional process mining tasks? How to design visualizations for the task of conformance checking? How to support human-centred process mining by constructing a unified process mining and visual analytics task taxonomy?* [3]. This article contextualizes VA and PM to derive challenges and opportunities thereof.

Digital Object Identifier 10.1109/XXX.0000.0000000

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org) identified by DOI: 10.1109/MCG.2024.3456916

SIDEBAR: Process Mining

Process Mining (PM) is a rapidly growing research discipline blending algorithm design, automated reasoning, machine learning and data mining concepts with ideas taken from the broader field of business process management (BPM), especially process modeling and analysis [1]. PM explores and utilizes event data (typically recorded in *event logs*, or logs for short) and process models to support various PM tasks, such as the (automated) discovery of process models (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing models and logs), and process enhancement (i.e., extending or improving models using event logs) [1].

Figure 1 displays an example of typical process mining inputs and outputs. The event log pertains to the management of road traffic fines [2], an excerpt of which occurs at the top of the figure. The process begins with the issuing of a fine. The offender can then pay the fine (in its entirety or in parts). A penalty is added if the payment does not take place in due time. However, after being notified per post, the offender can appeal in a court or at a prefecture. In case of a successful appeal, the case ends. If the offender still refuses to pay, the fine is handed over to a credit collection agency. As it can be seen in Fig. 1, PM strives to excel at extracting and analyzing workflow structures based on data. The graph-like diagram to the right is a Directly-Follows Graph (DFG) [1], which depicts activities as box-shaped nodes on paths from the start to the end vertices, whereby arcs denote an observed direct sequence between activities in the event log. The diagram at the bottom is a Business Process Model and Notation (BPMN) diagram [3], again depicting activities along paths from the start to the end events, but with dedicated nodes to depict gateways (diamonds), branching possible executions into exclusive choices (cross), parallel flows (plus sign), all-inclusive joins (circles). Practitioners often use DFGs to have a glimpse of how activities follow one another as per the recorded executions. However, DFGs fail to provide a thorough representation of the workflow. Alternatively, process models like BPMN diagrams have a richer notation and can be used for capturing the causal relationships, choices and concurrency among the activities in the process. Though less intuitive, they can be instrumental to draw more reliable conclusions on how the process actually unfolds [1].

In fact, most PM approaches create artifacts that are tied to visual representations of the extracted process behavior. The vast majority of them adopt a procedural approach, depicting the possible execution paths from start to end. More recently, the declarative approach emerged as an alternative representation [4]. It provides the rules that process executions abide by rather than depicting the end-to-end activity sequences that do so. Languages such as Declare [5] and Dynamic Condition Response Graphs [6] belong to this category. However, both with procedural and declarative settings, the control flow of the process (namely, its sequence of activities) is the main view, whereas additional data and metrics that can be analyzed (e.g., time, resources, conformance, networks) are less emphasized in the resulting visualizations. Over the last decade, research efforts in the field of PM have mainly focused on the development of algorithms and approaches for specific PM tasks, addressing each one separately from a technical perspective. However, less attention has been given to supporting the humans (experts, practitioners, laypersons, etc.) involved in the PM processes.

PM aims to extract information from event logs and process models, which can often exhibit unexpected behavior and complex relationships. Therefore, before and during the application of PM algorithms, the analyst needs to investigate and understand the data and models at hand in order to decide which analysis methods might be appropriate. Event-based multivariate network visualization [7]–[9] and principles of knowledge-assisted VA could ease that process. To this end, the combination of visual data exploration with PM algorithms makes complex information structures more comprehensible and facilitates new insights. Moreover, such event data or activity traces often have data quality issues [10] and exhibit complex relations, which can lead to unexpected behavior. Consequently, it is important for analysts to thoroughly explore, investigate, and understand the data before and during the implementation of interactive analysis methods, such as PM algorithms, to determine which methods are appropriate and to be able to make sense of the output generated by the selected methods.

References

[1] W. van der Aalst, *Process Mining: Data Science in Action*, 2nd. Springer Publishing Company, Incorporated, 2016, ISBN: 3662498502. DOI: 10.5555/2948762.

[2] F. Mannhardt *et al.*, "Balanced multi-perspective checking of process conformance," *Computing*, vol. 98, no. 4, pp. 407–437, 2016. DOI: 10.1007/S00607-015-0441-1.

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org) identified by DOI: 10.1109/MCG.2024.3456916

CaseID	Activity	End timestamp	Resource	Amount	Article	Expense	NotificationType	PaymentAmount	TotalPaymentAmount	VehicleClass
N26018	Create Fine	03/06/2000	538	31.30	157	NULL	NULL	NULL	0	A
N26018	Send Fine	20/07/2000	NULL	NULL	NULL	11.41	NULL	NULL	NULL	NULL
N26018	Insert Fine Notification	29/08/2000	NULL	NULL	NULL	NULL	P	NULL	NULL	NULL
N26018	Add penalty	28/10/2000	NULL	62.59	NULL	NULL	NULL	NULL	NULL	NULL
N26018	Send for Credit Collection	10/04/2002	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
N61937	Create Fine	08/08/2004	559	33.60	157	NULL	NULL	NULL	0	A
N61937	Send Fine	17/12/2004	NULL	NULL	NULL	16.6	NULL	NULL	NULL	NULL
N61937	Insert Fine Notification	28/12/2004	NULL	NULL	NULL	NULL	P	NULL	NULL	NULL
N61937	Payment	28/12/2004	NULL	NULL	NULL	NULL	NULL	44.60	44.60	NULL
N61937	Add penalty	26/02/2005	NULL	68.77	NULL	NULL	NULL	NULL	NULL	NULL
N61937	Payment	21/03/2005	NULL	NULL	NULL	NULL	NULL	5.60	50.20	NULL

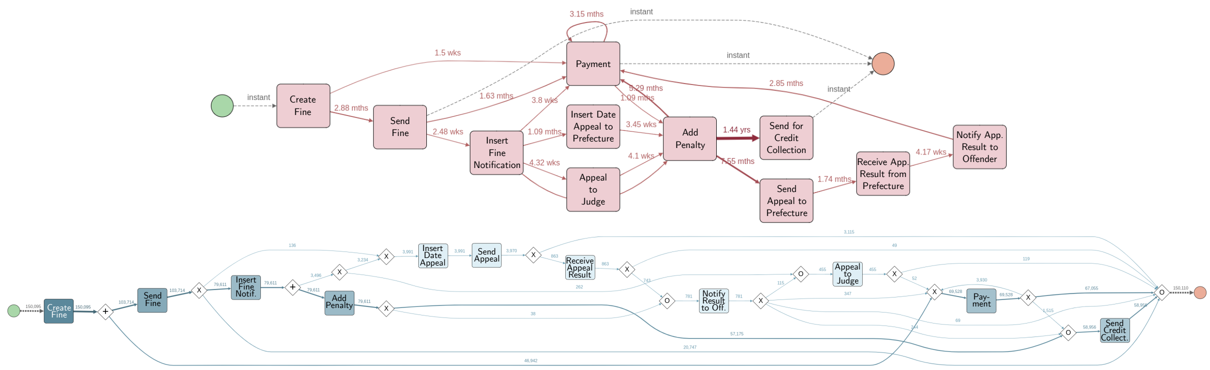

FIGURE 1. Examples of data and process visualizations used in PM, based on the road traffic fine event log. *Top*: An excerpt of the event log in a tabular format. *Middle*: A DFG mined from the log with arcs weighted by duration. *Bottom*: A BPMN diagram mined from the log with arcs and nodes weighted with case frequency. The diagrams were made by the authors via the Apromore PM suite. Activity labels are magnified and abbreviated for readability.

[3] G. Decker *et al.*, "The business process modeling notation," in *Modern Business Process Automation - YAWL and its Support Environment*, A. H. M. ter Hofstede *et al.*, Eds., Springer, 2010, pp. 347–368. DOI: 10.1007/978-3-642-03121-2_13.

[4] W. M. P. van der Aalst *et al.*, "Declarative workflows: Balancing between flexibility and support," *Computer Science - R&D*, vol. 23, no. 2, pp. 99–113, 2009. DOI: 10.1007/s00450-009-0057-9.

[5] M. Pesic *et al.*, "DECLARE: Full support for loosely-structured processes," in *EDOC*, 2007, pp. 287–300. DOI: 10.1109/EDOC.2007.14.

[6] T. T. Hildebrandt and R. R. Mukkamala, "Declarative event-based workflow as distributed dynamic condition response graphs," in *PLACES*, 2010, pp. 59–73. DOI: 10.4204/EPTCS.69.5.

[7] F. Beck *et al.*, "A taxonomy and survey of dynamic graph visualization," *Computer Graphics Forum*, vol. 36, no. 1, pp. 133–159, 2017. DOI: 10.1111/cgf.12791.

[8] C. Nobre *et al.*, "The state of the art in visualizing multivariate networks," *Computer Graphics Forum*, vol. 38, no. 3, pp. 807–832, 2019. DOI: 10.1111/cgf.13728.

[9] P. Simonetto *et al.*, "Event-based dynamic graph visualisation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 7, pp. 2373–2386, 2018. DOI: 10.1109/TVCG.2018.2886901.

[10] F. Zerbato *et al.*, "Supporting provenance and data awareness in exploratory process mining," in *Advanced Information Systems Engineering - 35th International Conference, CAiSE 2023*, M. Indulska *et al.*, Eds., ser. Lecture Notes in Computer Science, vol. 13901, Springer, 2023, pp. 454–470. DOI: 10.1007/978-3-031-34560-9_27.

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org) identified by DOI: 10.1109/MCG.2024.3456916

Open Challenges

Researchers have identified various challenges intertwining VA and PM [2], [4], [5]. Even more than ten years ago, van der Aalst et al. [2] mentioned in their Process Mining Manifesto the potential of VA to support PM. In the following, we are elaborating on the challenges of intertwining VA and PM to achieve more informed, traceable, trustworthy, and explainable investigations and decision-making processes. These challenges are not exhaustive, and there are other directions to explore. The following challenges are organized in accordance with the underlying workflow of decision-making, commencing with more general aspects and progressing to specific components.

Developing Mixed-Initiative Processes

A mixed-initiative process [6] is an approach in which both humans (also referred to as users) and systems can "take the initiative" and both contribute to the process. Such a mixed-initiative process could be utilized to enhance problem-solving and achieve a comprehensive understanding. Even though Horvitz [6] defined this concept in 1999, there are still unsolved issues regarding the tasks and roles of humans and systems as well as their interplay. In PM, the notion of mixed-initiative is not developed. Mostly, the initiative is by the human, who decides what kind of analysis to perform, and relies on the output generated by that analysis. In some cases, it is possible to explore the output by controlling the level of detail or by switching among different views and process perspectives that are available (e.g., [7]). However, such operations are still at the hands of the human. One research direction addressing the mixed-initiative is the concept of "guidance".

Providing Guidance

VA supports the information-discovery process from a data-dependent, task-specific, and user-oriented perspective. However, determining which VA methods are best suited for their specific data and tasks can be challenging for users, who are typically experts in their application domains but not in VA. Guidance is needed to assist users with the selection of appropriate visual means and interaction techniques, the utilization of analytical methods, as well as the configuration instantiation of these methods with suitable parameter settings and combinations thereof [8].

Guidance is derived from the user's knowledge and can take into account various inputs, including data, visualizations, interaction history, and domain knowledge. Guidance can be presented in different forms, such as visual cues and alternative options, to aid in data exploration, identification of interesting data nuggets and findings (e.g., anomalies, strange behaviors), and collection and grouping of insights to explore high-level hypotheses and gain new knowledge. Once a VA method and parameters have been selected, it is crucial to have guidance for exploring the data, identifying noteworthy findings (such as anomalous behaviors in event logs), and collecting and categorizing insights for further exploration. This guidance, rooted in Human-Computer Interaction, can be thought of as a mixed-initiative process (cmp. Fig. 2) where both the system and the user contribute to the analysis.

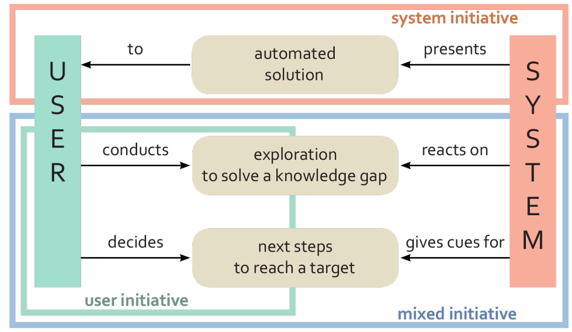

FIGURE 2. Guidance is a mixed-initiative process. On the one hand, the user explicitly or implicitly expresses his/her analysis target and a possible knowledge gap that hinders progression, by interacting with the system. On the other hand, the system reacts to the user's actions and gives cues that help to decide which steps to take to reach the target (Figure from [9]).

Consider our road traffic fines example; which visualization of the process model should be selected (for example, BPMN or Petri nets), how should the event logs be connected to the process model, how should the data quality be assessed, which PM algorithm should be applied, etc.

In summary, it is challenging to provide effective guidance that is timely, trustworthy, adaptive, controllable, and non-disruptive, which is tailored to the specific tasks in PM. Consequently, it is necessary to investigate models and concepts that are assisted by knowledge.

Incorporating Knowledge-Assisted Models

Consistent with the user-centered design tradition in Human-Computer Interaction, ensuring that a visualization system and its users share prior knowledge as a common ground for information exchange has been

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org)
identified by DOI: 10.1109/MCG.2024.3456916

identified as one of the top ten unsolved problems in visualization [4]. In particular, users need two types of prior knowledge to understand the intended message in visualization: *operational knowledge* (how to interact with the visualization system), and *domain knowledge* (how to interpret the content). Chen [4] states that while a focus on usability and perception- and cognition-aware design can alleviate the need for operational knowledge, domain knowledge cannot be easily replaced. Integrating operational and domain knowledge in the analysis process can improve not only the user-controlled analysis processes but also the automatic ones. Federico et al. [10] proposed a conceptual model of knowledge-assisted VA, extending van Wijk's model of visualization to a knowledge-assisted VA scenario by incorporating an explicit knowledge store and several knowledge-related processes.

Yet, most PM techniques rely mainly on event data as their primary input, augmented by knowledge captured in process models for specific tasks (e.g., conformance checking), but without incorporating any additional domain knowledge. Process models capture procedural domain knowledge, as they often specify organizational policies and required procedures. However, other forms of domain knowledge are lacking. In our road traffic fines example, rules exist for governing the allowed time between specific events (e.g., if a fine is not paid within 60 days after it has been sent, a penalty is added). These, however, cannot be visualized for current PM analysis, and this forms a limitation of existing analysis approaches. In general, it is important to determine the specific knowledge required for various PM tasks, including process discovery, conformance checking, temporal performance analysis, and predictive and comparative PM, within the framework of this conceptual model. Consequently, an adaption of the model of knowledge-assisted VA [10] is required to incorporate the PM parts from task-specific perspectives. The reliability of the data must also be addressed.

Capturing and Assessing the Data Quality, Uncertainty, and Provenance of Event Data and Process Models

Data quality and uncertainty is a well-known challenge in VA [1], [11] and PM [5]. Thus, it is just as important when it comes to the combination of both fields. Provenance in the VA literature roughly refers to tracking how data was generated or modified (called *data provenience*) and to how the users interacted with a VA approach (*analytic provenance*).

PM algorithms require event logs that contain well-grounded data of reasonable quality and are well-structured. In practice, they can often be erroneous and badly structured. Quite often, the event log contains missing, incorrect, imprecise, uncertain, or irrelevant data. Visual or VA approaches have been developed to address the issue of data quality (see, e.g., [11]–[13]). However, these approaches often fail to consider the specific needs of PM, such as case heterogeneity (i.e., mixed scenarios), diverse event granularity, or concept drifts.

Furthermore, data frequently includes temporal, spatial, and value uncertainty. For example, event logs may have uncertainties regarding which event type corresponds to each log entry or the precise timing of events. Visually communicating these uncertainties to the analyst (see, e.g., [11]) is crucial to facilitate better-informed reasoning.

In the context of PM, tracing the provenance supports the understanding of the broader context and environment within PM tasks. It is essential to incorporate VA methods to enable analysts to contextualize and interpret their findings, identify influencing factors, and make informed decisions. Several approaches have been proposed that use VA methods to assess data quality, uncertainty, and provenance (cmp. Fig. 3). However, these approaches usually deal with time-point-based data (instead of event sequences), do not incorporate domain knowledge and/or PM-related information, capture limited interaction techniques, etc., which are essential for providing appropriate, effective, and efficient VA solutions.

Exploring Event Data and Models over Time and Space

Usually, event data and models have an inherent temporal structure and, at times, also a spatial one. To this end, various approaches exist to visualize temporal, spatial, and spatio-temporal datasets [11], [14]. The models derived from the event data can be represented as networks or graphs, and techniques for visualizing temporal or event-based multivariate networks can be adapted [15]. For any such event data, at least the sequence of events and the observed locations are known and used to analyze paths and processes. However, time and space are a complex data domain with unique features that require special consideration [11]. In the event logs that serve for PM, timestamps are a mandatory feature, while spatial information is not commonly given. A recent exception that concerns spatial information as a contextual process layer is proposed in [16]. Time information mostly serves for activity ordering and is also used for performance

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org)
identified by DOI: 10.1109/MCG.2024.3456916

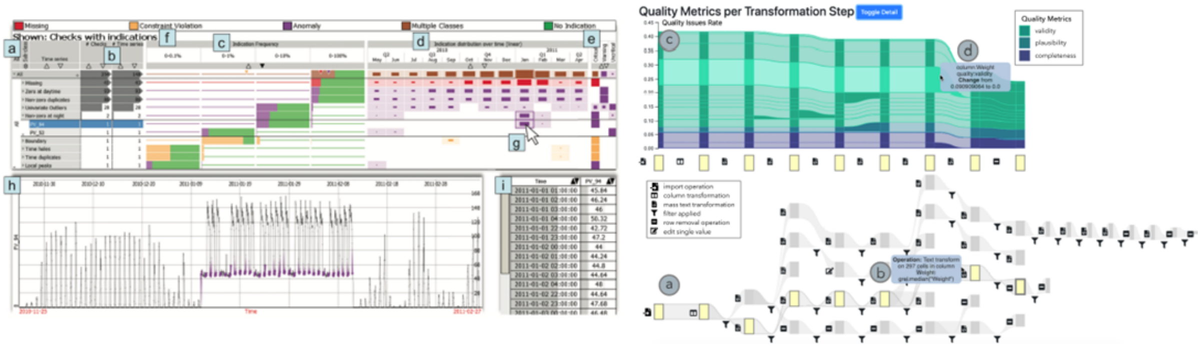

FIGURE 3. Two examples of VA approaches to support data quality and provenance. *Left-hand side*: Visplause consists of multiple linked views: (a) The DQ overview provides a hierarchical outline of results and additional information (b-g); linked views provide details of the selected data anomalies for validation (h, i) (Figure from [12]). *Right-hand side*: DQProv Explorer combines linked views: (a) The Provenance Graph view allows navigation of the individual data states; (b) On-demand mouseover displays information on the nodes and vertices; (c) The Quality Flow View shows the development over time for a selected wrangling branch; and (d) On-demand information on the Quality Flow View highlights the flow of the currently inspected metric and additional provenance information. (Figure from [13]).

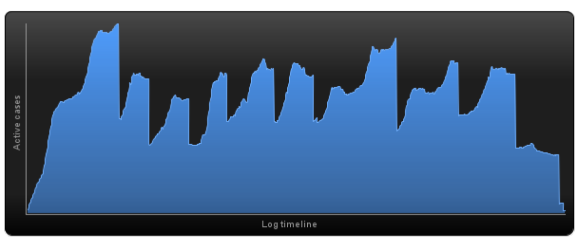

FIGURE 4. Active cases over time in the road traffic fine management log. A batch closing of open cases is observed through vertical drops in the total numbers

analysis. Nevertheless, insights into processes from data can be derived from more perspectives than the sole control flow.

Figure 1 shows typical process visualizations in PM (DFG and BPMN diagrams). As shown in Fig. 1, both modeling approaches can show execution frequency, costs or performance information about the activities or the transitions between consecutive ones via overlays and coloring. However, the insight provided thereby is limited because the displayed cues have a local perspective (to the single task or adjacent ones). Other analyses requiring a broader multi-perspective view can be desirable. In the road traffic fine process, e.g., a batch behavior of closing large numbers of cases can be observed (see Fig. 4), but this visualization is detached from other process perspectives, and requires further and separate investigation.

In summary, several open challenges can be indicated concerning time and space. First, to account for the various peculiarities of events (e.g., some cyclical behavior, events occurring with a certain delay, and shifting locations). Second, to visualize them adequately for the particular tasks at hand. Third, to incorporate time-based and space-based analysis into other process perspectives. Appropriate interaction techniques could also facilitate the VA process.

Providing Task-Specific and Tailored Interaction Techniques

Designing appropriate visual representations according to particular data and tasks supports humans in understanding data and models and extracting knowledge from data. Data exploration is an iterative process with trial and error loops. Suitable interaction techniques range from drilling down from overview visualization to investigating single events, allowing humans to focus on different parts of the data, explore the data from alternative perspectives, use different views, etc. Thomas and Cook [1, p. 30] already pointed out that "visual representations alone cannot satisfy analytical needs. Interaction techniques are required to support the dialogue between the analyst and the data." Typically, such an iterative process follows the *visual information seeking mantra: Overview first, zoom and filter, then details-on-demand.* by Shneiderman or the expanded VA mantra by Keim et al.: *Analyze first - show the important - zoom, filter and analyze further - details on demand* [1].

To illustrate, consider our road traffic fines example;

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org)
identified by DOI: 10.1109/MCG.2024.3456916

first, the analyst aims to gain an overview of the event log, including an assessment of its data quality and uncertainties; then, the process model should be visualized and important sub-aspects should be highlighted and filtered, which can then be analyzed and visually explored in the subsequent steps. The iterative process should facilitate a return to the raw data, an examination of the raw data in conjunction with the process model, and so forth.

In the context of time-oriented data, Aigner et al. [11] presented basic interaction concepts, including temporal navigation, direct manipulation, brushing & linking, and dynamic queries as well as beyond basic interaction, such as interactive lenses, natural visual comparison, guidance, event-based visualization, and interaction beyond mouse and keyboard. However, the authors conclude that advanced interaction methods have not been fully exploited. They suggest that future work should focus on better adapting existing interaction techniques, particularly to address the needs of time-oriented and event-based visualization.

In summary, interaction techniques are one of the key factors in the VA processes, yet they have not been extensively investigated in PM. However, these interaction techniques available in VA need to be tailored toward the various PM tasks and goals to facilitate the discovery of process models, conformance checking, and process enhancement. To this end, the particular characteristics of time and space need to be taken into account as well.

Deriving a Task Taxonomy for Supporting Human-Centered PM with VA

While the technical aspects of PM have evolved considerably in the last decades, there has been comparatively little emphasis on supporting the human element within the process [3]. To overcome this gap interdisciplinary efforts are needed. Especially the intersection of human-centered PM activities, cognitive aspects, and visual support offers opportunities to overcome this gap. To realize this opportunity, it is first vital to establish a common vocabulary and terminology that span the VA and PM areas, where often similar problems are referred to using different terms. This could be facilitated by a taxonomy, mapping analytical tasks in PM, cognitive tasks, and VA tasks. Such a mapping can be developed on the basis of existing taxonomies in both areas. The resulting task taxonomy would provide a common language to facilitate communication between different communities, aid in design decisions for process visualizations, identify research gaps, and provide a basis for evaluating process visualizations. However, the endeavor to establish such a task taxonomy remains an unsolved challenge.

Visual Sensemaking of Multi-Faceted Process Information

Process information often extends across a multitude of information systems and includes a number of data dimensions beyond the typical sequence of recorded actions that characterize the classical event log examined in the PM literature. It does not come as a surprise that the theoretical abstraction of a mere bag of activity sequences goes under the name of "*simple* event log" [5]. Reality, in fact, is often more complex than that. For example, several temporal constraints are exerted on the road-traffic fine management process (e.g., the date by when the fine can be sent to be still valid, the payment due date, the time to appeal), knowledge about the geographic location in which the activities take place would inform where most fines were issued or most cases were appealed to a court, resources involved in the majority of cases could highlight possible bottlenecks.

However, even if the finest PM algorithm processed the full range of data, time, resource, and control-flow dimensions, the results presented in a bi-dimensional, graph-based diagram, such as the typical process model would likely be either too convoluted to provide an effective visualization, or overly simplistic to encompass the full spectrum of interconnected findings. Making sense of multi-faceted information flattened in that way encumbers the work of process analysts, let alone final users or stakeholders. The interactive investigation of factual evidence stemming from information mined from data sources is a typical prerogative of VA, and several studies on the problem of representing complex, multi-faceted phenomena have been conducted [15]. The interplay of VA and PM could be of great benefit to both areas to this end. A framework leveraging that ensemble has been theorized only recently by Alman et al. [16], employing a multi-layered view of interconnected information anchored over a base visual cue from the domain in which the process unfolds (the *backdrop*). Nevertheless, multiple aspects remain unresolved, ranging from the practical implications and field studies to the technical challenges of enabling interactivity, run-time filtering and aggregation of views, and more, thus paving the path for future endeavors.

Providing Explainability

Exlainability of results is vital for PM, as the main motivation for PM is to make sense of processes

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org) identified by DOI: 10.1109/MCG.2024.3456916

and improve them. In general, artifacts generated by PM techniques, such as process models, are well understood and lend themselves to sense-making in a discovery context. However, highlighting the behavioral characteristics due to which the recorded process executions diverge from the expected unfolding of the model (which is typical in conformance checking settings) is far from being resolved. Similarly, with the increasing use of prediction techniques for PM purposes, the need for explanations becomes crucial, as understanding the root causes of predicted results is vital for process improvement.

To this end, various techniques for obtaining explainable results have been proposed. These mainly rely on rule-based explanations as well as counterfactual explanations In terms of visualization, rules are sometimes visualized using declarative constraint languages [17], while counterfactual explanations are visualized in a decision tree-like manner. Nonetheless, more VA approaches should be explored to increase the explainability. In pursuit of this goal, appropriate visual representations and interaction techniques should be used to understand and explore the event logs as well as process models (e.g., event-based multivariate network visualization [15]). Furthermore, the principles of knowledge-assisted VA [10] could be utilized to generate context-specific explanations according to particular tasks and users. However, the effectiveness of these representations for sense-making has not been established so far.

Easing Exploration and Comparison of Process Model

Exploration is an essential step in the PM process. However, in practice, PM analysts mostly rely on discovered process models with very few other visualizations that are specifically intended to support exploration. This aspect becomes especially pivotal in the context of highly flexible behavior, such as that of knowledge workers' processes. Indeed, these processes are typically depicted through declarative specifications, which bear an inherent cognitive complexity making them hard to be fully captured by the involved users, and thus call for new exploration strategies. Exploration is often aimed at a root-cause analysis: a relevant question for the road traffic fine process, e.g., would be why do fines remain unpaid. Such questions require a complex multi-perspective exploration. Comparison of different process behaviors is also essential for gaining insights. It can be pursued, e.g., with techniques comparing execution traces via alignment [7]. While some visual representation exists for trace alignment, its cognitive effectiveness has not been thoroughly tested and leaves room for improvement.

Considering this from the VA side, process models can be seen as event-based multivariate networks changing over time and space (i.e., network dynamics, nodes/edges addition and removal, are modeled after events, which have real-time coordinates and a finite duration). Some approaches of event-based drawing methods [15] could be utilized. However, there is still much unexploited potential for VA to better support pattern discovery, exploration, fine-tuning, and comparing process models, as well as identifying conformance problems and finding alternative solutions.

Supporting Qualitative and Quantitative Evaluation

Applying quantitative or qualitative evaluation strategies (or a combination thereof) [18], [19] to assess the effectiveness and appropriateness of proposed solutions is an open challenge in VA and PM, because VA and PM aim to support process-oriented analyses and exploration and human reasoning, which are qualitative dimensions and hard to quantify. The evaluation strategies need to take the exploratory nature of the tasks according to the appropriate visualizations, the type of data (e.g., event-based, irregularly sampled data), and the added value of the visualizations (e.g., data quality assessment, potential discoveries, and comparison of process models) into account. However, the evaluation of gained awareness and insights, as well as the transfer of actions from insights to be able to implement improvements (in particular in PM), is still an open challenge.

The work by Beerepot et al. [19] suggests a three-staged method on how to evaluate PM insights. The work emphasizes the importance of context. Insights that appear surprising when taking a single process perspective into account might no longer be surprising when other perspectives (e.g., temporal or organizational) are considered. This clearly indicates the need for a multi-perspective analysis of PM and VA insights. Moreover, the authors point out the need and importance of techniques that allow us to visually compare behaviors and include context, which is an opportunity for VA research. However, a targeted effort to develop such techniques, emphasizing the visual support of evaluation, is still pending.

Conclusion

The research disciplines of Visual Analytics (VA) and Process Mining (PM) have been separately active for

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org) identified by DOI: 10.1109/MCG.2024.3456916

years. As similar problems are sometimes relevant for both, it appears that the combination of PM algorithms with interactive visual data analysis and exploration, offered by VA, can make complex information structures more comprehensible and facilitate new insights. In this paper, we have illustrated the concepts of VA and PM to show how their combination can support the extraction of more insights from complex event data. We further acknowledged and elaborated on the challenges and opportunities for analyzing process data with VA methods and enhancing VA methods using PM techniques.

The VA and PM communities have a unique opportunity to set out for a new joint research path that is poised to deliver unprecedented results, which would be otherwise unreachable with the endeavors narrowed within the individual areas of investigation. Our analysis yielded a broad set of open challenges with a clear prospect of potentially promising results. With this paper, we hope to provide clarity, objectives, and perspectives for the work to come.

ACKNOWLEDGMENT

This work was initiated by Dagstuhl Seminar 23271 "Human in the (Process) Mines" [3].

References

[1] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, U.S. Department of Homeland Security, 2005, isbn: 0769523234.

[2] W. van der Aalst *et al.*, "Process mining manifesto," in *Business Process Management Workshops*, F. Daniel *et al.*, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 169–194, ISBN: 978-3-642-28108-2. DOI: 10.1007/978-3-642-28108-2_19.

[3] C. Di Ciccio *et al.*, "Human in the (Process) Mines (Dagstuhl Seminar 23271)," *Dagstuhl Reports*, vol. 13, no. 7, pp. 1–33, 2024, ISSN: 2192-5283. DOI: 10.4230/DagRep.13.7.1.

[4] C. Chen, "Top 10 unsolved information visualization problems," *IEEE Computer Graphics & Applications*, vol. 25, no. 4, pp. 12–16, 2005, ISSN: 0272-1716. DOI: 10.1109/MCG.2005.91.

[5] W. van der Aalst, *Process Mining: Data Science in Action*, 2nd. Springer Publishing Company, Incorporated, 2016, ISBN: 3662498502. DOI: 10.5555/2948762.

[6] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proceeding of the CHI '99 Conference on Human Factors in Computing Systems: The CHI is the Limit*, M. G. Williams and M. W. Altom, Eds., ACM, 1999, pp. 159–166. DOI: 10.1145/302979.303030.

[7] F. Mannhardt *et al.*, "The multi-perspective process explorer," in *13th International Business Process Management Workshops (BPM 2015)*, 2015, pp. 130–134. [Online]. Available: https://ceur-ws.org/Vol-1418/paper27.pdf.

[8] D. Ceneda *et al.*, "Characterizing guidance in visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 111–120, 2017. DOI: 10.1109/TVCG.2016.2598468.

[9] D. Ceneda *et al.*, "Guidance or no guidance? a decision tree can help," Eurographics Digital Library, 2018, pp. 19–23, ISBN: 978-3-03868-064-2. DOI: 10.2312/eurova.20181107.

[10] P. Federico *et al.*, "The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics," in *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017, pp. 92–103. DOI: 10.1109/VAST.2017.8585498.

[11] W. Aigner *et al.*, *Visualization of Time-Oriented Data*, Second Edition. Springer, 2023, ISBN: 978-1-4471-7527-8. DOI: 10.1007/978-1-4471-7527-8. [Online]. Available: https://timeviz.net.

[12] C. Arbesser *et al.*, "Visplause: Visual data quality assessment of many time series using plausibility checks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 641–650, 2017. DOI: 10.1109/TVCG.2016.2598592.

[13] C. Bors *et al.*, "Capturing and visualizing provenance from data wrangling," *IEEE Computer Graphics & Applications*, vol. 39, no. 6, pp. 61–75, 2019. DOI: 10.1109/MCG.2019.2941856.

[14] A. Yeshchenko and J. Mendling, "A survey of approaches for event sequence analysis and visualization," *Information Systems*, vol. 120, pp. 102–283, 2024, ISSN: 0306-4379. DOI: 10.1016/j.is.2023.102283.

[15] V. Filipov *et al.*, "Are we there yet? a roadmap of network visualization from surveys to task taxonomies," *Computer Graphics Forum*, vol. 42, no. 6, e14794, 2023. DOI: 10.1111/cgf.14794.

[16] A. Alman *et al.*, "Tiramisu: A recipe for visual sensemaking of multi-faceted process information," in *ICPM Workshops*, Apr. 2024, pp. 19–31. DOI: 10.1007/978-3-031-56107-8_2.

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org)
identified by DOI: 10.1109/MCG.2024.3456916

[17] C. Di Ciccio and M. Montali, "Declarative process specifications: Reasoning, discovery, monitoring," in *Process Mining Handbook*. 2022, pp. 108–152. DOI: 10.1007/978-3-031-08848-3_4.
[18] H. Lam *et al.*, "Empirical studies in information visualization: Seven scenarios," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1520–1536, 2012, ISSN: 1077–2626. DOI: 10.1109/tvcg.2011.279.
[19] I. Beerepoot *et al.*, "From insights to INTEL: evaluating process mining insights with healthcare professionals," in *56th Hawaii International Conference on System Sciences, HICSS 2023*, T. X. Bui, Ed., ScholarSpace, 2023, pp. 3245–3254. [Online]. Available: https://hdl.handle.net/10125/103030.

Silvia Miksch is University Professor and head of the Visual Analytics research unit (CVAST) at the Institute of Visual Computing and Human-Centered Technology, TU Wien. Her main research interests are Visual Analytics, https://www.cvast.tuwien.ac.at/team/silvia-miksch.

Claudio Di Ciccio is an associate professor at the Department of Information and Computing Science of Utrecht University (Netherlands). His research interests include process mining, formal methods, automated reasoning, and blockchain technologies. For more information, visit http://claudio.diciccio.net/.

Pnina Soffer is a Full Professor and the former Head of the Information Systems Department at the University of Haifa. Her main research area is process mining, combining technical and human-related issues. For more information, visit https://cris.haifa.ac.il/en/persons/pnina-soffer.

Barbara Weber is Full Professor for Software Systems Programming and Development at the University of St. Gallen (HSG), Switzerland. Barbara's research interests include human and cognitive aspects in software and process engineering, process modeling and mining. For more information, visit https://www.alexandria.unisg.ch/entities/person/Barbara_Weber.

Pre-print copy of the manuscript published by IEEE (available at ieeexplore.ieee.org) identified by DOI: 10.1109/MCG.2024.3456916

This document is a pre-print copy of the manuscript
(Miksch et al. 2025)
published by IEEE (available at ieeexplore.ieee.org).

The final version of the paper is identified by DOI: 10.1109/MCG.2024.3456916

References

Miksch, Silvia, Claudio Di Ciccio, Pnina Soffer, and Barbara Weber (2025). “Visual Analytics Meets Process Mining: Challenges and Opportunities”. In: *IEEE Computer Graphics and Applications* 44.6, pp. 132–141. DOI: 10.1109/MCG.2024.3456916.

BibTeX

```
@Article{        Miksch.etal/VisViews2025:VisualAnalyticsProcessMining,
  author        = {Miksch, Silvia and Di Ciccio, Claudio and Soffer, Pnina
                  and Weber, Barbara},
  journal       = {IEEE Computer Graphics and Applications},
  title         = {Visual Analytics Meets Process Mining: Challenges and
                  Opportunities},
  year          = {2025},
  number        = {6},
  pages         = {132--141},
  volume        = {44},
  doi           = {10.1109/MCG.2024.3456916},
  keywords      = {Process mining; Visual analytics},
  publisher     = {IEEE}
}
```