

A Taxonomy for Conformance Checking Visualizations

Marie-Christin Häge^[0009–0006–8031–8044] and Jana-Rebecca
Rehse^[0000–0001–5707–6944]

University of Mannheim, Mannheim, Germany
haege@uni-mannheim.de, rehse@uni-mannheim.de

Abstract. Conformance checking is a sub-discipline of process mining, which compares process execution data with predefined process models to identify deviations between them. Although recognized as the most important feature of process mining tools, conformance checking is currently not widely applied in practice. One reason for this lack of adoption is the absence of process-mining-specific visualizations, which can effectively communicate conformance checking results to practitioners. Although researchers have identified the need for such visualizations, they have left their development to the tool providers, such that available visualizations are highly different and difficult to compare. This inhibits the opportunities to conduct empirical research on conformance checking visualizations, which would be crucial to understanding user preferences. To address this issue and establish a foundation for future empirical research, this paper provides an overview of the existing breadth of characteristics of conformance checking visualizations in the form of a taxonomy. This taxonomy consists of six dimensions, which highlight in a structured manner what information is displayed in conformance checking visualizations and how this is visualized in different academic and commercial tools. Our research enhances the comprehension of visual analytics in process mining, particularly for conformance checking, and highlights promising avenues for future empirical research.

Keywords: Conformance Checking · Process Mining · Visual Analytics · Taxonomy.

1 Introduction

Conformance checking, one of the sub-disciplines of process mining, compares process execution data with a predefined to-be process model. By analyzing intended and observed process behavior, it identifies deviations between them [1]. This can help organizations to detect violations of internal or external regulations or find process improvements [6]. These capabilities make conformance checking highly relevant in practice. This was underlined by a recent study, which found that practitioners see conformance checking as the most important feature of process mining tools and expect it to grow the fastest in the coming years [1, 7].

Despite this established relevance, conformance checking often fails to deliver the desired outcome [6] and is currently not widely applied in practice [17, p. 39]. This discrepancy in the adoption is caused by different technical and organizational aspects [5]. One of these aspects is the absence of suitable visualizations, which could help to effectively communicate conformance checking results to the user [9]. This would significantly support the analysis process by allowing users, experts and non-experts, to acquire new insights more quickly, draw conclusions more effectively, and make more knowledgeable decisions [11].

Researchers have acknowledged the need for such visualizations [9, 12, 16]. However, so far, their development has mostly been left to the process mining tool providers, who have developed their own visualizations, based on the context and goals of the respective tools. Consequently, the visualizations differ highly in the displayed data and their representation [16]. These differences make it difficult to compare the visualizations and generate a structured understanding. However, this is crucial to gain insights into the users, their preferences, and the effects the different visualizations have on them [8, 9, 16]. Such insights would enable significant improvement potential for conformance checking visualizations and, consequently, for process mining and its adoption overall.

To address these issues, this paper provides a detailed overview of the existing breadth of characteristics of conformance checking visualizations in the form of a taxonomy, i.e., a multi-dimensional classification of a set of entities that allows to better relate and structure them. We created this taxonomy through a comparative analysis of academic and commercial process mining tools that include conformance checking visualizations. It consists of six dimensions and various characteristics, which present, in a structured manner, the information that is currently depicted in conformance checking visualizations. In addition, it highlights how this information is visualized across the different tools.

In the following, the necessary background and related work are presented in Section 2. Section 3 outlines the method used to develop the taxonomy, which itself is introduced in Section 4. We discuss the results, its limitations, and future implications and conclude in Section 5.

2 Related Work

In this paper, we investigate visualizations in process mining and thereby connect process mining research with the field of visual analytics. In this section, we provide the relevant background information on the two research fields and elaborate on the work connecting visual analytics and process mining.

Business Process Models. Organizations typically capture business processes as business process models, which support them in communicating, analyzing, documenting, redesigning, improving, monitoring or implementing processes [4]. For this purpose, different process model notations can be used. The most common ones are DFGs (Directly-Follow-Graph), Petri Nets, and BPMNs (Business Process Model and Notation) [21]. Each notation has a different focus, leading to differences in their depiction. In Figure 1, these process model notations are

used to visualize the same process. All models show the allowed process executions for different cases. Each case is an individual occurrence of the process, which is shown in the model through a path going from start to end. Each path contains activities that can or must be completed [24]. To visualize single cases, a suitable alternative notation are chevron diagrams, which depict the sequence of executed activities [17].

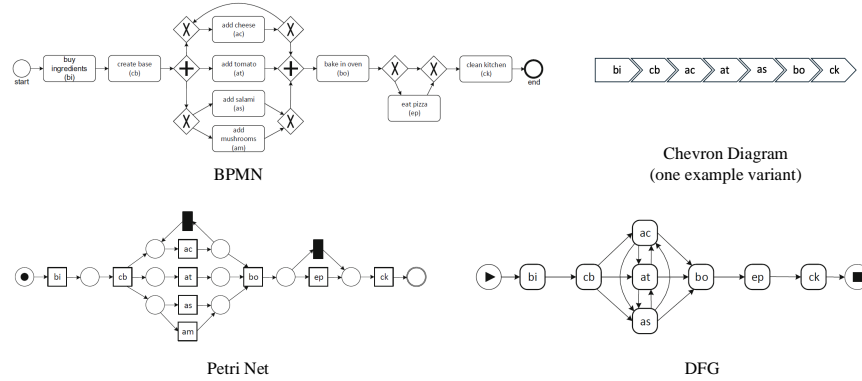


Fig. 1. Example Process depicted in All Four Process Model Notations [22, p. 10 ff.]

Although process models provide many advantages to organizations, they have certain limitations. Such challenges are the reduction of the process complexity in the models, the discrepancy between reality and the process model, the maintenance, and the restrictions of the different notations [4, 21, 24].

Process Mining. In order to gain data-driven insights into their processes, organizations can leverage process mining, a family of techniques aimed to analyze event data available from different information systems. The most common sub-disciplines are process discovery, conformance checking, and process enhancement [22]. They all can consider different process perspectives. The most common perspective is the control-flow, which represents the ordering of activities. Other perspectives may refer to time, data, or resources [22]. To conduct process mining, we require an event log, which is a collection of events. A single event contains at least a case identifier, an activity name, and a timestamp or another ordering mechanism. Events can also include additional attributes, such as resources or costs. A single trace is made of a sequence of events that describe a process execution. If a timestamp is available, for each trace the throughput time can be measured, which is the time from start to end. Multiple traces can follow the same activities, which can be then summarized as a variant. In the end, the results are often visualized as process models [21].

Conformance Checking. In this paper, we focus on the process mining sub-discipline of conformance checking, which considers event logs and existing process models. It aims to compare these two inputs and analyze how accurately they align. The process model usually represents the intended process, whereas the event log contains the real-world process execution. Through this compari-

son, deviations between them can be detected [1], which can help organizations to identify violations of internal or external regulations or find novel opportunities for process improvement [6]. There are three main conformance checking techniques: *token replay*, *rule checking*, and *alignments*. The latter two are the current state-of-the-art and can also be applied to other perspectives than the control-flow [1]. While rule checking techniques analyse if behavioural rules defined in a model are violated by certain traces in the log, alignment-based techniques differentiate between log, model, and synchronous moves. They compare a trace of an event log directly with a process execution in the model and identify possible alignments. The goal is an optimal alignment where the costs of log and model moves are kept to a minimum. To quantify the degree of conformance between a log and a model, a fitness measure is typically used, which indicates how well the model covers the recorded behavior in the event log [1].

Visual Analytics. The second relevant research area for this paper, visual analytics, can be defined as the “science of analytical reasoning facilitated by interactive visual interfaces” [20]. It aims to improve human understanding, reasoning, and decision-making with regard to provided data sets. Compared to other fields, such as information visualization, visual analytics emphasizes the importance of including the user throughout the analysis process. The goal is to achieve an intuitive and useful interpretation of the data rather than a cognitive or information overload when presenting only results [2, 11]. Moreover, visual analytics always considers the task that the user wants to perform by means of the visualization, which ensures that the visualization provides the desired insights [11, 13]. Through interactive capabilities, the user can see dynamics within the data, gain hidden insights, and adjust the complexity to their needs. As a consequence, they can fulfill tasks more effectively and efficiently [9, 13].

To systematically evaluate visualizations, Munzner [13] defines an analysis framework with three questions: ‘Why’, ‘What’, and ‘How’ (Figure 2). ‘Why’ analyzes the reasons for using the visualization (tool) and identifies the performed task. ‘What’ specifies the kind of data the users see in the visualization. ‘How’ describes the design of the visualization and its possible interaction options [13]. By answering these questions, a fitting visualization for a given situation can be identified.

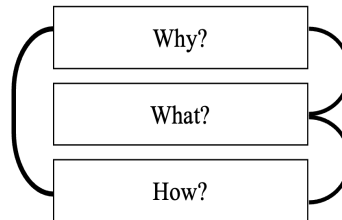


Fig. 2. Analysis Framework [13, p. 17]

Visual Analytics in Process Mining. Van der Aalst et al. [23] emphasized the relevance and opportunities of combining process mining and visual analytics already in 2011. Since then, there have been a few research initiatives in this direction. Researchers have assessed, developed, or categorized visualizations in process mining and established frameworks that consider also visual analytics [10, 12, 16, 18]. Klinkmüller et al. [12] identified the visualizations used for different process mining domain problems, as mentioned in analysis reports.

These domain problems are kept very general and represent the different process mining techniques. Rehse et al. [16] applied the three-part analysis framework by Munzner [13] to identify conformance checking tasks that are addressed by the currently available visualizations in academic and commercial tools.

Both papers find that generic visualizations are used predominantly, instead of those specific to process mining. Furthermore, the same visualizations are often used for different tasks and different information [12, 16]. These findings are a starting point for improving process mining through visual analytics, but many challenges and research gaps still need to be tackled. According to Gschwandtner [9], these challenges include, among others, the visualization of time-oriented data, the question of visualization evaluation, and the issue of scalability and aggregation. On the positive side, an interplay of both research fields could provide an improved understanding of the processes and an increased support of the user’s analysis process [9].

3 Method

The objective of this paper is to provide a structured overview of existing conformance checking visualizations through a taxonomy as a foundation for future research. Such a taxonomy allows the organization of knowledge and the identification of relationships among the underlying concepts. Its aim is the multi-dimensional classification of a set of entities that allows to better relate and structure them. It consists of a set of dimensions, where each dimension contains multiple characteristics. To develop such a taxonomy, we followed the iterative development method by Nickerson et al. [14], which is based on the design science paradigm in the information systems field. This structured approach suggests to develop a taxonomy over multiple iterations, following either an empirical-to-conceptual or a conceptual-to-empirical approach. An iteration with a conceptual-to-empirical approach derives the information from theory and literature, whereas an empirical-to-conceptual iteration uses empirical data. Before these iterations begin, we set the meta-characteristic that defines the goal and purpose of the taxonomy and the ending conditions that define the criteria for ending the development process. Then, in each iteration, one of the approaches is followed and the collected data will be structured, common characteristics will be identified, and lastly, grouped into dimensions for a taxonomy. This process will be repeated until the ending conditions are fulfilled [14].

Meta-characteristic and Ending Conditions. In the first step, we defined the meta-characteristic. This purpose of our taxonomy was to *identify what data and visualizations of conformance checking results are currently used in the existing process mining tools*. As ending conditions, we adopted the five objective and five subjective conditions from Nickerson et al. [14]. The objective ending conditions included that (1) no new dimensions or characteristics were added in the last iteration, (2) no new dimensions or characteristics were merged or split in the last iteration, (3) every dimension is unique and not repeated, (4) every characteristic is unique and not repeated within its dimension, and (5) each di-

mension consists of collectively exhaustive characteristics. Nickerson et al. [14] also suggest the ending condition that each dimension should have mutually exclusive characteristics, meaning only one characteristic per dimension is applicable when categorizing an object within the taxonomy. We excluded this condition due to the nature of visualizations and the underlying principles, e.g., visual expressiveness, which suggests the use of multiple variables. This does not completely contradict the ending condition as each combination of the characteristics and dimensions will be mutually exclusive and, therefore, unique [15, 19]. The subjective ending conditions are that the taxonomy needs to be concise, robust, comprehensive, extendible, and self-explanatory [14].

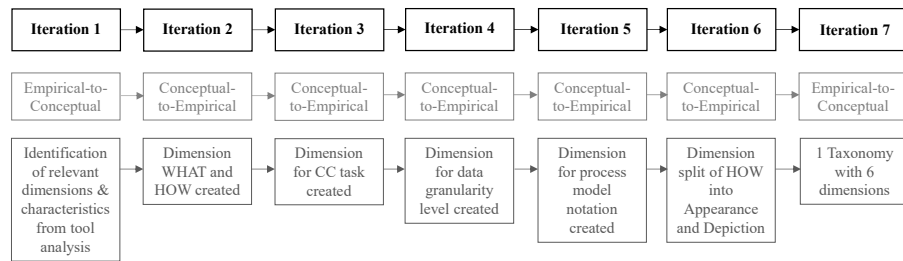


Fig. 3. Taxonomy Development Process

Iterations. We developed the taxonomy in seven iterations, outlined in Figure 3. The identified dimensions and characteristics are described in Section 4. As preparation, we selected all tools supporting conformance checking from the current tool landscape [16]. In total, we were able to get access to 11 academic and commercial tools: Appian, Apromore, ARIS, Celonis, IBM, Microsoft Power Automate, mpmX analytics, ProM ‘Replay a Log on Petri Net for Conformance Analysis’, ProM ‘Replay alignment on Performance/Conformance Checking’, ProM ‘Mine with Inductive Visual Miner’, and SAP Signavio Process Intelligence. For mpmX analytics, we only had screenshots available. To allow a comparison between the different visualizations of these tools, we used the same data set¹. We pre-processed it and generated a BPMN process model by applying different ProM plug-ins, allowing conformance checking in all tools.

First Iteration (empirical-to-conceptual). Because academic literature on conformance checking visualizations is scarce, we could not perform the more common conceptual-to-empirical first iteration. Instead, we analyzed the existing visualizations in the selected process mining tools in an empirical-to-conceptual fashion. These visualizations² had similar characteristics that could be clustered into groups, although each tool had a different focus when conducting conformance checking. For example, many tools used some type of process model notation, which we grouped together to process models. For each of these groups, we de-

¹ Data set: Road Traffic Fine Management Process

² Screenshots available upon request.

veloped fitting abstract dimensions resulting in the first version of the taxonomy that was then further specified throughout the next iterations.

Second iteration (conceptual-to-empirical). Next, we applied the framework by Munzner [13], shown in Figure 2, and added the questions ‘What’ and ‘How’ as dimensions in our taxonomy. For those dimensions, fitting characteristics were derived through the examples given by Munzner [13], which were transferred to process mining, and the characteristics that could be identified in the different tools. For the ‘What’, one characteristics group we identified was everything around conformity, e.g., the conformance rate. For the ‘How’, identified characteristics were, for example, ‘Number’, ‘Text’, or ‘Table’.

Third iteration (conceptual-to-empirical). Because the visualizations differed significantly, we divided them into high-level groups, according to their main purpose (the ‘Why’). For the characteristics, we relied on the four conformance checking task types currently supported by process mining tools: ‘Quantify Conformance’, ‘Break Down and Compare Conformance’, ‘Localize and Show Deviation’, and ‘Explain and Diagnose Deviation’ [16].

Fourth iteration (conceptual-to-empirical). In particular, the insights from literature provided by Rehse et al. [16] and others, such as De Weerd and Wynn [3], regarding the granularity level of process mining data, were used to differentiate between the data levels used in the visualizations. This allows a better classification. In detail, process mining data can be analyzed and visualized on log, variant, deviation, deviation category, or trace level in conformance checking [16]. We conceptualized these levels as characteristics under one dimension.

Fifth iteration (conceptual-to-empirical). For the task type ‘Localize and Show Deviation’, we recognized that the same visualization, a process model, was always used in the tools, but in different notations. Based on process modeling literature [21], we hence added this as a dimension, with DFGs, BPMNs, Petri Nets, and chevron diagrams as characteristics. While chevron diagrams only show a single trace, the others are more expressive. Nevertheless, all notations create a visualization of a process. As the other tasks do not necessarily use process models, we also added the characteristic ‘No Model’.

Sixth iteration (conceptual-to-empirical). Next, we went back to Munzner’s framework [13], with the goal of increasing the conciseness of the taxonomy. The framework divides the ‘How’ into two aspects [13]: graphical elements and visual appearance. Graphical elements describe primitive objects in a visualization and are characterized through their visual appearance. Together, they define the design space of visualizations [13]. Following this, we split the ‘How’ dimension into two individual ones: Depiction (Marks) and Appearance (Channels). By splitting up this dimension, the taxonomy with its dimensions and characteristics increased in comprehensibility and conciseness.

Last iteration (empirical-to-conceptual approach). Finally, our goal was to verify the taxonomy with all its different dimensions and characteristics. Again, we examined the visualizations in all tools and compared them with the developed taxonomy, confirming the completeness of the different dimensions and characteristics. Throughout this process, no adaptations were made, leading to the

fulfillment of the objective ending conditions. Regarding the subjective ending criteria, we did a further assessment, which showed that the taxonomy is concise and comprehensive as the dimensions and characteristics are kept to a minimum and are clearly defined. It can also be extended by adding new dimensions or characteristics with no issues. As the dimensions and characteristics are labeled precisely, as simple as possible, and on the needed differentiation level, the taxonomy appears also to be self-explanatory and robust. Hence, we concluded the development process as all subjective and objective ending conditions were fulfilled from our perspective.

4 Conformance Checking Visualization Taxonomy

The final version of our conformance checking visualization taxonomy is shown in Figure 4. It has six dimensions, with four to 42 characteristics per dimension. In the following, we describe each dimension and its characteristics in detail.

Task Type. Any conformance checking visualization within the tools aims to achieve one of the four conformance checking tasks: *Quantify Conformance*, *Localize and Show Deviation*, *Break Down and Compare Conformance*, and *Explain and Diagnose Deviation* [16]. Even though the visualization will have one task in focus, it might also be able to support another task. Therefore, the characteristics are not necessarily mutually exclusive.

Data Granularity Level. Data is needed as input for each visualization. This data will be refined to a certain granularity level, as shown within the visualization. For conformance checking data, six granularity levels are used: *Log*, *Variant*, *Single Trace*, *Deviation*, *Deviation Category*, and *Attribute* [1,16]. Consequently, this dimension is mutually exclusive.

Process Model Notation Type. Many, but not all, conformance checking visualizations incorporate process models. Due to their dominance and importance across the different tools, this dimension includes their use and a detailed differentiation based on their notation. If a visualization does not contain a process model, it falls under the characteristic *No Model*. Otherwise, the fitting notation will be selected, which is a *DFG*, *BPMN*, *Petri Net*, or *Chevron Diagram*.

What (Data Abstraction). This dimension describes what information is shown within the visualization. As one visualization can depict multiple pieces of information, selecting multiple characteristics is possible, so they are not mutually exclusive. Due to the different information depicted across tools, this dimension includes many characteristics. To simplify this, the characteristics are clustered into several groups:

- (1) Conformity information, such as the *Conformance Rate* and the *Number of Conform or Non-Conform Cases*.
- (2) Deviation information, including *Deviations*, their *Types*, and more *Details*.
- (3) Variant information, including *Variants*, and more *Details*.
- (4) Fitness information, including *Average*, *Maximum*, *Minimum*, or *Per Month*.
- (5) Activity information, including their *Names*, and more *Details*.
- (6) Frequency information, either for the *Case*, *Path*, *Variant* or *Activity* level.

| | | | | | | | | | | | | | | | | | |
|--|--|--|---|---|---|--|---|--|---|---|---|---|--|--|--|--|--|
| Task Type | <table border="1"> <tr> <td>Quantify Conformance</td> <td>Localize & Show Deviation</td> <td>Break Down & Compare Conformance</td> <td>Explain & Diagnose Deviation</td> </tr> </table> | Quantify Conformance | Localize & Show Deviation | Break Down & Compare Conformance | Explain & Diagnose Deviation | | | | | | | | | | | | |
| Quantify Conformance | Localize & Show Deviation | Break Down & Compare Conformance | Explain & Diagnose Deviation | | | | | | | | | | | | | | |
| Data Granularity Level | <table border="1"> <tr> <td>Log</td> <td>Variant</td> <td>Single Trace</td> <td>Deviation Category</td> <td>Attribute</td> <td>Deviation</td> </tr> </table> | Log | Variant | Single Trace | Deviation Category | Attribute | Deviation | | | | | | | | | | |
| Log | Variant | Single Trace | Deviation Category | Attribute | Deviation | | | | | | | | | | | | |
| Process Model Notation Type | <table border="1"> <tr> <td>DFG</td> <td>BPMN</td> <td>Petri Net</td> <td>Chevron Diagram</td> <td>No Model</td> </tr> </table> | DFG | BPMN | Petri Net | Chevron Diagram | No Model | | | | | | | | | | | |
| DFG | BPMN | Petri Net | Chevron Diagram | No Model | | | | | | | | | | | | | |
| What (Data Abstraction) | <table border="1"> <tr> <td data-bbox="493 758 651 1125"> (1) Conformity Conformance Rate (# of) Conform Cases (# of) Non-conform Cases # of Conform vs. Non-conform Cases # of Selected Cases Case Conformance Per Month </td> <td data-bbox="678 758 836 947"> (4) Fitness Avg. Fitness Max. Fitness Min. Fitness Fitness Per Month </td> <td data-bbox="886 758 1044 877"> (7) Violation Violations Root-cause Analysis </td> <td data-bbox="1094 758 1252 982"> (10) Perspective Resource Perspective Time Perspective Control-flow Perspective Data Perspective </td> </tr> <tr> <td data-bbox="493 1146 651 1314"> (2) Deviation Deviation Deviation Type Deviation Details </td> <td data-bbox="678 968 836 1171"> (5) Activity Activities Activity Name Activities per Case Activity Details </td> <td data-bbox="886 898 1044 1056"> (8) Move Model Moves Log Moves Synchronous Moves </td> <td data-bbox="1094 1010 1252 1335"> (11) Granularity Level Adjustable Granularity Level for Paths Adjustable Granularity Level for Activities Adjustable Granularity Level for Variants </td> </tr> <tr> <td data-bbox="493 1335 651 1440"> (3) Variant Variants Variant Details </td> <td data-bbox="678 1199 836 1440"> (6) Frequency Case Frequency Path Frequency Variant Frequency Activity Frequency </td> <td data-bbox="886 1077 1044 1255"> (9) Path Start/End Paths Paths (excl. Start/End Paths) Path Details </td> <td data-bbox="1094 1356 1252 1440"> (12) Additional Throughput Time Case Cost Alignment Statistics </td> </tr> <tr> <td data-bbox="493 1461 651 1493"></td> <td data-bbox="678 1461 836 1493"></td> <td data-bbox="886 1272 1044 1440"> (13) Visualization explanation </td> <td data-bbox="1094 1461 1252 1493"></td> </tr> </table> | (1) Conformity Conformance Rate (# of) Conform Cases (# of) Non-conform Cases # of Conform vs. Non-conform Cases # of Selected Cases Case Conformance Per Month | (4) Fitness Avg. Fitness Max. Fitness Min. Fitness Fitness Per Month | (7) Violation Violations Root-cause Analysis | (10) Perspective Resource Perspective Time Perspective Control-flow Perspective Data Perspective | (2) Deviation Deviation Deviation Type Deviation Details | (5) Activity Activities Activity Name Activities per Case Activity Details | (8) Move Model Moves Log Moves Synchronous Moves | (11) Granularity Level Adjustable Granularity Level for Paths Adjustable Granularity Level for Activities Adjustable Granularity Level for Variants | (3) Variant Variants Variant Details | (6) Frequency Case Frequency Path Frequency Variant Frequency Activity Frequency | (9) Path Start/End Paths Paths (excl. Start/End Paths) Path Details | (12) Additional Throughput Time Case Cost Alignment Statistics | | | (13) Visualization explanation | |
| (1) Conformity Conformance Rate (# of) Conform Cases (# of) Non-conform Cases # of Conform vs. Non-conform Cases # of Selected Cases Case Conformance Per Month | (4) Fitness Avg. Fitness Max. Fitness Min. Fitness Fitness Per Month | (7) Violation Violations Root-cause Analysis | (10) Perspective Resource Perspective Time Perspective Control-flow Perspective Data Perspective | | | | | | | | | | | | | | |
| (2) Deviation Deviation Deviation Type Deviation Details | (5) Activity Activities Activity Name Activities per Case Activity Details | (8) Move Model Moves Log Moves Synchronous Moves | (11) Granularity Level Adjustable Granularity Level for Paths Adjustable Granularity Level for Activities Adjustable Granularity Level for Variants | | | | | | | | | | | | | | |
| (3) Variant Variants Variant Details | (6) Frequency Case Frequency Path Frequency Variant Frequency Activity Frequency | (9) Path Start/End Paths Paths (excl. Start/End Paths) Path Details | (12) Additional Throughput Time Case Cost Alignment Statistics | | | | | | | | | | | | | | |
| | | (13) Visualization explanation | | | | | | | | | | | | | | | |
| How – Depiction (Marks) | <table border="1"> <tr> <td>Number</td> <td>Text</td> <td>Percentage</td> <td>Bar Chart</td> <td>Dotted Graph</td> <td>Line</td> </tr> <tr> <td>Symbol</td> <td>Table</td> <td>Filter Bar</td> <td>Donut Chart</td> <td>Bullet Graph</td> <td>Legend</td> </tr> </table> | Number | Text | Percentage | Bar Chart | Dotted Graph | Line | Symbol | Table | Filter Bar | Donut Chart | Bullet Graph | Legend | | | | |
| Number | Text | Percentage | Bar Chart | Dotted Graph | Line | | | | | | | | | | | | |
| Symbol | Table | Filter Bar | Donut Chart | Bullet Graph | Legend | | | | | | | | | | | | |
| How – Appearance (Channels) | <table border="1"> <tr> <td>Color Hue</td> <td>Color Luminance</td> <td>Color Saturation</td> <td>Size</td> <td>Shape</td> <td>None</td> </tr> </table> | Color Hue | Color Luminance | Color Saturation | Size | Shape | None | | | | | | | | | | |
| Color Hue | Color Luminance | Color Saturation | Size | Shape | None | | | | | | | | | | | | |

Fig. 4. Conformance Checking Visualization Taxonomy

- (7) Violation information, including *Root-Cause Analysis*.
- (8) Move information, if it is a *Model*, *Log*, or *Synchronous Move*.
- (9) Path information, such as *Start* or *End*, and more *Details*.
- (10) Perspective information, differentiating between *Control-Flow*, *Time*, *Resource*, and *Data*.
- (11) Granularity level information, adjustable for *Paths*, *Activities*, or *Variants*.
- (12) Additional information, such as *Throughput Time*, *Case Cost*, and *Alignment Statistics* providing further details on different granularity levels.
- (13) Visualization Explanation.

How Dimensions. The How Dimension is divided into two sub-dimensions for a better structure and higher comprehension. However, they are interconnected, as one depiction can be supported by a specific appearance.

How – Depiction (Marks). This sub-dimension of ‘How’ shows what graphical elements are used for the visualization. For conformance checking visualizations, 12 different characteristics can be applied. The simple ones include a *Number*, *Text*, *Percentage*, *Symbol*, *Line*, or *Table*. More complex depictions are charts, such as a *Bar Chart* or *Donut Chart*, or graphs, such as *Dotted Graphs* or *Bullet Graphs*. As multiple of these can be selected, they are also not mutually exclusive.

How – Appearance (Channels). This sub-dimension supports the previous one and emphasizes the used depiction. Characteristics of this dimension are visual channels used for conformance checking visualizations, including *Color Hue*, *Color Luminance*, *Color Saturation*, *Size*, and *Shape*. However, none of these possible visual channels does not have to be used, leading to the additional characteristic *None*. These characteristics are also not mutually exclusive.

5 Discussion & Conclusion

The aim of this paper was to generate an overview of the used conformance checking visualizations, cluster the information systematically, and provide a foundation for future empirical research in order to reduce the current research gap of visual analytics in process mining. With the developed taxonomy, researchers can now categorize the visualizations they use or develop. Moreover, the taxonomy highlights how diverse and complex conformance checking visualizations are. This taxonomy is one of the first artifacts that combines the foundations of visual analytics and process mining. It provides detailed insights on the depicted conformance checking visualizations across the different tools and the visualization variables used. We hope that the developed taxonomy allows researchers to deepen their understanding of visualizations and their relevance, and helps them identify information gaps that have not been shown so far.

Our work is subject to multiple limitations. First, we focused on procedural conformance checking and excluded declarative and rule-based conformance checking visualizations, which are not used at all by industry tools. Moreover, the results are based on data acquired only during a specific time period. Therefore, all findings and information, especially from the tools, may have been subject to modifications. Furthermore, this information was gathered from a selection

of tools. As we did not have access to all possible tools, important insights from these tools might be overlooked. Although the tool selection was done thoroughly, due to the fast-changing market of the tool providers, a tool might have been missed. However, with the covered tools in this paper the majority of the market for conformance checking solutions was considered. Fourth, the taxonomy only includes information on currently existing conformance checking visualizations. This means that it entails information derived from existing literature and the existing visualizations in the tools. Consequently, newly developed visualizations are not covered by this taxonomy and might include information that is not included so far. Furthermore, this also implies that it only covers what is currently available within the tools and does not cover what the user might need. In addition, the taxonomy development methodology used still leads to a certain degree of bias. Sixth, relevant literature, visualization variables, or information within the tools could have been missed and might not be included in the taxonomy, even though we conducted several iterations.

In conclusion, the taxonomy provides the foundation for future empirical research. We identified multiple potential future paths. Research should focus on developing and verifying the taxonomy further, leading to an increased objectivity and significance. It would be especially important to verify the different dimensions and characteristics through empirical data. Therefore, interviews with other experts who know either one of the research fields or both well should be conducted. Moreover, the operationalization of the taxonomy should be analyzed by applying it to further use cases. Although we tried to reduce the research gap of visual analytics in process mining, we identified further questions and problems that should be considered. First, visual analytics emphasizes focusing on the user and including them in the analysis process. However, so far, there are no user studies for visual analytics in process mining, also not for other process mining sub-disciplines. This provides an opportunity for empirical research to include the user when considering process mining visualizations and conduct research on it. Such research needs to check if the existing visualizations fulfill the users' needs for the different tasks, especially for conformance checking. Thus, future researchers could conduct experiments where users evaluate visualizations to test the user's comprehension of the process. With this knowledge, researchers could improve and standardize conformance checking visualizations. This could also include the adaptation or new development of the underlying algorithms, precisely solving the users' needs.

Overall, this taxonomy is an important contribution, as it provides a foundation for future empirical research on conformance checking visualizations, offering an initial understanding to be elaborated on. Following this path of research would have a significant impact on the adoption of process mining and would advance the research of connecting visual analytics and process mining.

References

1. Carmona, J., van Dongen, B., Weidlich, M.: Conformance checking: foundations, milestones and challenges. In: *Process Mining handbook*, pp. 155–190. Springer

- (2022)
2. Cui, W.: Visual analytics: A comprehensive overview. *IEEE Access* **7**, 81555–81573 (2019)
 3. De Weerd, J., Wynn, M.T.: Foundations of process event data. In: *Process Mining Handbook*, pp. 193–211. Springer (2022)
 4. Dumas, M., Rosa, L.M., Mendling, J., Reijers, A.H.: *Fundamentals of business process management*. Springer (2018)
 5. Dunzer, S., Stierle, M., Matzner, M., Baier, S.: Conformance checking: a state-of-the-art literature review. In: *S-BPM ONE*, pp. 1–10. ACM (2019)
 6. Emamjome, F., Andrews, R., ter Hofstede, A.H.: A case study lens on process mining in practice. In: *OTM*, pp. 127–145. Springer (2019)
 7. FAU: Process mining survey (2021), https://www.processmining-software.com/wp-content/uploads/2021/FAU2021_Process_Mining_Survey.pdf
 8. García-Bañuelos, L., Van Beest, N.R., Dumas, M., La Rosa, M., Mertens, W.: Complete and interpretable conformance checking of business processes. *TSE* **44**(3), 262–290 (2017)
 9. Gschwandtner, T.: Visual analytics meets process mining: challenges and opportunities. In: *SIMPDA*, pp. 142–154. Springer (2017)
 10. Kaoui, A., Theodoropoulou, G., Bousdekis, A., Voulodimos, A., Miaoulis, G.: Visual analytics in process mining for supporting business process improvement. In: *NIDS*, pp. 166–175. IOS Press (2021)
 11. Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., Melançon, G.: *Visual analytics: Definition, process, and challenges*. In: *Information Visualization: Human-Centered Issues and Perspectives*. Springer (2008)
 12. Klinkmüller, C., Müller, R., Weber, I.: Mining process mining practices: an exploratory characterization of information needs in process analytics. In: *BPM*, pp. 322–337. Springer (2019)
 13. Munzner, T.: *Visualization analysis and design*. CRC press (2014)
 14. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. *EJIS* **22**(3), 336–359 (2013)
 15. Oberländer, A.M., Lösner, B., Rau, D.: Taxonomy research in information systems: A systematic assessment. In: *ECIS* (2019)
 16. Rehse, J.R., Pufahl, L., Grohs, M., Klein, L.M.: Process mining meets visual analytics: the case of conformance checking. In: *HICSS*, pp. 5452–5461. ScholarSpace (2023)
 17. Reinkemeyer, L.: *Process mining in action*. Process mining in action (2020)
 18. Sirgmetz, M., Milani, F., Nolte, A., Pungas, T.: Designing process diagrams—a framework for making design choices when visualizing process mining outputs. In: *OTM*, pp. 463–480. Springer (2018)
 19. Szopinski, D., Schoormann, T., Kundisch, D.: Criteria as a prelude for guiding taxonomy evaluation. In: *HICSS* (2020)
 20. Thomas, J., Cook, K.: *Illuminating the Path* IEEE. CSP (2005)
 21. Van Der Aalst, W.: Foundations of process discovery. In: *Process Mining Handbook*, pp. 37–75. Springer (2022)
 22. Van Der Aalst, W.: Process mining: a 360 degree overview. In: *Process Mining Handbook*, pp. 3–34. Springer (2022)
 23. Van Der Aalst, W., de Leoni, M., ter Hofstede, A.: Process mining and visual analytics: Breathing life into business process models. *BPM* **17**, 699–730 (2011)
 24. Weske, M.: *Business Process Management - Concepts, Languages, Architectures*. Springer (2019)