UNIVERSITY OF TARTU
Institute of Computer Science
Conversion Master in IT

Marit Sirgmets

A Visualization Framework for Designing Process Mining Diagrams

Master’s Thesis (15 EAP)

Supervisors:
Fredrik Payman Milani
Taivo Pungas

Tartu 2018

A Visualization Framework for Designing Process Mining Diagrams

Abstract:

Event logs hold valuable information about the health of business processes. In order to access this information, raw data must be transformed to a comprehensible format. Process mining tools use various diagrams to support visual exploration of process logs. Designing such diagrams is not an easy task because oftentimes neither the developer nor user know where interesting or intriguing information lays. Therefore, the diagrams require thoughtful designs that on the one hand allow flexible exploration, and on the other hand, are simple and intuitive to use for analysts as well as non-experts. This work takes a look into existing solutions of process mining visualizations and the design decisions the visualizations are based on. A framework is proposed to simplify and improve the design process for process mining diagrams. It is based on data visualization theory as well as visualization practices in process mining. The effectiveness of the framework is tested in a case study.

Keywords:

Process mining, data visualization, process visualization, framework

CERCS: P170 Computer science, numerical analysis, systems, control

Protsessikaeve Diagrammide Kujundamise Raamistik

Lühikokkuvõte:

Sündmuslogid sisaldavad väärtuslikku informatsiooni äriprotsesside seisundi kohta. Informatsioonile ligi pääsemiseks peab andmestiku viima arusaadavale kujule. Protsissikaeve tööriistad kasutavad erinevaid diagramme, mis toetavad sündmuslogide visuaalset uurimist. Nende diagrammide kujundamine ei ole lihtne ülesanne, sest tihti ei tea arendaja ega kasutaja kus huvipakkuv informatisoon võib asuda. Seepärast peavad diagrammid olema paindlikud, kuid samas lihtsad ja intuitiivsed, et nii analüütikud kui ka mitteasjatundjad saaksid tööriista kasutada. Antud töö uurib olemasolevate protsessikaeve diagrammide kujundusi ja kuidas need kujundused on autorite poolt põhjendatud. Töös tutvustatakse ka raamistikku, mis on välja töötatud selleks, et lihtsustada ja täiustada protsessikaeve diagrammide kujundamist. See põhineb andmete visualiseerimise teoorial ja visualiseerimise praktikatel protsessikaeves. Raamistiku tõhusust on katsetatud juhtumuuringus.

Võtmesõnad:

Protsessikaeve, andmete visualiseerimine, protsesside visualiseerimine, raamistik

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Table of Contents

Appendix

1. Introduction

The pace of data generation is constantly increasing. Currently, 2.5 quintillion bytes of data are generated in a single day, which means in relative terms that 90% of data of the mankind has been created in past two years alone [1]. Numerous algorithms and platforms have been developed to turn the mounting data into usable form of information. Every day new companies implement those tools to leverage the power of data to escalate revenue and growth.

In practice, the situation is not as triumphant as in theory – the process of data transformation results often in struggle instead of success. Companies store large amounts of data that is not put in everyday use. One of the reasons for that is interpretation. Even though data is captured automatically, it is oftentimes meant to be used by people – operational workers, analysts, decision-makers – and the human interpretation aspect makes data vulnerable to misunderstandings.

The field of data visualization has taken a leading role in finding ways to tackle this issue. Data visualization is conveying the meaning of data in a graphical form, which is an intuitive way of comprehending non-concrete ideas. Our brain is very effective at visual and spatial thinking and therefore, interpreting clear data visualizations does not require much effort from the users.

Design of visualizations plays a crucial role in helping the user to focus on right aspects of the data to solve their tasks. The more specific is the task the user is solving, the easier it is to design a suitable visualization for it. However, sometimes neither the user nor the developer know where interesting or problematic points are located in the data. Designing visualizations for such tasks is a challenge. One field facing this challenge is process mining.

Process mining uses event log data to extract knowledge about business processes [2]. The knowledge can be retrieved from many angles. For example, analyzing performance metrics, comparing process flows or looking at the process from organizational perspective, which means analyzing process interactions between resources or departments.

Processes differ from one another and problems and opportunities can lay in various places. Therefore, most of process mining techniques do not offer ready-made answers, but function as a general guidance in the analysis process. They serve as a platform to empower the user in exploring processes and extracting insights. Process mining tools must be flexible to enable such exploration. They must allow the user to manipulate both data and visualizations.

The flexibility of tools is essential to process mining, but it has a cost – the more flexible the tool, the more complex it is for the user. Multiple visualizations and options for interactive exploration impose a great cognitive load on the user – he/she has to learn the meaning of different visual encodings and how to navigate in the landscape. It is designer's responsibility to help the user by creating clear and intuitive visualizations that do not compromise on the complexity of the data. In other words, the user should be able to focus on data exploration, not on figuring out the rules of the presentation.

Current process mining tools can be labelled as commercial or academic tools. These two types of tools have a clear gap [3]. Commercial tools, such as Fluxicon Disco[1], Celonis[2] or Aris PPM[3] have successfully developed presentations of process logs that are easy to use and understand. They have put a great effort into the visual and interaction design. On the other side stand academic tools that have equally great or even better algorithms than the commercial tools, but oftentimes lack the focus on design aspects [4]–[7]. The tools seem to be too complicated for non-expert users and not worth the effort of passing the learning curve. This stops wider audience from adopting the tools.

The thesis explores opportunities to improve the visualization practices in process mining field. To achieve that, it aims to identify the weak points in current practices of process mining visualizations in academia context and develop a framework that would help to overcome those problems.

The weak points are identified by analyzing various studies on process mining visualizations. Descriptions of visualizations and reasoning behind the design decisions in study papers give an overview of what has been the focus of developers and where could it be supported by a framework.

The need of frameworks for data visualization has been recognized. Indeed, several frameworks have been proposed for data visualization [8], interaction design [9] and user experience design [10]. However, to the current knowledge of the author, there is no framework designed specifically for process mining visualizations. The thesis aims to create a process mining visualization framework on the basis of the findings from process mining visualization practices as well as existing data visualization theory.

The structure of the thesis follows the steps of the framework development process. Section 2 lays the groundwork by giving a brief introduction about data visualization and process mining fields. Section 3 describes the state of art and identifies the aspects the framework has to improve. Data visualization theory that the framework is based on is described in the section 4 along with a full overview of the framework and its development process. Section 5 discusses the application and validation methods of the framework as well as summarizes the results of the validation. Section 6 concludes the paper.

[1] https://fluxicon.com/disco/

[2] https://www.celonis.com/

[3] https://www.softwareag.com/jp/products/aris_alfabet/bpa/aris_ppm/default

2. Background

This section describes the general background of the two main fields the thesis is based on – section 2.1 gives an overview of process mining and section 2.2. of data visualization.

2.1 Process Mining

Process mining uses (semi-)automated analysis methods to gain insights about real-life executions of business processes. It is a relatively young area, which can be placed between data mining and machine learning on one side, and process analysis and modelling on the other [11].

Process data is recorded by process-aware information systems and is stored in the form of event logs [12]. Event is a singular action that takes place in a specific time and place. A sequence of specific events forms a process instance – a one-time execution of a process. For example, a process instance is a sequence of events that are performed to process a loan application number 3465, and one event in this instance is employee X submitting the filled application on Tuesday 9:31 am. Similar types of process instances can be grouped into process variants – instances that execute the process in a similar way. For example, loan applications that are accepted follow a certain path, which is different from the loan applications that are rejected. A process includes all the possible variants of the execution.

Event logs vary from one another, but usually the minimum data they contain is the execution time of the event (timestamp), which process instance the event belongs to (case id) and the name of the activity this exact event performed (activity name or id) [11]. Event logs can be seen as network datasets because the events are connected to one another through the time sequence and the instance they belong to. Hence, an event log can be reconstructed into a process model – a conceptual representation of a process. Process models can show the topology of the process as well as its performance metrics, such as average duration or the number of process instances that were executed.

Process models can be analyzed through many lenses. For example, detecting bottlenecks through performance analysis or checking the compliance to business rules through conformance checking. Various process mining techniques and tools have been developed to guide analysts in diagnosing the problems and discovering opportunities. Process mining practices combine human and computer intelligence to improve business processes [4].

2.2 Data Visualization

Data visualization presents data in a graphical form, such as maps or diagrams. One of the oldest and most prominent data visualization areas is cartography that dates back to prehistoric notations of hunting and fishing maps [13]. Other popular types of graphical forms for visualizing data, such as bar-, pie- and line charts, were first introduced in 18th century [14].

Technological advancements have increased the volume and possibilities of data visualizations. Today, a data visualization is not simply a static image that represents a dataset. Many tools that use visual depiction are built for exploration purposes – the user has been enabled to actively manipulate the underlying data as well as its presentation. Consequently, visualization design is closely connected to user interface design and interaction design. Both, user interface and interaction design belong to the human-computer interaction research field, which addresses questions such as how to make the meeting with machines easy and pleasurable for the users [9].

Interactive functionalities give data visualizations flexibility, but also increase the complexity of designs. The developers of tools that use depiction of data must design visualization rules that can be applied to any dataset. For example, automated process mapping tools must be able to visualize any event log, such as a loan application process, a logistics process or a medical care process. The notations must be generalizable, but should not overly simplify the data. This has triggered experimentation on the data visualization field and emergence of new types of notation languages and visualizations.

Data visualization theory aims to guide the designers in making and assessing the design decisions. The theory takes input from several other disciplines, such as color theory, perception theory, human-computer interaction and requirements engineering. The rules of visualization are usually in a form of guiding principles and are not rigid. Therefore, designing tools that use visualizations requires a fair amount of creativity. Successful designs can give answers to questions that neither the developers nor users knew they were looking for [4].

3. The State of Art

This section presents the methodology and results of a literature review describing the state of art of the visualization techniques of process mining. The literature review is conducted by largely following Kitchenham's guidelines [15]. As such, the research questions were formulated (section 3.1), research sources and search string chosen and defined (section 3.2), selection criteria developed (section 3.3) and relevant information extracted from the papers (section 3.4). Section 3.5 unfolds the results of the research, including answers to each of the research questions of the study.

3.1 Research Questions

The research was conducted to lay groundwork for the topic of the thesis and pinpoint the improvement opportunities in the design process of process mining visualizations.

The following research questions were formulated:

RQ1: Which process mining techniques require visualizations?
Most of process mining techniques are supported by visualizations, but not all. Some techniques are executed on an algorithm level and do not need human interpretation, and therefore, do not require visualizations. This question helps to identify the process mining techniques that are visualized. It also helps to understand a relative importance of the visualizations in the process mining field – the more techniques require visualizations, the higher is the impact of the visualizations on analysis practices.

RQ2: How are process mining techniques visualized?
In order to have an overview of the common visualization practices on the field, the basic question of "how?" is raised. The answer to this question will highlight the current visualizations and design practices in process mining field. The visualizations can be studied on two levels – firstly, which diagrams are used in process mining, and secondly, which visual and interactive elements are used on the diagrams. These aspects show which level of complexity is required for the design of process visualizations – is it a matter of selecting a chart and using a ready-made solution or does the chart require heavy adjustments and creative input.

RQ3: How are the design choices for process mining visualizations made?
For the development of a framework, the guiding principles, which are currently used when choosing a visualization of process mining output, must be identified and examined. If the decisions are justified in a systematic way in most cases, there is no need for the framework of visualization design in the process mining field as the developers of those visualizations are already aware of the decisions they take. If the design approach is superficial or not present, the need for guidance exists and the framework shall be developed. This research question aims at gathering information on the thought-process behind the design decisions.

3.2 Research Sources and Search String

The available literature was explored in the Web of Science and Scopus databases. These electronic databases were chosen as they cover the main venues (conference proceedings and journal papers) where research in process mining is published. Google Scholar was considered, but then excluded because the search results are limited to the first thousand hits, therefore, many relevant works could have been left out due to that limitation.

The first search string was "process mining" AND ("visualization" OR "visualisation"). This search string resulted in 33 papers in the Web of Science and 88 papers in Scopus. The search string was too narrow to find enough relevant articles. Therefore, a broader search string was used in addition "visual" AND "process" AND "mining". The second search string resulted in 1105 articles in the Web of Science and 1518 in Scopus. With such a wide search, most of the results were irrelevant. However, many important papers were found that were not detected with the first search string. The complete search string for the literature review was as follows:

("process mining" AND ("visualization" OR "visualisation")) OR ("visual" AND "process" AND "mining")

The state of art research was conducted from December 2017 to January 2018 (incl.).

3.3 Selection Criteria and Study Selection

The selection process took place in three rounds. In the first round, the duplicates were removed from the initial papers. If two or more papers had the same title, were authored by the same persons, and had the same publishing year, they were regarded as duplicates.

Once the duplicates were removed, filtering by title was conducted. In this filtering, papers clearly out of scope such as papers about "coal mining" or "data mining" were discarded as they did not address the topic of process mining.

The remaining papers were filtered by reading the abstract and if needed, the introduction of the paper. A set of inclusion and exclusion criteria were applied in filtering these papers. The exclusion criteria were as follows:

- the paper is less than 3 pages in length - the papers that are less than 3 pages are too short to convey enough information for close examination, and therefore may be misjudged in the research study;
- the paper is not accessible in full length online through the university subscription of databases - the paper must be accessible in order to extract data from it;
- the paper is not in English - the paper must be understandable in order to analyse it;
- the paper is published more than 10 years ago - the papers that are older than a decade, may contain outdated information.

The inclusion criteria were based on the research questions. They were as follows:

- the paper is about process mining - the words "process" and "mining" are used in the context of process mining, not any other context, such as "mining industry" or "process of text mining";
- the paper includes at least one visualization for process mining outputs - the paper includes a clear proposal of how to visualize the outputs of introduced techniques, e.g. a process diagram for process discovery or performance dashboards for performance analysis;

- the proposed visualization(s) include design decisions that are made by the developers – the authors of the visualizations have taken design decisions to present the data in an understandable visual format, for example using visual channels, such as color hue or size. This leaves out papers that have not addressed the design aspect of visualizations at all and present the process mining outputs only in a form of technical modelling languages, such as Petri nets or BPMN.

The first two inclusion criteria clarify the scope and meaning of the keywords "process", "mining" and "visual" in the search string. The last criterion is added in order to find papers, where the authors have been intentionally or unintentionally creative with the design solutions and where the new visualization framework could have been used. It is important to investigate such articles to gain confidence in the awareness for design decisions or the lack of it (RQ2, RQ3).

3.4 Data Extraction

After three rounds of filtration, 28 papers were considered relevant to the thesis. These studies were read and data was extracted in order to answer the research questions. The following data was extracted:

- general information – title, author(s), publisher, pages and year;
- brief description – one sentence what the article was about;
- the process mining technique that is described – name(s) of the specific technique(s) mentioned in the paper, such as organizational mining or process performance analysis;
- platform, where the visualization is applied – name of the platform or tool that is the subject of the paper, such as ProM plugins or packages;
- the proposed visualization – images, detailed descriptions and general names of charts or graphs, such as chord diagram or a process map;
- the reasoning for selecting the proposed visualization (or an element of the visualization) – a brief textual overview why the proposed visualization solution was chosen;
- evaluation of the proposed visualization(s) – name and/or description of the validation of the tool and/or visualization, such as questionnaires or interviews;

The general and brief content overview were captured to be able to easily navigate amongst the studies and to reference the papers later. The process mining technique and name of the platform were documented to answer the research question 1 – which process mining techniques require visualizations. Images of the visualizations, their textual descriptions and chart names were documented to answer the research question 2 – how are process mining techniques visualized. The reasons of the design choices answered the research question 3 – how are the design choices for process mining visualizations made. In addition, the evaluation of the visualization was extracted in case the evaluation triggered changes in the design of the proposed visualization and could give additional information to answer the research question 3.

3.5 Results of the Review

The following section presents of the state of art review, including the overview of the studies (section 3.5.1), answers to the research questions (sections 3.5.2, 3.5.3, 3.5.4) and the conclusion of the state of art research (section 3.5.5).

3.5.1 Overview of the Studies

After the filtering, 28 papers were eligible for data extraction. The full list of the research papers is included in Appendix I. Figure 1 shows the distribution of publishing years of the articles that were included to the research.

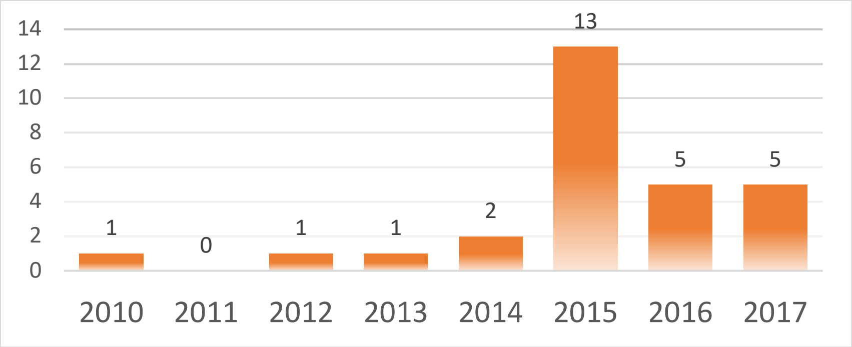


Figure 1. *Distribution of the publishing years of the scientific papers in the state of art research.*

The papers about process mining and visualization had a slow start from 2010 to 2014, with no papers at all to maximum two papers per year. Year 2015 was exceptional – 13 articles that were included to this research were published that year. Years 2016 and 2017 land in the middle – 5 papers were published each year. None of the papers were published 2018 as the research was conducted in December 2017 and January 2018.

3.5.2 RQ1 Which process mining techniques require visualizations?

Findings that answer the first research question are presented in the following categorization: visualizations for a single technique or function, multi-purpose visualizations and visualizations for general exploration. The papers that give a clear singular aim of the visualization are categorized as visualizations for a single technique or task (section 3.5.2.1). Multi-purpose visualizations papers are similar to the previous group in a way that these papers specify the purpose of the visualization, but instead of one technique, they are tailored for a variety of techniques and tasks (section 3.5.2.2). The third group, visualizations for general exploration, has studies where the purpose of the visualizations are expressed in high-level terms, without specifying any concrete techniques. Some of those papers claim that their visualizations are developed for a pre-technique exploration that helps the analyst to choose the right technique to continue the analysis with (section 3.5.2.3). All the previously mentioned sections introduce the used techniques and give an overview of the study papers. Section 3.5.2.4 summarizes various techniques that were mentioned in the studies.

3.5.2.1 Visualizations for a Single Technique or Function

13 papers described a visualization for one specific technique or function, out of which five were about process comparison, three about organizational mining (more specifically social

network analysis), two about performance analysis, one paper about prediction, one about signature discovery and one about root cause analysis for anomalies.

Process comparison means comparing two or more processes or process variants to one another [11]. Comparison can be done from many angles. Tool Process Comparator was built to compare process activities based on statistical difference between their performance metrics, such as frequency – how many process instances pass through an activity – or duration – how long does the execution or queuing take time [2]. Difference Graph gives a coarser point of view by visualizing differences in a categorical manner, marking differences as new, deleted, increased or decreased [16]. Similarly to Difference Graph, another prototype was built to show differences as categories, but this time it was only three levels – new, deleted or changed [17]. Process Profiler 3D takes also a look at the performance metrics, but visualizes the actual values of two or more process variants [18]. Process Profiler 3D was the only tool amongst the proposals that allowed comparison of more than two process variants. This was due to its visualization approach – instead of visualizing the results of the comparison (i.e. the value indicating the difference, such as Cohens d), it visualized the performance values of each process element to enable the user to do the comparison himself/herself in a simple way.

There was one paper about process comparison, which did not explain a visualization approach through a prototype. This study asked a general question of how to visualize differences of processes on a process map and the team researched it by conducting a literature review and a survey [7]. According to their findings, the most effective visual channels to indicate differences are color and shape [7]. Later these findings were implemented in the Difference Graph tool mentioned before [16].

Organizational mining uses event logs to explore the structure of organizational units and connections between them [11]. All three studies focused on a sub-technique of organizational mining – social network analysis. Social network analysis aims to find patterns of resources' communications and actions. One paper proposed a new graph – behavioral graph –, which visualizes similarities in the behavior of the performers in the system [19]. Behavior in this case means the resource's occurrence in various activities [19]. Another paper applied social network analysis techniques on logs from e-learning environment Moodle to extract a student network [20]. The last paper addressed specifically a visualization issue and proposed a chord diagram as a visualization technique for social network analysis (see figure 2) [21].

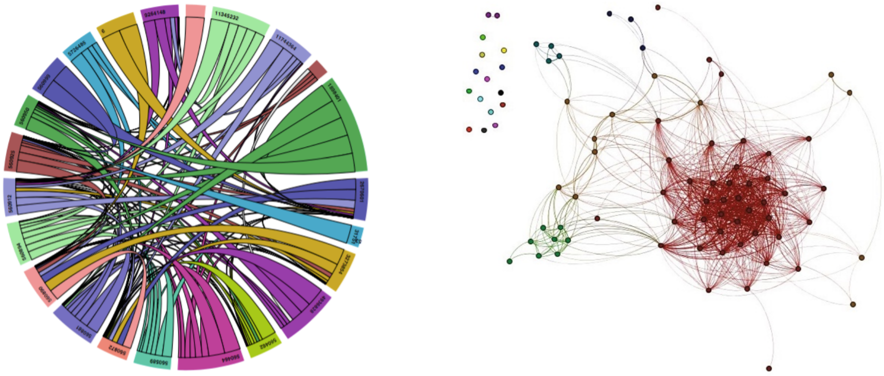


Figure 2. A *chord diagram (left) and a weighted node-link diagram (right) are both used to show connections between resources. Chord diagram image from* [21]*, weighted diagram image from* [20].

Performance analysis explores performance aspects of processes [11]. Common performance metrics are time and frequency as these variables can be extracted from most event logs. Usually, performance metrics are visualized as a part of a multi-purpose diagram, but there are some exceptions. For example, one study introduced a tool, which was developed to improve scheduling in train traffic [22]. In this case, the real timing of trains was taken as a primary metric to visualize the current issues that could be used for deeper analysis [22]. Another performance-focused study proposed a general approach for visualizing several performance metrics at once on the same process diagram to give an alternative to the usual way, where only one metric is shown at the time [5].

One study paper proposed a tool Nirdizati for predictive process monitoring [23]. Predictive analysis shows also performance metrics, but it includes predictive values to those metrics, such as the expected outcomes of the processes [23].

One paper addressed a technique signature discovery, which observes the trace patterns in event logs in order to detect differences and predict the class of uncategorized instances [24]. This technique has been used in fraud detection or indicating errors in machinery, such as X-ray machines [24]. Another study focused on detecting patterns in process instance traces, but from a slightly different angle – to conduct a root cause analysis of anomalous process instances [12]. The first technique gives the analyst a predictive view, the second framework gives the analyst a retrospective view.

3.5.2.2 Multi-purpose Visualizations

Ten studies were researching visualization possibilities for several techniques or functions. Eight studies had implemented the visualizations in a specific tool or prototype, one paper described three examples of different tools and one paper proposed a set of hand-drawn visualization concepts. Most of papers focused on a unique combination of techniques, therefore, each study is described separately below.

InterPretA is a tool for process oriented analysis [25]. The authors name performance and compliance analysis as the main techniques of the tool [25]. The last is an analysis of the alignment between the real-life execution of the process and the expected execution of the process, such as process according to business rules or protocols [25]. In addition, the paper includes guidelines about how to use the tool for deviance analysis, bottleneck analysis and frequency- or performance-oriented compliance analysis.

Event Streamer is a tool specialized in online discovery of declarative process models [26]. Declarative process models are based on a rule that every move that is not forbidden, is allowed, which is contrary to the usual models, where only the moves portrayed on the model are allowed [26]. In addition to process discovery the tool can be used to detect concept drifts [26], where the execution of the process has changed over time [11].

Multi-perspective Process Explorer is built to simplify the multi-perspective analysis of processes [27]. The tool “[integrates] existing data-aware discovery, conformance checking and performance analysis techniques” [27]. It aims to help the analyst by giving access to all the important tools for process analysis in one place.

PMCube Explorer produces several models (process variants) from one event log and those models can be juxtaposed or consolidated into one view for comparison purposes [28]. It combines process discovery, conformance checking and process comparison [28].

The main focus of Inductive visual Miner is to provide tools for process exploration, which in the authors' terms means "iteratively performing process discovery" [3]. The user can configure settings for filters and visualizations to discover the best representation of the process or to find problematic points in the process. In addition, the tool supports deviance mining. Deviance mining locates the points where process instances have skipped a step or taken additional unexpected steps in the process execution [3]. Inductive visual Miner was built to bridge the gap between commercial and academic tools. Two papers in the state of art research were based on this tool [3], [29].

One tool was developed specifically for medical industry for analysing performance, variation and conformance of care processes [30]. This tool has the highest number of different types of diagrams compared to other tools presented in the state of art research studies. It includes a treemap for understanding hierarchies of the activities, a flowchart and Sankey diagram for visualizing sequences and conformance and dashboards, such as bar charts and a scatterplot, for exploring performance metrics [30]. It allows analysis of one process as well as a comparison of two process variants [30].

One paper presented two complimentary diagrams that combine discovery and performance analysis [31]. The algorithm clusters the activities and the activity nodes on the process diagram include a list of all the activities that are combined into one, which is how the decomposition aspect of the process discovery is captured in a visualization [31]. Performance elements are included through visual channels, such as color hue, area and length [31].

One study was about challenges and opportunities in combining process mining and visual analytics [4]. This study presented three examples how this combination has been done successfully: Guideline Conformance tool is used for conformance checking and analyzing performance aspects of the processes; Plan Strips visualization is used to understand hierarchies of processes; EventExplorer is developed for a pre-technique exploration of the process [4].

The last study in this category was researching various options for process cohort comparison [32]. As a result, the authors designed three potential visualizations – a general model for exploring the performance and resource perspective; a superimposed model for visualizing differences and similarities on the level of activity match between cohorts; and a side-by-side model visualizing the actual execution and waiting times of the activities [32].

3.5.2.3 Visualizations for General Exploration

Five papers researched solutions for general exploration. Two of the studies claimed that their solutions can be used as a first exploration point of the process analysis, where the exact technique is not yet decided upon. These tools are called here pre-technique tools. Other three named the purpose of their visualizations in general terms without pinpointing the exact techniques.

The first pre-technique tool is Log On Map Replayer that visualizes instance traffic in an animated form and allows the user choose, which map it can be visualized on [33]. For example, the user can display the visualization on a geographic map, deadline map or a process flow map (see figure 8) [33]. Another pre-technique tool proposed turtle graphics method as a way to visualize process instance paths and their performance [34]. Turtle graphics is a method in computer graphics, where the images are drawn using a relative cursor. The cursor is called the turtle. In an example, the turtle was given directions according to events – each event had a direction up, down, right or left – and the turtle moved according to the events the process instance passed through, leaving behind the traces of the process (see figure 3) [34].

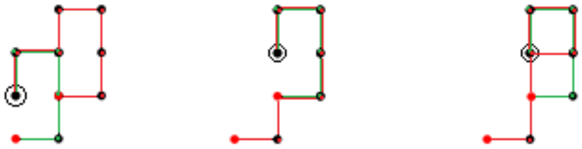


Figure 3. *Traces drawn with a turtle graphic approach show the topology of deviances. Colors represent different process variants, direction of the trace shows the activity that was executed* [34].

One study explored a possibility of using coordinated multiple views visualization in process exploration [35]. Coordinated multiple views means that the user is given multiple views on the dataset through several (juxtaposed) diagrams and he/she can interact with the views in a coordinated way, such as applying cross-filtering, where changes on one diagram trigger changes on all the diagrams [35].

Another study was concerned about the overloading of the process models with visual elements and proposed new ways to visualize data flow and process flow simultaneously [6]. The authors designed four different visualization concepts as alternatives to BPMN modelling language [6].

One general exploration visualization was proposed for ERP usage logs [36]. These visualizations give the operational users of ERP a process-, task- and context-related information during their active use of the system [36].

3.5.2.4 Summary of Process Mining Techniques

According to K. Oruste literature review [11] process mining techniques can be divided to eleven different categories: process discovery, performance analysis, process optimization, conformance checking, performance prediction, organizational mining, decomposition, model repair, deviances, concept drift and process comparison. Most of those techniques were mentioned in the studies in the state of art research. No visualizations were specifically developed for process optimization, but one tool visualized conformance checking with the intention to support analysis for process optimization [25]. Another technique that did not have any dedicated visualizations is decomposition, which is usually done on an algorithm level and therefore, does not require specialized diagrams. The most popular topics were visualization of process performance [5], [18], [22], [25], [27], [30], [31] and comparison (both conformance checking as well as comparing two or more processes to one another) [2], [4], [7], [16]–[18], [25], [27], [28], [30], [32]. In addition to Oruste's list, two additional techniques were clearly defined as process mining techniques – process exploration (i.e. repeated interactive process discovery) [29] and signature discovery (i.e. pattern detection) [24].

3.5.3 RQ2 How are process mining techniques visualized?

The data was extracted from two perspectives to answer this question. Firstly, which type of diagrams are used in the research papers to visualize process mining data? And secondly, which visual and interactive elements were described in the research papers? The diagrams were explored only through the study papers (text and images), not through immediate manipulation of the diagrams. The actual use was not included due to the accessibility of the tools and prototypes – some are easily available, while others are not. Therefore, the scope of the sources of this research question was narrowed down to the information accessible through the study papers.

Section 3.5.3.1 describes the types of diagrams presented in the research papers and section 3.5.3.2 gives an overview of the visual and interactive elements mentioned in the research papers. Section 3.5.3.3 summarizes the results from the perspective of both, diagram types as well as visual and interactive elements used on the diagrams.

3.5.3.1 Diagram Types

Overall, eleven different types of diagrams were introduced in 28 studies. These eleven types were: a node-link diagram, a treemap, a pie chart, a bar chart, a line chart, a parallel coordinate plot, a box plot, a scatterplot, a gauge chart, a graph for an online instance stream and a trace map drawn with turtle graphics. All these diagrams are listed in table 1.

Table 1. Diagrams Used for Process Mining Visualization

Diagram type	*Number of studies*	*References*
Node-link diagram	24	[2]–[7], [16]–[22], [25]–[33], [35], [36]
Bar and triangle charts	7	[18], [23], [28], [30]–[32], [35]
Pie chart	4	[5], [23], [24], [33]
Line and area charts	3	[4], [25], [26]
Treemap	2	[4], [30]
Scatterplot	2	[24], [30]
Parallel coordinate plot	1	[12]
Box plot	1	[25]
Gauge chart	1	[35]
Instance stream graph (custom)	1	[26]
Turtle graphics trace map (custom)	1	[34]

A node-link diagram was the most common way to visualize process data in the studies – 24 out of 28 papers presented node-link diagrams in one form or another. Directed node-link diagrams were used to visualize techniques such as process discovery, conformance checking, process comparison and performance analysis. Undirected node-link diagrams were used to visualize social networks. Additional support diagrams, such as bar-, pie- and line charts, were used to visualize performance metrics. Hierarchical process relations were shown on treemaps. Correlations were visualized on scatterplots. Other charts were a parallel coordinate plot for pattern detection [12], box plot for value distribution [25] and gauge chart for performance metrics visualization [35]. All those charts were used only once in the studies. An instance stream graph [26] and turtle graphics trace map [34] were both custom-built diagrams and also used only once. All the types of diagrams and their use in process mining are further described below.

A node-link diagram is a typical visualization for network data. It consists of nodes (items) and links (connecting marks between items). In the study papers, node-link diagrams were usually in a form of process maps, where nodes represent activities and links the succession of the activities (see figure 4). Another option was a weighted node-link diagram that is often used in social network mining (see figure 2) [19], [20]. A special version of a node-link diagram proposed for social network mining is a chord diagram (see figure 2) [21]. One node-link diagram was for a declarative process discovery, which means that the visualization does not show the sequence of the activities, but the links represent different types of relations between the activities (see figure 5) [26]. In one version the node-link diagram was presented as a value chain – visualizing only nodes with directional shapes without links [27]. One node-link diagram was a specific type of a flow chart – a Sankey diagram (see figure 6) [30].

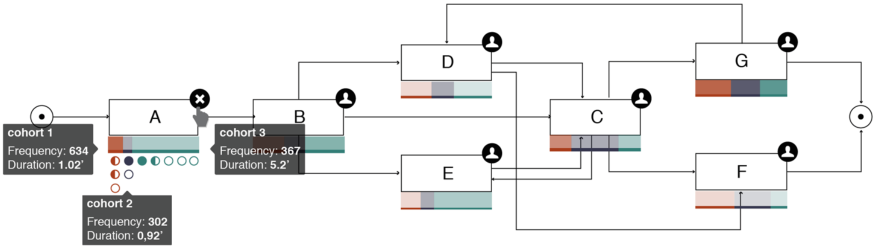


Figure 4. *A process map, where nodes represent activities and links the succession. The bottom of the nodes are designed as stacked bar charts that show performance metrics* [32].

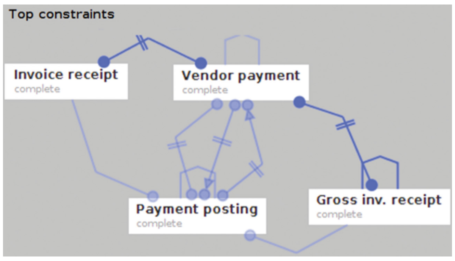


Figure 5. *Declare notation shows the types of connections between the activities instead of their succession. It is used for a declarative process discovery. Image from* [26].

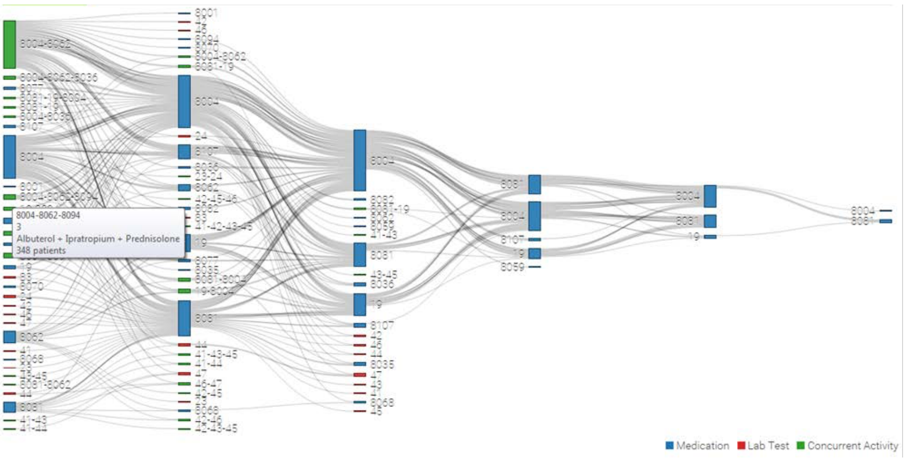


Figure 6. *A process flow shown in a Sankey diagram. Image from* [30].

A treemap is a nested diagram, which visualizes hierarchies – larger containers, which are on top of hierarchy, hold smaller containers, which represent lower levels of the hierarchical structure (see figure 7). Treemaps were presented in two papers, in both cases it was used to visualize hierarchies of care processes [4], [30].

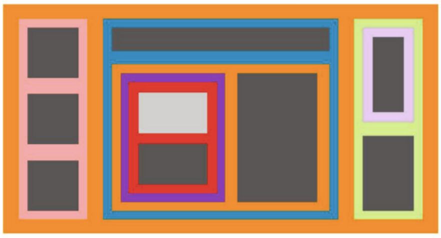


Figure 7. *An example of a treemap, where the care process activities are nested according to their hierarchy – the further back the rectangle, the higher level of hierarchy. Image from* [4].

A pie chart is a circular chart, which can be divided into areas, whereas the size of each area (also called slice) shows a relational value. Pie charts were mentioned and/or visualized in four research studies [5], [23], [24], [33]. In two cases the pie chart was dynamic – the slices were changing on the visualization. In Nirdizati it was done due to the use of a dynamic dataset – the visualization is constantly evolving, while new or updated data is flowing in [23]. Log On Map Replay used animation effect on a static dataset – the data is shown in a time sequence to replay the execution of the process (see figure 8) [33].

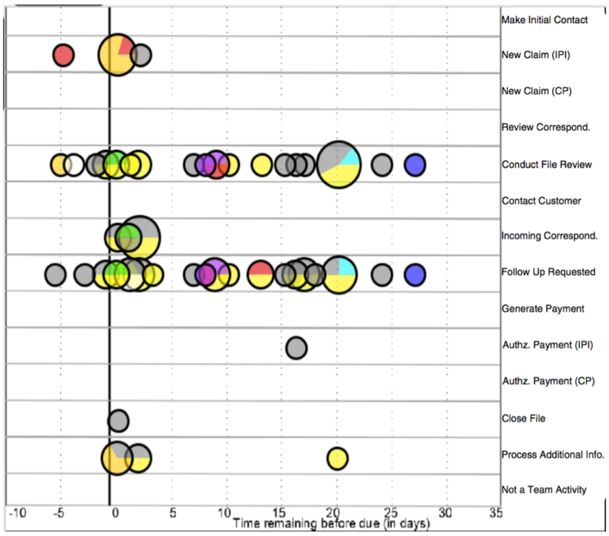


Figure 8. *Log On Map Replay portrays animated pie charts on any background map. On a current map the x-axis shows the time to the deadline, while the color of the slices of pies show the state of process instances. Image from* [33].

Bar chart is a diagram, where values are shown with the length and/or width of rectangles or lines. In the studies, bar charts were used as stand-alone charts [23], [28], [30], [35] as well as charts integrated into process diagrams [18], [31], [32]. The last version was implemented also with using a variation of a bar chart – a triangle chart, where the height of the tip of the triangle and width of the base of the triangle show values [18], [32]. In some cases, bar charts were stacked – the bar is divided into areas, where each area shows a relational value of a sub-item (see figure 4) [30]–[32].

A line chart is a diagram that shows value points with a connected line. It is usually used to show trends and/or time-based data. A version of a line chart is an area chart, where the area between the bottom of the chart and the line is filled with color. Line charts were used in two tools [4], [26] and an area chart in one tool [25].

A parallel coordinates plot shows the patterns through lines like line charts, but the idea behind the visualization is different. A parallel coordinates plot can be seen as a visual table, where rows are represented by lines and attribute columns are replaced with a y-axis of the value range of the attribute. Each line (row) crosses this axis at the point, which represents the value that otherwise would have been written in the cell of that row (see figure 9) [12]. The parallel coordinate plot is often used to analyze multi-dimensional data, such as correlations between several attributes. In process mining, it has been used to detect anomalous patterns and find out their root causes [12].

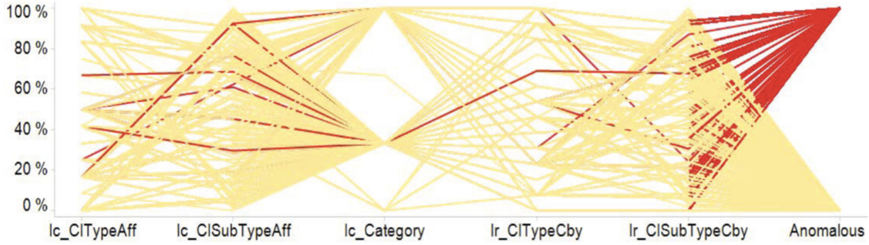


Figure 9. *A parallel plot is used to highlight patterns of anomalous incidents. Image from* [12].

A box plot is a bar-chart-like diagram, where bars are replaced with box elements that show quantiles and outliers. The line in the middle of the box shows median, the ends of the box show location of the 1st and 3rd quantiles and the lines (whiskers) running out of the box end, where maximum and minimum values lay. Any marks laying beyond the lines are outliers. This type of a chart was mentioned only once in the research papers for the use of compliance or bottleneck analysis [25].

Scatterplot visualizes relations between two variables. Items are shown as points, where the value of the first variable is shown by the location on the x-axis and the second variable by the location on the y-axis. It can be used to visualize correlation [30] or clusters [24].

A gauge chart is usually visualized in a form of a speedometer, where values are aligned clock-wise in a half-circle and a needle is pointing to the current value. A gauge chart was used in one prototype diagram to show a total number of instances that passed through a process [35].

One diagram was specifically built for an online stream of events – the x-axis showing the time and points on the graph showing new events coming into the dataset [26]. Y-axis did not carry any other meaning, except aligning events that were streamed at the same time [26].

The last type of the diagram is a graph that was formed after the use of turtle graphics to show process topology and performance metrics. It is similar to a process diagram because it draws the path of the process, but instead of nodes showing the activities, the direction of the line is used – the lines directed left, right, up and down represent four different activities (see figure 3) [34].

3.5.3.2 Visual and Interactive Elements

Most of chart types were used in a traditional way, using the visual channels that the charts allow. For example, in the case of stacked bar charts, the length channel was used to visualize the values and color hue to differentiate between sub-groups. None of the research studies tried to challenge these common combinations between chart types and the visual channels they usually use. The only exception is the visualization of node-link diagrams, which was depicted in a different manner in each study that represented a version of this diagram. In addition to the portrayal of the base topology, other visual channels, such as shape, color or size, were used to show additional data. As the state of art research is interested in the creative input of the developers and designers (see inclusion criteria in section 3.3), this section describes which visual channels and interactive elements were used in design of node-link diagrams, not any other chart types presented in the papers.

The visual and interactive elements of node-link diagrams can be divided into the following sub-sections: layout, faceting, visual channels and interactions.

Layout of a diagram presents the base structure of the network data. Only few papers described the details of the layout decisions. Ordering was mentioned in four papers – two papers clarified that the order of the nodes represents the sequence of the process [4], [30] and two papers emphasized the undirected nature of the diagram [20], [26]. Alignment was described in four different concepts – alignment that avoids edge crossings [17], [36]; matching alignment amongst several juxtaposed diagrams [16]; alignment separating activity sequence and data flow (see figure 10) [6]; and alignment of parallelism [18], [26]. Another aspect of layout is dimensionality. It was mentioned in two papers, which used three dimensions instead of typical two-dimensional layout [6], [18]. One paper brought out a deterministic design of the layout of its diagram – every time the same dataset is loaded, the nodes and edges are placed to the same location [17].

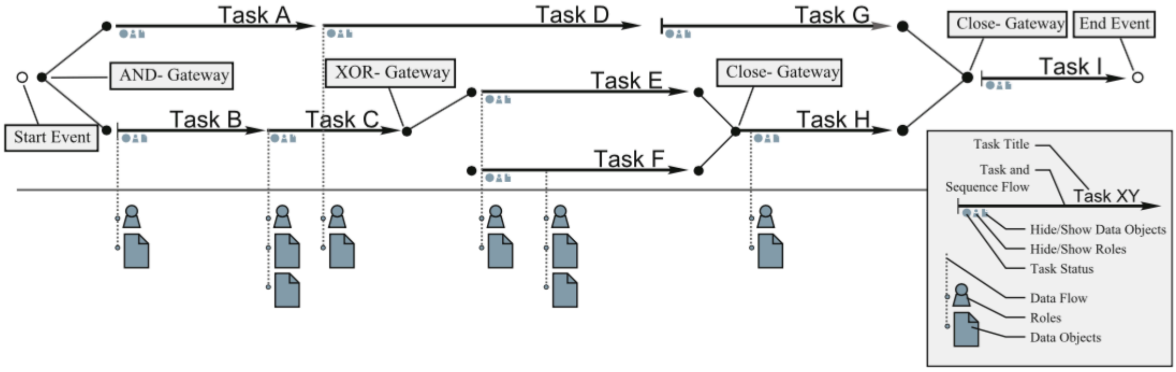


Figure 10. *Thin line diagram concept uses vertical alignment to separate process flow elements from data flow elements. Image from* [6].

Faceting means designing the structure of the various views of the diagram. Faceting can be for example superimposition of layers, juxtaposition of multiple views or embedding data that can be revealed by user interactions. Complex faceting is often used when one view cannot visualize all the data that the user might be interested in.

In the research studies, superimposition was used to layer markings of additional data, such as color hue or saturation, or different type of charts, such as pie charts or bar charts, on top of the base node-link diagram [5], [16], [18], [22], [25], [28], [29], [32], [33]. A special type of superimposition is a process instance animation, where the instances are moving in relation to the base diagram. This was introduced in two papers [29], [33]. Embedding data was a common way to make additional data, such as sub-processes or details of values, accessible on demand [2], [4], [18], [30], [33], [36]. Juxtaposition was used for comparison purposes or for offering different views on the same process [4], [16], [28], [32], [33], [35]. In the first version, different process variants with the same visualization were placed side-by-side (see figure 11); in the second, different visualizations with the same data were placed side-by-side.

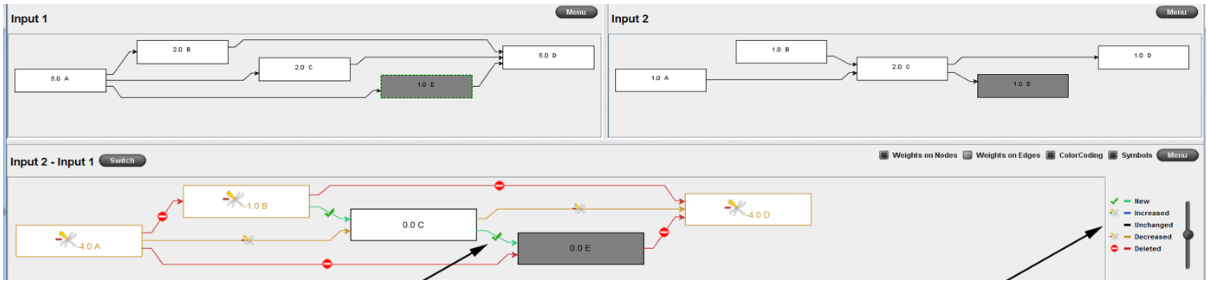


Figure 11. *Differencegraph uses juxtaposed faceting as well as superimposition. Two original process models are visualized on the top, while the integrated model that combines both of the models is visualized in the bottom. Differencegraph uses linked highlighting – if an activity node is selected on one of the three diagrams, the selected activity is highlighted on all the diagrams. Image from* [16].

Visual channels are the visual properties that are used to express the values of the data [37]. The following visual channels were described in the studies: color hue, color saturation and luminance, shape, size, length, spatial position and transparency.

Color hue was used on 18 different node-link diagrams. It was often used to show element's belonging to a certain group, for example to differentiate process variants [18], [22], [32], types of activities [30], types of events [4], a status of process instances [23], types of differences [2], [16], [17], a level of duration (high, medium, low) [31] or resources and their connections [20], [21]. This is not a complete list as the color coding can be also customizable by the user [33]. In addition, several highlights were marked with a different hue of the selected element [6], [16], [35], [36]. Hue was also combined with other channels. For example, with a shape channel [18], [22], [29] or color saturation [2], such as shades of blue visualize frequency and shades of red visualize duration [25].

Color saturation and/or luminance were used to visualize numeric variables, such as frequency of occurrences [25], [26], duration [25], level of similarity or differences [2], [7], [25], [27]. Similarly to color hue, the attributes this channel communicates can be customizable by the user [33].

Shapes were often used on the diagrams that were presented in the state of art studies. Shapes can be a modification of nodes, a variation of the links (e.g. dashed) or additional symbols layered on top of the diagram. Nodes were modified to show a direction of the process flow [27], [38] or a type of a difference between process flows [4], [7], [16], [17]. In addition, the shape of the bordering line of the node can carry meaning, for example dashing [7], [17] and levels of blur [32]. The shape of links can communicate direction (arrows) or type of the flow [6], [7], [17], [29], [36]. It can also be more complex and express different rules between activities as it is done in the visualization of a declarative process model using Declare encodings (see figure 5) [26]. Additional symbols were used to visualize gateways [6], [18], types of activities or resources [6], [30], [32], [35], differences in process flow and performance [7], [16], data objects [6], [35] or indicators showing where more data is embedded into the diagram [6], [32].

Size and length were used to show ordered variables. The most common use of these channels was on the thickness or length of the links or nodes to show variables, such as frequency or duration [2], [4], [16], [21], [27], [30]–[32], [36]. Also, size and length were used when additional charts were integrated into the node-link diagram, such as bar/triangle charts, pie charts or treemaps [18], [32], [33], [36]. In one paper the increase in size highlighted a selected node [6]. Font size of labels was explored as an option to visualize differences between processes [7].

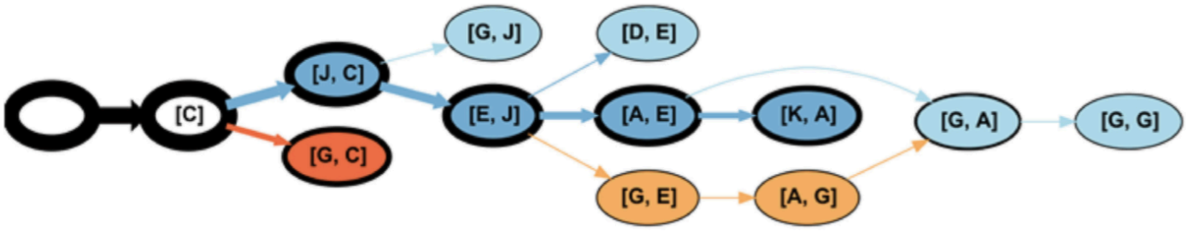


Figure 12. *Process Comparator uses shape, size, color hue and saturation. Shapes of link endings show the direction, thickness of the links and node borders show frequency, shades of red show negative effect size and shades of blue positive effect size. Image from* [2]

Spatial position was shown either on a common scale, for example placing a time axis to the process diagram [4], [32], or on an unaligned scale, for example in the case of weighted node-link diagrams to visualize clusters [19], [20]. Another way to take advantage of the spatial position is to use animation, where the tokens of the instances flow over the edges of the map and communicate meaning through their location on the map [3].

Transparency was used once to express values of frequency (see figure 4) [32]. It was also used to manage occlusion (see figure 8) [33].

Interaction design was generally not as common topic as the descriptions of visual design in the state of art studies. The interactive functionalities that were mentioned were panning [16], [22], zoom [16], [22], [33], brushing [16], aggregation [3], [33], filtering [3], [18], [22], [27], [33], [35] and supporting tools for animation, such as regulation of speed and stop/play controls [33]. User actions that were specified in the studies, were hover and click. Hover was used to reveal embedded detailed data or to show labels [21], [30], [33]. Click was also used to reveal embedded detailed data [2], [6], [32], [36]. In addition, it was used as a shortcut on the diagram for event filtering [3], [38]. One diagram used interactive linked highlighting – when a node is selected on one diagram, it is also highlighted on other diagrams that are shown in the same view (see figure 11) [16]. Besides manipulating the dataset by filtering functionality, some diagrams allowed the user to change visual encoding – a color scheme [33], the base chart [33], visual channels expressing the values [16], [17], rotation of the axis of the diagram [22], faceting [28] and alignment [17].

3.5.3.3 Summary of Process Mining Visualizations

Eleven different types of diagrams were presented in the studies of state of art research: a node-link diagram, a treemap, a pie chart, a bar chart, a line chart, a parallel coordinate plot, a box plot, a scatterplot, a gauge chart, a graph for an online instance stream and a trace map drawn with turtle graphics. The most popular type of a diagram was a node-link diagram, which was shown in 24 papers. It was used either to show relations between activities and the topology of the process, or else, relations between resources. The first used a process diagram, while the second used a weighted network diagram or a chord chart.

All the node-link diagrams in various studies were presented in a different way, using a unique combination of visual and interactive elements to express the data. It shows that the design of node-link diagrams for process mining does not have a set standard – the developers and designers of the tools use various visual channels to communicate different attributes. This means that the development of process mining tools and diagrams includes creative, complex and experimental design tasks.

3.5.4 RQ3 How are the design choices for process mining visualizations made?

Most of the studies were with a practical outlook and therefore used the article space to present the results of the design without going into the details of the design process and decision-making. Even so, fragments of the thought process behind the design decisions were collected where possible. As the arguments were often well hidden into the text, the following section does not attempt to give a comprehensive list of reasoning, but rather give a general overview of the sources the authors based their ideation on. Section 3.5.4.1 describes various ways the reasoning of design decisions was done in the study papers. 3.5.4.2 summarizes the findings that answer the third research question.

3.5.4.1 Reasoning for Design Choices

Only one paper in the state of art research studies refers clearly to a systematic design process framework – Design Science Methodology [18]. The authors have adjusted the methodology to their project. They include three sources for knowledge input to their design process: process mining knowledge, visualization principles and evaluation of visualizations [18]. In the search of visualization theory principles the authors have gone great lengths and picked input from several scattered sources. It seems they did not find one clear source to get input for visualization design principles in process mining.

Other papers have not based their design process on a specific framework or a methodology. Even so, they have gathered input from various sources and based some of their design decisions on this input. The sources can be divided to four types: existing practice, domain input, theory and logical argumentation. Each type with examples is described separately below.

Existing practice source was used when a literature review or a comparison with existing tools was conducted to uncover the shortcomings or draw inspiration from the existing solutions and approaches. The first option – uncovering the shortcomings – was more common than building on existing visual solutions from the industry. The papers usually described the external context overview through the lens of criticism in order to justify the development of their own work. For instance, Bachhofner et al) [5] pointing out that current solutions visualize only one performance metric at the time on the process map, hence, there is need for visualizing several performance metrics on the same map. In some cases several papers use the same arguments, for example, both Bolt et al [2] and Pini et al [32] claim that current comparison tools do not take in consideration differences in performance metrics. Some papers went a step further from a general overview and took specific features of existing tools as a benchmark to compare their own tool against [3], [19], [21]. This way the differences in the design were clearly highlighted on a very concrete level. Some studies mixed the critical and inspirational lens and used the existing solutions as a base to their own design [3], [7], [18], [26], [31].

Domain input refers to design decisions that are based on real-life tasks, requirements or user feedback. Five papers listed clear requirements [3], [6], [18], [30], [32] and two added use case descriptions [16], [36]. Some works were based on a real-life case study [3], [6], [30]. Several studies used user feedback as an evaluation tool [6], [7], [16], [18], [30], [32], [33], but only some of them used it to gather design ideas for further development [16], [18], [30], [33]. One tool for care process analysis can be brought out as a good example of a domain input led process – the tool is built for a specific hospital and the primary users were involved in the project before and during the development process [30]. In fact, user feedback was so influential in this project that the first version of the tool was completely discarded after the visualizations did not prove to be as helpful as expected [30].

Theoretical sources were the least common input for the design decisions. Usually, theoretical basis was presented in detailed choices, such as selecting visual channels based on the variability they can show [18] or the reason for adding control elements for an animated playback [33]. However, some papers had extracted general visualization principles, such as "[visualized] variable has to preserve the structure of underlying data" [5], or referred to larger research bases, such as graph theory [20] or research for coordinate multiple views [35].

The most common source used for explaining design decisions was logical argumentation. Some argumentations were simply general opinions, for example "by watching the displays' content and simultaneously performing selection on the business process model, … differences in the selected sets of data become intuitively visible …".[35] or "we chose this representation because it makes comparisons more natural for the user" [25]. Other arguments stemmed from visualization problems - a design decision was taken because it seemed like a good solution to solve some kind of a problem, such as using transparency to manage occlusion [33]. Typically such arguments were given without exploring any alternative options, common practices or supportive theory.

3.5.4.2 Summary of Reasoning of Design Choices

Overall, one study followed a structured scientifically verified approach in their design process. Other studies included various types of inputs to their decision-making unsystematically. Inputs for decision-making were existing practices, domain input, theory and logical argumentation. This list is not comprehensive because the reasoning for different design choices was available only in a fragmented way and some reasons that were expressed in a subtle manner may have been overlooked.

3.5.5 Conclusion of the State of Art Research

The answer to the first research question showed high importance of visualizations in process mining field. Nearly all process mining techniques (according to the list by Oruste [11]) were visualized in the state of art research studies. Some visualizations were direct outputs of a technique, for example a process model for a process discovery, while others were mentioned as an input for a technique, such as visualizations for performance analysis and process comparison.

The answer to the second research question revealed the most used visualization for process mining techniques – a node-link diagram. In addition, the listing of various visual and interactive elements showed that the node-link diagrams do not have a set standard in process mining and therefore, design of such diagrams requires execution of complex design tasks.

The findings from the third answer showed that most of studies do not use any framework or methodology to support their design process. Instead, design choices are based on multiple sources, such as domain input, existing practices, theory and logical argumentation. The logical argumentation is the most popular way to reason the design decisions, while theory the least common method.

In conclusion, the visualizations are essential to various process mining techniques (RQ1) and the design process of the visualizations is complex (RQ2). Even so, oftentimes the practitioners do not apply a systematic approach to the visualization design (RQ3). This shows a gap in the visualization design practices in process mining field – a complex and important aspect of tool development is not supported by any structured guidance. Hence, there is a need for visualization framework that is specifically tailored for designing process mining diagrams.

4. Framework

The following chapter introduces a framework for guiding developers of process mining tools in data visualization decision-making. The framework is built to increase design awareness and raise effectiveness of visualizations in process mining. The scope and high-level goals of the framework are described in section 4.1 and the basis of the structure and content of the framework are introduced in section 4.2. The process of developing the framework is described in section 4.3. Section 4.4 defines and reasons each part of the framework. The framework can be found in full length with instructions in Appendix II and III.

4.1 Scope and Goals of the Framework

The nature of a design process is complex as there are many aspects that shape it. The aim of the framework is not to facilitate the whole design process, but to focus on the gap the state of art research detected – the lack of systematic reasoning of design decisions in the visualizations developed for process mining techniques. This finding narrows down the scope in two ways – firstly, which field and users the framework is developed for (section 4.1.1), and secondly, which section of the design process is improved by the framework (section 4.1.2). Other constraints on the framework are described in section 4.1.3 and the main goals in section 4.1.4.

4.1.1 Context and Users

The framework is developed for process mining techniques and it is meant to be used, when designing diagrams for those techniques. The high-level framework that identifies areas of design decisions can be used for any diagram for process mining, including charts for performance metrics and networks for organizational mining. The detailed version of the framework, which includes sub-questions and alternative solutions with their strengths and weaknesses (see Appendix III), is tailored for designing process maps. This focus is chosen for two reasons. Firstly, there is plenty of in-depth work on designing performance dashboards [39]–[41], whereas limited guidance in designing process maps [42]. Secondly, the state of art research revealed a heavy use of this type of diagram – most of process mining techniques require understanding of the topology of processes, which is usually aided by node-link diagrams, i.e. process maps. The expansion of the framework to all types of diagrams can be done in future development of the framework.

The primary audience is developers of process mining algorithms, who do not have professional experience in design field. In addition, the framework can be used by designers, who do not have previous experience in designing process mining visualizations.

4.1.2 Design Process

The framework aims to improve the systematic reasoning of design decisions made for visualizing process diagrams. This boarders the section of the design process, where the framework can be used (see figure 13).

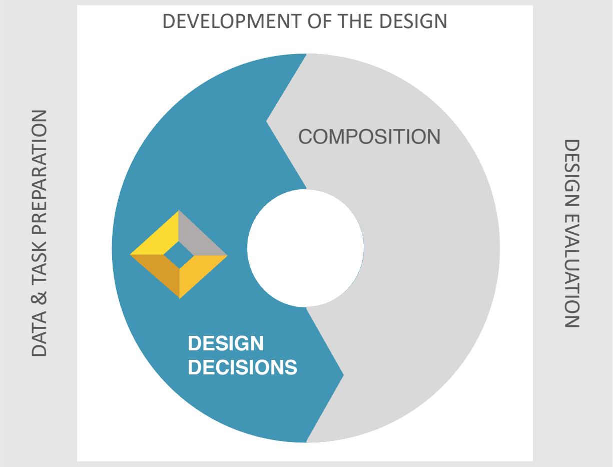


Figure 13. *The scope of the framework – design decisions are in the scope (blue), while composition, preparation of data and tasks and design evaluation are out of the scope (grey).*

Data visualization process is a design process, which starts from the need to visualize data for presentational or analytical purposes. The design process has three steps – data and task preparation, design development and design evaluation.

Data and tasks are prepared by identifying possible user tasks and selecting and/or deriving data that helps to achieve those tasks. For example, in process discovery the aim is to visualize process activities and succession of the activities, so that the analyst could study the topology of the process. It is expected that this preparation is done before the framework is used, meaning that the developer should know what data needs to be visualized and which core purposes is the diagram designed for.

The next step of the design process is to transform data and tasks into an effective visualization. Design development has two parts. It starts form agreeing on a set of design rules, and then composing those rules into a visualization. For example, at first it is decided that the frequency of activities is shown with shades of blue on the nodes and average duration with shades of red. Secondly, those decisions are implemented on an actual diagram, either hand-drawn or coded. Once the full composition is sketched out, conflicts between separate visual mappings are revealed and design decisions must be revised. For instance, shades of blue and red cannot be shown simultaneously in an overlapping way on the activity nodes. The designer has to go back to the design decisions and specify the choices by either changing the visual channels or find a way how to facet the encoding, so that it would not overlap. Therefore, the composition and design decisions are shown as an iterative process.

The framework aims to improve systematic reasoning behind design decisions. Hence, it is placed to the part of the design process, where designer considers various options for visualization rules – the section named “design decisions”. This placement clarifies the output of the framework – a set of decisions that a designer can use to compose a visualization, not a ready-made composition or a mock-up.

The last part of the design process is to evaluate the composed visualization. This is out of the scope of the framework. The framework does not intend to help with validating the quality of process mining diagrams, it only helps with the design development part. There are several existing works dedicated to validation techniques for data visualization in academic field, such as [8], [43], [44].

4.1.3 Other Constraints

Some aspects of design were excluded or constrained in the framework. This was done to keep the framework concise and to focus only on the most important areas of design. Constraints described in this section can be included in further development of the framework.

Process logs can be either static or dynamic. Static dataset is complete and fully available when it is visualized, while dynamic dataset changes in real time, adding new or updated data to the visualization [37]. Dynamic process logs visualizations are not very common. Only two studies from the state of art research were handling dynamic datasets, whereas only one of them used a process map for visualization [23], [26]. Therefore, the framework is built for more common practice of the two – static datasets.

The framework touches visual encoding as well as interactivity aspects of the design, but it does not intend to help with the design of user interface. User interface aspects of the framework are limited to identifying few elements and functionalities of user interface that are relevant to the diagram design, such as widgets for control panel or navigation bars between several views, but does not attempt to give input on how to design those elements.

Process maps are usually designed as two-dimensional diagrams. 3D visualization has been explored in process mining [18], but is not recommended because of the problems that arise from perspective distortion and occlusion [37]. The elements on the map are not perceived precisely if they are placed in a three-dimensional space. The framework does not include any 3D design requirements and therefore expects the designer to develop the visualization in a 2D space.

The framework is meant for designing diagrams for laptop and desktop use, because the process mining visualizations that are currently available are targeted to be used on those devices. Process diagrams get quickly large and managing those complicated diagrams on handheld devices brings a whole new set of challenges that are beyond the scope of this work. Therefore, characteristic elements and limitations of touchscreens are left for further development of the framework and are not included to the current version.

The current format of the framework is not digital – it is meant to be used on a hard copy. The focus of this thesis is to develop content for the framework. Therefore, the format is secondary and will be improved when the framework is developed further.

4.1.4 Goals

The framework targets several goals. Three most important overarching goals are described in this section and accompanied with high-level explanations how these goals are met in the framework.

The first goal of the framework is to raise awareness of design decisions – which questions need to be asked in order to design a thought-through diagram. The framework identifies necessary areas, starting from high-level questions about encoding and interaction in general, and breaking them down to detailed sub-questions tailored for process diagrams. In addition to hierarchical sequencing of the questions, the linear sequencing of sub-questions is added to help the user orient and navigate in the framework structure.

The second goal is to aid the user in justifying the decisions. The framework functions as a tool to give confidence to the user by allowing them to consider trade-offs each design decision brings. This is done by giving alternative answers to sub-questions and supporting the decision-making with explanations of strengths and weaknesses of those alternatives where possible.

The third goal is to ensure the comprehensibility of the design decisions. The users can come from either coding or design background, therefore some of the design or process mining terminology can be unfamiliar to users. This problem is reduced by adding illustrative visualizations for alternative answers.

4.2 Basis of the Framework

The framework is based on the findings from the studies that are introduced in the state of art section. The studies propose many examples and aspects to keep in mind in the design process, but do not provide enough information to shape a comprehensive framework out of it. Therefore, another cornerstone is selected – data visualization theory. It is briefly introduced in this section.

The base principles of visualization theory are retrieved from work developed by Tamara Munzner [37]. Her theory is chosen for several purposes. Firstly, the book “Visualization & Design” [37] proposes an overarching framework for designing and analyzing data visualizations. It touches all the aspects of the process from domain and data analysis to validation. The main focus of the book is on the same part of the design process as the framework of this thesis – how to visualize data. Secondly, Munzner has built her work on the foundation of existing data visualization theory, including works by Leland Wilkinson [45], Edward Tufte [46] and Collin Ware [47]. Lastly, Munzner’s works have been cited nearly seven thousand times according to Google Scholar, placing her amongst the top data visualization theory authors.

Munzner presents data visualization process as a nested model, where the output of one layer is an input to another [37]. The four layers are domain situation, data/task abstraction, visual encoding/interaction idiom and algorithm [37]. The question “how to visualize?”, which is the interest of the framework of this thesis, lays in the third layer – visual encoding and interaction idiom. Munzner breaks that question down to several sub-questions and each of those sub-questions have their own sub-questions forming an hierarchical tree of design decisions [37]. Some, but not all questions are answered with possible solutions, theory behind the preferred options and specific examples from the data visualization field.

The theory provides a method to analyze any data visualization with any scope and content. This ambition makes it a good guidance for designers, who want to expand their awareness of different visualization possibilities for various purposes and fields. It is not a good source for developers, who want to design a diagram for a specific purpose and context, such as process mining. The spectrum is wide and there is too much irrelevant information if one is searching help for a specific task. Going through all the material before composing a single visualization would be more confusing than constructive for a developer. Therefore, in the framework of this thesis, relevant information is extracted from the general theory and adjusted according to process mining visualization practices in order to enable a quick and relevant overview.

4.3 Process of Developing the Framework

The process of developing the framework followed the sequence of goals listed in section 4.1.4. The first step was to identify the areas of design decisions and order the identified areas and questions into a clear narrative to guide the user in the sequence of the process (section 4.3.1). The second step addressed the issue of lack of justified decision-making by listing several possible answers and their trade-offs to each question (section 4.3.2). The last step improved the comprehension of the framework by adding visual illustrations to the alternative answers (section 4.3.3).

4.3.1 Identifying Areas of Design Decisions

The areas of design decisions for visualizing process mining outputs depend on design theory as well as process mining visualization practices. The design theory brings to focus general decision points that might be overlooked when designing for a narrow purpose. The process mining examples bring to focus decision points that are characteristic to process mining and are too detailed to be covered in a general theory.

Firstly, design questions were extracted from the Munzner's theory. The questions were extracted from all the main sections of the book – "What?", "Why?", "How?", "Evaluation" and "8 Rules of Thumb" [37]. They were captured in their hierarchical structure. The primary interest of the framework was in the section "How?". The rest was extracted to not to miss out relevant questions that were placed in a different section of the book. The first set of questions included 121 questions in total.

Secondly, information about various process mining techniques was gathered. Various types of techniques with their definitions and aims were taken from a literature review on process mining techniques by Oruste [11]. The set included eleven different techniques. In addition, examples of the diagrams were gathered from the state of art studies. Also, commercial tools Disco [48] and Celonis [49] were explored to enrich the pool of examples with commercial designs. These tools were chosen because several papers of the literature review mentioned those specific tools [3], [29], [18], [35] and their trial versions are accessible for free.

The data visualization questions and the visualization practices of process mining techniques were combined to identify relevant questions of the framework. Data visualization theory questions were measured against process mining visualization practices – only the questions that were applicable to process mining techniques were selected. Additional questions were included when there was no question to cover an essential design aspect of diagrams used in process mining techniques.

Finally, the scope of the framework was taken in consideration, and only the questions that fit into the scope were selected (section 4.1). The focus of the framework – process flow diagrams – shaped the sub-questions, whereas the high-level design questions remained technique-independent to enable further development of the framework. In the final selection, 62 questions remained relevant to the framework.

Due to the selecting and adding the questions, the structure of the hierarchy of the questions was not complete anymore. The questions required reorganising and simplification of the structure. This was done with a top-down approach. Two main areas were identified – encoding and interaction. Each of these areas was divided into two – encoding was divided into arrange and map, interaction was divided into reduce and change. The rest of the questions were divided between those four sub-questions.

After the dividing, the dependencies between questions were identified in each section, i.e. one question cannot be answered before some other decisions are already made, for example the decision about the base of the diagram must be done before designing details. These dependencies defined the sequence and hierarchy of the questions in most of cases. The questions that were dependent on one another, were placed nearby and in the sequence they required. In cases without dependencies to rely on, the initial data visualization theory version of the sequence [37] was applied to the order and hierarchy of the questions.

4.3.2 Identifying Answers

Once the questions and their sequence were identified, the second goal – helping the user to justify the decisions – was addressed by giving the questions alternative answers. Most of the answers were extracted from the data visualization theory [37]. If the theory did not offer an answer that could fit to the process mining context or the answer was missing, the examples from the process mining diagrams were used to identify potential answers. In some cases, supporting theoretical material was searched if neither of the base sources offered relevant answers (see section 4.4).

The answers were either mutually exclusive or complementing one another. The marking for the first type of answers is radio buttons and second type is marked with select-boxes. Some answers have to be filled in by the user – he/she has to identify or categorize attributes or visual elements that are relevant to their dataset. This type of answers are marked with additional ": …" in the end of the answer. Most of the answers include the option "other" to emphasize openness of a design process – there are suggestions, but no strict rules and all the questions can be solved differently if the purpose is to experiment with the visualizations.

The first two questions were answered because the rest of the framework was built according to those choices. However, the alternatives were given, so that the user can justify the choice or experiment with completely different type of visualizations if needed.

The strengths and weaknesses were extracted together with the specific answers from the visualisation theory or inspired by general principles from the theory [37]. In cases, where dualistic pros and cons were irrelevant, common practices with brief reasoning were listed instead of theoretical trade-offs. The common practices were taken from the diagrams of the literature review studies as well as the selected commercial tools [48], [49].

4.3.3 Adding Illustrations

Supporting visualizations were added, where answers needed to be explained due to the industry-dependent terminology. All the drawings are inspired by the examples from the state of art studies, commercial tools or the data visualization theory. Certain examples were inspired by modelling languages, such as BPMN [50]. The visual examples are mostly used in the encoding section because the interaction cannot be clearly captured on static non-interactive images. In the further development of the framework, interactive examples can be included.

4.4 Structure and Content of the Framework

The following chapter introduces the framework and the reasoning behind each section of the framework. Figure 14 illustrates the first three levels of the framework hierarchy.

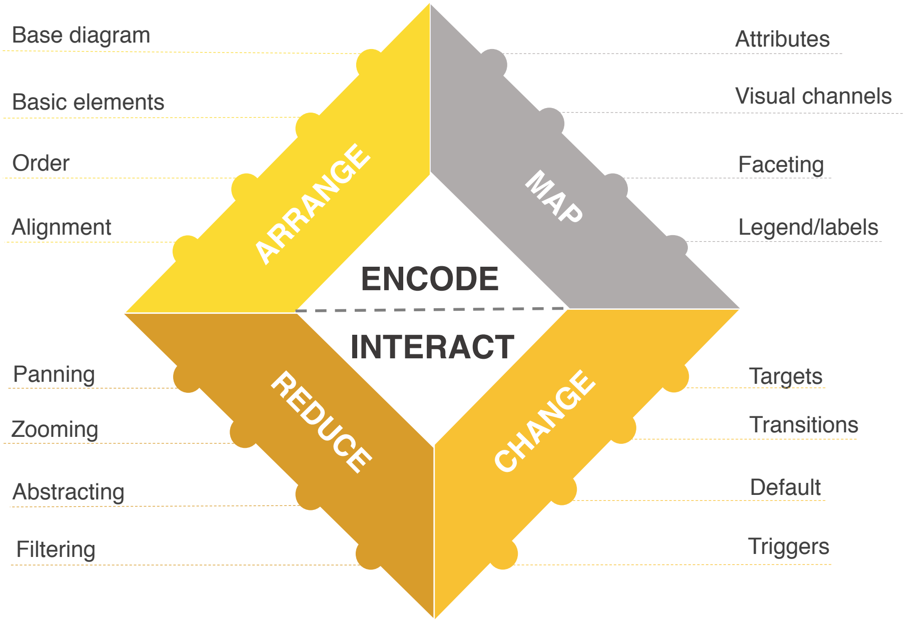


Figure 14. *The base layout of the framework. The model can be read inside out – the highest level design questions are in the centre, the second level questions in the colored boxes and the third level listed aside.*

The overarching question – how to design data? – is first divided into two areas – encoding (section 4.4.1) and interaction (section 4.4.2). Encoding means making design decisions about what is seen and interaction addresses questions about how the visualization can be manipulated by the user.

Both, encoding and interaction are divided in two. Encoding is a sum of questions about arrangement (section 4.4.1.1) and mapping of the data (section 4.4.1.2). Arranging section helps to decide on the basic structure, order and alignment of the diagram. Mapping focuses on assigning visual channels as well as organizing them on the diagram. It also includes a section about how does the user know the meaning of the mapping – placing a legend and/or labels.

Interaction includes questions about changing (section 4.4.2.1) and reducing (section 4.4.2.2) the data by the user. Changing brings out questions about what can be changed, how the change appears, what is the default look and how the user can trigger the changes. Reduction section introduces possible ways to see less data on the view. It includes panning, zooming, abstracting and filtering.

Table 2. Hierarchy of the Questions in Process Visualization Framework

Level 1	*Level 2*	*Level 3*
How to encode data?	How to arrange data?	What is the base diagram? What are the basic elements of the diagram? How are the basic elements ordered? How is the diagram aligned?
	How to map data?	Which attributes are shown on the diagram? Which channels express the attributes? How is the data faceted on the diagram? How does the user know the meaning of the channels?
How to design interaction?	How can the user change the visualization?	What can be changed on the diagram? How do the changes appear? What is the default appearance? How can the changes be triggered?
	How can the user reduce data?	Does the diagram need panning? Does the diagram need zooming? Does the diagram need abstracting? Does the diagram need filtering?

Table 2 shows the hierarchy of questions in the framework. It is aligned with the keywords presented in figure 14. The framework should be used by answering all the lowest level questions, because an answer of each high-level question is the sum of the answers of its sub-questions. For example, the question “how to arrange data?” is answered when the designer has answered all its sub-questions - “what is the base diagram?”, “what are the basic elements of the diagram?”, “how are the basic elements ordered?” and “how is the diagram aligned?”. For further instructions see Appendix II.

In the full framework (see Appendix III), the third level questions are also broken down to more detailed sub-questions to simplify the use of the framework. In addition, specific alternative proposals for solutions are given for each deeper level sub-question. The lower level questions and answers are focused only on diagrams visualizing process flows (reasoning of this focus is given in the scope description, section 4.1.1).

In the following sections each question of the framework is described in further detail (section 4.4.1, 4.4.2). Section 4.4.3 summarizes the detailed description of the framework design. The main body of the thesis does not include the full unfolded version of the framework, it is attached in Appendix III.

4.4.1 How to encode data?

Working with event logs means navigating in a vast amount of detailed data. Visual analytics field aims to transform the data overload into an opportunity to gain deeper insights about various phenomena [4]. Visual encoding is the core of transforming data into a visual form to achieve that goal.

Visualizations and semantics are not considered a strong aspect in process mining tools [4]–[7]. Therefore, the first question is chosen to draw attention to this basic shortcoming in current solutions. The sub-questions help to arrange the layout of the visualization –

arrangement (section 4.4.1.1) – and assign meaning to visual elements on the diagram – mapping (section 4.4.1.2).

4.4.1.1 How to arrange data?

Separating data arrangement from mapping is inspired by the design principle of separating content from presentation [51]. Firstly, the building blocks of the diagram are agreed upon and organized, which clarifies the abstract version of the diagram. Secondly, the visual channels determine the actual looks of the diagram, which transforms the abstract version into a concrete visualization. In the context of this framework, the separation is not between the data and the visualization, but between the abstract and concrete versions of the diagram. For example, firstly it is decided that the diagram consists of nodes and connecting marks, and secondly, it is decided what are the visual rules that determine the aesthetics – the thickness of links, shade of the nodes etc.

The mapping depends strongly on the arrangement, because decisions in the structure constrain the visual channels. For example, a bar chart and a node-link diagram allow completely different sets of visual channels to convey values of ordinal data – a bar chart uses mostly the length of the bars [37] versus node-link diagram can communicate similar ordinal values with color saturation [2]. This is the reason why arrangement is placed before mapping in the framework.

What is the base diagram?
This sub-question presents possible solutions for the basic structure of the diagram. This decision shapes all the other answers in the framework, because different diagram types bring different design opportunities.

There are three ways to visualize network data – a node-link diagram, adjacency matrix and enclosure [37]. A node-link diagram is marked as a default option for process flow maps, because this type of a diagram is best suitable for tasks that require understanding of topology of the data [37]. The rest of the framework is based on this decision. Other options are brought out to help the designer justify this decision through alternatives or in case he/she wants to experiment with different types of process model visualizations. Enclosure and adjacency matrix are rare, but not completely foreign in process mining, for example Plans Strips visualization uses enclosure diagram to portray medical processes (see figure 7) [4].

What are the basic elements of the diagram?
The choice of a node-link diagram identifies the basic elements of the diagram – nodes and links. Nodes are items and links are connection marks between the items [37]. One important decision on the structural level about the nodes and links is if they are separated or merged. The separated version is more common, but both have been used in process mining field, sometimes even in the same tool – merged nodes and links show high level process flow, when separated version shows a more complex version of the process, allowing the user to drill down to details [27].

The framework is helping to design process flows with separated nodes and links as they are more complex and require more attentive design. The complexity lays in two reasons – a separated version allows the use of visual channels on nodes as well as on links and the spatial layout of the separated version requires a careful ordering and alignment to avoid occlusion and visual clutter. The default option is chosen for the user as the rest of the framework is built around those complexities. However, the answer is given as a multiple choice option, so the user can add the merged version as an additional layer if necessary.

How are the basic elements ordered?
A process can be shown as a sequential flow or a hierarchical set of actions. A comparative study on those two types of maps suggests their complementary nature – using both maps gives a more comprehensive overview on the process than using only one [52]. Therefore, the options are given as a multiple choice.

The sequential map is chosen as a default option and is explored with additional guiding questions because it is more often used in process mining than the hierarchical maps [52]. In the state of art study, only two process models were presented in a hierarchical structure, all the others in a linear process flow.

The sequential nature of a process can be emphasized in several ways. Three common ways extracted from the state of art research are listed as options – orientation of the diagram, directional shapes of diagram elements and special encoding for the start and end of the process. The orientation can be from left to right or from up to down. Commercial tools use both up to down orientation [48], [49], when most of academic tools studied in the state of art research propose left to right layout. The sequence can be shown with directional shapes of diagram elements. It is usually done by placing an arrow to the end of the connection marks, but can be also done by shaping the nodes as arrows [35], [27]. A simple way to apply the third option – differentiating the start and end – is to start and end the diagram with a circular unlabelled nodes, as it is done in Disco [48] and Celonis [49] as well as academic tools, such as ProcessProfiler3D [18] and Inductive Visual Miner [3].

How is the diagram aligned?

The elements of the diagram are separated by different actions and ordered by their sequence of appearance in the process (or hierarchy if an hierarchical model is chosen). The alignment is optional, but as there is a high risk of occlusion and cognitive load for the user in complex process models, intentional planning of the alignment can increase the scalability of the node-link diagrams [37].

The first step to plan the alignment is to know what needs to be aligned – is it just elements in one process model or are there diagrams of several processes. Visualizations for detecting differences in processes were the most researched and tested area in process mining visualizations according to the state of art research. Several processes are aligned in process comparison as well as conformance checking [11].

The differences of two or more models can be either juxtaposed in a side-by-side or matrix view [28] or else, superimposed into one view [7]. Several models can be also split to completely different views [37], but this is not recommended for comparison purposes. It would impose a great cognitive load on the user as he/she has to rely on their memory to be able to compare the processes in different views. Instead, this option could be used for showing different views on the same process – for example one hierarchical and one sequential version of the process shown in different views [30].

The rest of the questions about alignment are for both, aligning diagram elements as well as aligning several diagrams. The first question addresses the semantic meaning of the alignment – what is the alignment based on? The most common way to align node-link diagrams is to do it by the best fit to the screen, which brings a downside of irregularity in the meaning of proximity of nodes – sometimes it carries a meaning, but sometimes not [37]. Another way is to intentionally encode meaning into the alignment, for example showing parallel activities side-by-side [32] or separating activities into resource pools as it is done in BPMN language [50]. This organizes the diagram into an easy-to-follow format, but might be space-consuming for complex processes.

Another aspect that Munzer brings out about network diagrams is the choice of non-deterministic or deterministic layout – does the layout of the process model change when the event log is reloaded or does it stay the same [37]. Deterministic layout takes advantage of the user's visual memory and makes it easy to refer to elements on the process diagram, such as "the stuff in the upper left corner" [37], but it is more complex to code than a proximity based layout.

4.4.1.2 How to map data?

Visual mapping means deciding on which visual vocabulary is used to present values of specific data points [53]. Mapping is necessary, when the designer wants to convey more data from the event log than bare basic topology. For example, if two processes are compared, it can be done with just basic side-by-side maps of the process models or it can be mapped with a visual pop-out of differences in the process flows [16].

The outcome of this part of the framework is a bundle of rules that will determine the looks of the diagram based on the given data. For instance, it is not decided that the thickness of the connecting mark is 2.5 points, but the decision is about how the line width can show different values, such as aligning line width with different levels of throughput [36].

Analytical tasks often require analyzing more attributes than available channels can visualize in one view. It is common to assign same channels for several attributes [25] or embed detailed data into an overview diagram [4]. This requires splitting the mapping into multiple views or layers [37]. Last questions of this section help to decide on how to facet all the necessary data and how to help the user understand the complex visual landscape.

Which attributes are shown on the diagram?
It is expected from the user of the framework to already know which data they want to encode (see section 4.1.2). In this section, the selected data should be divided to categorical and ordered values. Categorical values mean that the values can be grouped by levels, but these groups cannot be ordered, such as process resources. Ordered values can be put in a relative or absolute sequence such as timestamp or throughput. The classification is necessary to be able to assign the selected attributes right visual channels [37].

Which channels express the attributes?
The designer has to assign visual channels to attributes according their type – categorical attributes are shown by identity channels and ordered attributes by magnitude channels [37]. Both types have specific channels that are ranked according to their effectiveness. Figure 15 shows Munzner's ranking without 3D channels as 3D design is not included to the framework (see 4.1.3) [37]. More important attributes are usually mapped with more effective channels.

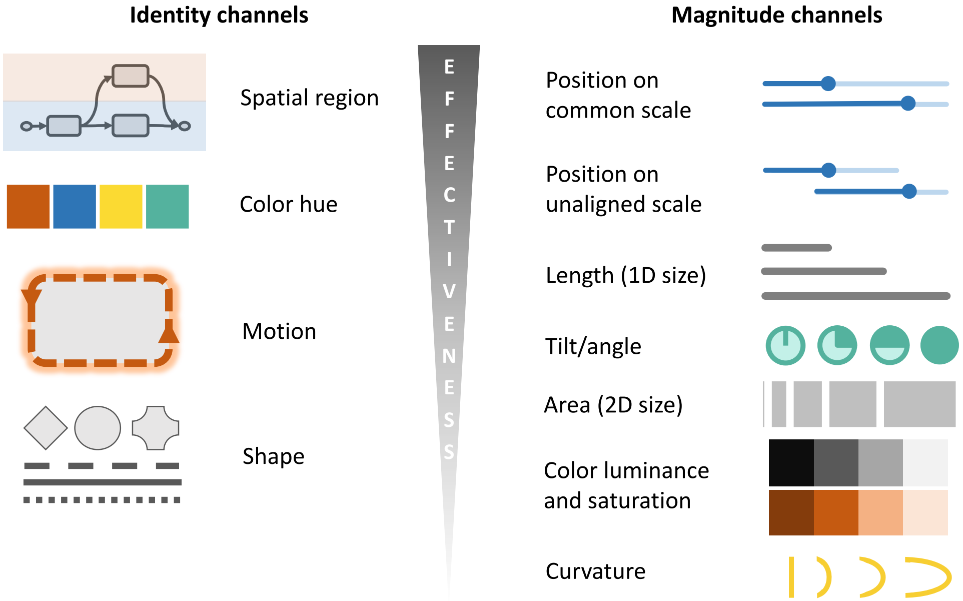


Figure 15. *Ordered attributes are visualized through magnitude channels and categorical attributes through identity channels. Channels are ordered by their effectiveness, more effective on the top and least effective in the bottom* [37].

Munzner's synthesized list and order of the channels is based on many previous perception studies. For example, the psychophysical power law of Stevens, which proves a difference between the perception and actual intensity of channels – some channels are magnified more, when the intensity is increased, such as saturation, and some channels are perceived less intense than they are when they are increased, such as area and brightness [54]. Length is one measure that is considered completely accurate – if the length of a line is increased two times, it is perceived twice as long as the first version [54]. Another study Munzner has incorporated measures the error rate when using magnitude channels [55]. The results of the initial study as well as the re-testing by different scientists decades later [56] match with the Steven's findings and support the final ranking – the most effective magnitude channels are measures of lengths, then angle, followed by area, color luminance and saturation and curvature (figure 15).

The practice of the process mining visualizations does not always follow the suggestions of the visualization theory. Therefore, example solutions from the process mining tools are brought out for the most popular channels. For identity channels, the popular channels are shape, spatial region and color hue, for magnitude channels, it is color luminance and saturation, length, area and positioning on common or unaligned scales. The common practices are supported with the reasoning, why certain channels are preferred over the others even though they are less effective. The argumentation is inspired by general design concerns pointed out by Munzner, such as a concern for limited screen-space and colliding visual channels [37].

There is a third option to encode the attributes in addition to visual channels – textual sets. This is included because of the common use of listed statistics as an overview of the whole process [36] or about each element embedded into the process diagram, such as pop-up windows in Disco [48] and Celonis [49].

How is the data faceted on the diagram?
Often there is more data to be shown than what can be seen in one view. One way to handle complexity and richness of data is to facet the display in multiple views or layers [37]. Three ways to facet data in process models is to superimpose, embed or show data in a separate area of the view. The designer has to choose, which way of faceting is the best for each attribute he/she has selected for visualizing.

Superimposed layers are often used to visualize performance metrics on process models. For example, in a tool InterPretA saturation channel is assigned to show three different variables – fit of the model, throughput and time –, whereas first two are configurable to obtain even more views [25]. The layers have the same visualization, but the data that is shown changes. To make the changes visible and clear to the user, a categorical channel of color hue is assigned to each layer – fit of the model is in shades of green, throughput in shades of blue and time in red [25]. The first task designer has when superimposing, is to identify channels and attributes of each layer.

One specialized layer is an animated layer that needs separate attention. Munzner does not recommend animation visualization, except for short transitions, because it imposes a strong cognitive load on the user – the user has to remember the states of the process during the flow of animation in order to extract important information [37]. Despite of this recommendation, the question is included, because it is a tool that is often used in process mining. It is used to show the flow of process instances through different states and activities of the process [3], [33]. The memory aspect is reduced by including playback control widgets, so the user can access any point in the animation and capture still images of it.

Another way to facet data is to embed it [37]. In the context of the framework, embedding means eliding information that can be interactively revealed and hid again. In commercial tools, this is used to make the detailed statistics about diagram elements accessible to the user. The questions to compose embedded data are as follows: what is embedded, where it is embedded, how does the user know that there is something embedded and where does the data appear when it is revealed. The last question needs a careful consideration, because when embedded data is revealed, more data must fit on the screen. If the data is shown on top of the diagram, some data will be hidden underneath [37]. On the other hand, if data is shown in a separate area of the view, all the data is visible, but the view becomes crammed.

“Off the diagram” option for placing excess data is not mentioned in Munzner’s theory, but is extracted from the practice. Some general statistics of the whole process can be shown in a separate area of the diagram view, such as number of instances or average processing time. It is a common practice in commercial tools, but was also used in some state of art studies [23], [30].

How does the user know the meaning of channels?
Once, the designer has made decisions about what to show and how to show it, it is important to take a look at the diagram from the user’s perspective – how does the user know what is meant by those different visualizations. This topic was not raised in the Munzner’s framework, but was included because labelling and legends are commonly used on process diagrams.

The first question about the legend aims to identify, which values and visual channels must be shown on the legend. The second question is about the placement of the legend. A conventional placement is nearby the diagram as a separate pane, which is often accompanying static non-interactive diagrams [57]. Interactive visualizations have a wider variety of options, for example, in process mining tools, the legend can be integrated into the control panel [25]. If the designer still prefers a conventional way, there is a space-conserving option – a dynamic legend, which shows channels and attributes that are relevant only to the selected layer of the diagram [57].

Diagrams and visual channels are good for detecting patterns, but often do not generate trust for precision tasks. For example, in a development process of a process mining tool in medical field, users asked the data to be shown in a table format in addition to the graphical layouts [30]. Labels can help to show precise values and make the diagram function also as a source for extracting detailed data.

The questions about labels, include identifying which labels are visible and when, where are the labels placed and how to ensure the readability. The last question is a special concern that has not been successfully dealt with in practice – the labels are usually in a fixed size and when a complex diagram is visualized, the labels are too small to be comfortably readable. This problem is addressed with some potential answers that are extracted from Munzner's framework, such as semantic zooming or using magnified glass functionality [37].

4.4.2 How to design interaction?

Event log data can provide several insights about a process. A single view, that is selected by the designer of the diagram, cannot include all the possible angles the user may need for the analytical tasks [37]. The visualization design has to enable the user to explore data in various ways. Interaction design helps to facilitate changing as well as reducing or increasing the amount of data that is seen.

The main concern of interaction design is to come up with interactive solutions that are easy, effective and pleasurable for the user [9]. Interaction design lingers between data visualization and user interface design. The user interface design is a large field by itself and does not fit into the scope of this framework (see section 4.1.2). The framework includes some stepping stones in the interaction design, but does not give a comprehensive overview. Questions are included where interaction design is essential for the diagram design, such as identifying user actions that relate to changes in the visualizations. The questions single out few requirements for user interface design, but do not aim to propose specific design solutions, such as composition of a menu bar or a control panel.

4.4.2.1 How can the user change the visualization?

Changing the view helps to handle visual clutter that can rise from complexity of data – the user cannot see useful information on the diagram if it is overloaded with visual expressions [37]. In the context of this framework, change means switching the data and/or visual channels that are shown on the diagram.

Designing change, means designing the targets, transition and actions of the change. Firstly, various versions of the diagram must be clarified – which encodings can the user switch to or from. Secondly, the time aspect of the change is defined – how does the transition appear, when changing from one to another version. Thirdly, the starting point is agreed upon – the default appearance. And finally, the user actions are defined – which actions trigger changes.

What can be changed on the diagram?
Identifying changing elements is a continuation of faceting data from the previous section. Therefore, identification is divided in two groups that were already introduced before – embedded data and superimposed layers. In this section embedding and superimposing are looked from the perspective of the manipulation – how can the user access all the various encodings that the designer wants to show. The third way of faceting – placing additional data on a separate panel in the view – is not addressed here as it can be visible all the time and does not require interaction.

Different types of faceting trigger different types of changes – when embedded data is revealed, the essence of change is showing additional data on top of existing data; when superimposed layers are changed, the visualization stays similar, but attributes are changed [37].This is another reason the division is applied on the data – the designer has to take in count those differences when he/she designs change transitions or assigns user actions.

The changes that come from reducing data are designed in the last section of the framework (see 4.4.2.2). This section is focused on the changes of visual encoding.

How do the changes appear?
Whenever there is a change on a screen, there are two ways to make the change appear – jump cuts or animated changes. Often this is overlooked and it is taken for granted that the jump cut is the only way to move from one frame to another. In fact, animated transitions prove to be a better solution in some cases of data visualization. It is considered a preferred choice when a continuity and connectedness between objects and features are expected from the changes [58]. The user does not have to make the connections between before and after image himself/herself, but he/she is guided in the transition.

On the other hand, animated transitions can easily get overloaded, when many changes are happening simultaneously. It can lead to change blindness, where even significant changes are not noticed, because the focus is distracted [37]. Also, it can lead to false conclusions if the animation does not follow the semantics of the data [58]. It is a strong tool that requires a careful use.

What is the default appearance?
The default appearance is important because it is the first image that introduces the user to the process. The known information-seeking mantra for user interface design states "overview first, zoom and filter, then details on demand", meaning that the full collection of data should be presented in a summarized form before anything else, followed by zooming into interesting items and filtering out uninteresting items, and finally, allowing the user to drill down to details [59]. This suggests the default appearance to be a combination of channels and settings, which best summarize the data at hand. In this question, the user is expected to select the first overview image according to the task and data the visualization is designed for.

How can the changes be triggered?
User's journey in the change interaction is broken down to three questions: where to change; how does the user get visual feedback; and how to undo the changes. Answers to these questions determine the very base of interaction design – how is the user going to communicate with the program. Human-interaction design is a well-studied field and this

framework touches only the very surface of it by directing the designer to think about some essential questions.

The location of user actions depends on the type of faceting that is being changed. Embedded data triggers are usually placed on the diagram, such as pop-up windows that appear when clicking on diagram elements [48], [49]. Change of superimposed layers takes usually place on control panel [2], [25]. Control panel takes up space, but it gives a clear indication to the user where and what is possible to change. Changes on the diagram take less space, but it is not as clear as a control panel – the user has to experiment with different interactions on the diagram to discover the embedded data. Similar strengths and weaknesses can be applied to changes triggered by keyboard shortcuts – the user has to experiment with various combinations to reveal the shortcuts or read the help manual and memorize the shortcuts.

Changes on the diagram are specified with the question how to change. It is brought out because the direct interaction of the diagram requires attentive composition. The answer lists the most common user actions, such as click, hover, scroll, drag and touchpad gestures. Touchpad gestures and scroll are more common for navigation and zoom, but they are listed in case the designer wants to explore a wider variety of interaction possibilities.

User feedback is brought out to emphasize the importance of dynamic queries, which is considered one of the key techniques in information visualization [60]. Communication between users and technology is sensitive – if the user does not get any feedback after the user action, they most likely assume that the action did not trigger anything or that the program is not working. Options for user feedback are retrieved from Munzner's theory – immediate response, highlight and showing a progress indicator in case the user has a longer wait ahead [37].

Once the user has done changes to the diagram, it is important to know how he/she can undo those changes. For superimposed layers, it means deciding if the users have to reset settings of the previous version or there is a shortcut. For embedded data, the user has to know how to hide the data again, for example a close button or clicking elsewhere. From algorithm perspective, it is important to know if the history should be built into the algorithm and if it accessible to the user, for example the use of back button [59].

4.4.2.2 How can the user reduce data?

In process diagrams, it is common to conduct a visual search to identify patterns, singular objects or their features. Reduction of data helps the user in such search tasks. It is difficult to spot necessary elements on the diagram when a lengthy process is fitted into one static view and cannot be manipulated with panning, zooming or abstraction.

The reduction of data is also necessary for analytical tasks. The process diagrams should help to detect problems that the user nor the designer of the tool did not know existed [4]. The ability to shape the dataset can reveal important patterns about the process – a problem can be hidden in a small section of the process flow and therefore cannot be noticed if all the data is in the view. This is where filtering can help to locate critical issues.

Does the diagram need panning?
In panning (also scrolling) user moves the view along the diagram from up to down or side to side [37]. It is a necessary tool, especially when combined with zooming to show process

models that cannot be fitted to the screen without losing the readability of labels and visibility of visual channels.

Panning is usually constrained in process mining visualizations – user cannot scroll endlessly out of the boarders of the diagram. It helps the user to avoid getting lost. The designer has to decide how far the user can pan in all the directions.

Common user actions for scrolling are dragging the scrollbars, using touchpad gestures (two fingers) or arrows on the keyboard. Gestures specific to touchscreen are not proposed as it is out of the scope of this framework (see section 4.1.1).

Actions are supported with navigation elements, such as scrollbars or buttons. In order to scale the diagram for complex processes, it might be useful to add an overview-detail pane, where detail pane is on the main view and a separate smaller pane shows the overview of the whole diagram with the current location of the main view [38]. This was not done in any process diagrams that were presented in the state of art studies, but it is a common element used in other types of data visualization [61].

Does the diagram need zooming? Does the diagram need abstracting?
Zooming and abstracting are often used hand in hand in process mining visualizations. Both techniques help to navigate between overview and details. Abstracting makes the process model coarser (with less nodes and links) or more complex (with all the possible instance paths) and zoom helps to see the whole picture (full diagram in one view) or take a closer look to sections of the diagram (a selected part of diagram in the view) [37].

In the framework, zooming and abstracting share the same layout of questions. These sections follow three fundamental questions of interaction design listed by Harrower and Sheesley: "(1) what type of interactivity is needed (kind of control), (2) how much interactivity is needed (degree of control), and (3) how should this interactivity be implemented (method of control)?" [62].

The types of zooming are semantic zooming and geometric zooming. Geometric zooming is mimicking the action of getting closer to and further from objects in real life – the objects simply become larger or smaller [37]. In semantic zooming more details appear when zooming in and disappear when zooming out [62]. Semantic zooming was listed also in the previous section of the framework in the context of handling readability of labels.

Abstraction has at least two variations that can be used simultaneously – reduction or increase of number of activities (nodes) or paths (links). Usually, the abstraction on the level of activities and paths can be controlled separately. The abstraction is based on filtering, for example, the least popular activities or paths are left out in the coarse version of the diagram.

The degree of control for both, abstraction as well as filtering, is asked to be set by the designer. Designer has to identify the minimum, maximum and default level of abstraction and zoom. The question is included, because it is expected that those functionalities are not built in an unconstrained way. Unconstrained navigation can be confusing and the user can get lost in the visualization [37].

User actions and control elements for manipulating zoom and abstraction are listed to address the third question of interaction design – how to implement the interactivity [62]. Slider and buttons are common widgets for controlling zoom and aggregation [48], [49]. Control elements and actions are intertwined – sliders require dragging and buttons clicking. Touchpad gestures option is included, because a diagonal movement from center to corners is often used as a gesture for zoom.

Does the diagram need filtering?
Filtering enables users to discard uninteresting items and focus on the necessary items [59]. The filtering section in the framework follows roughly the same structure of questions as zooming and abstraction sections, identifying the type and number of filters user can apply and where/how the filters can be controlled.

Datasets can be filtered by reducing attributes or items [37]. In process mining, filtering of attributes happens on the level of switching superimposed layers of the diagram, for example selecting encoding of throughput or time. A separate filtering functionality is used only for item filtering.

Filtering can be executed in several locations in the user interface. A separate display area in the same view with the diagram could be a good option for dynamic querying, where the user gets immediate response to their selections [60]. A separate filtering view works better for advanced filtering, where all the possible filtering options are accessible for the user [59]. Shortcuts are proposed to enable quicker interaction with the visualization for frequent users [63]. Shortcuts could be placed on the diagram for each element, for example amongst the embedded data in pop-up windows as it is done in Disco [48] and Celonis [49].

The degree of filtering is identified by setting how many filters can user add – is it one set of criteria at time or can the user add new filters on top of existing ones. In both cases it is important to summarise the applied filters to the user, so that he/she does not have to memorise the state of the dataset. The most convenient for the user is to include a list of applied filters in the same view as the diagram, but it might be space-consuming. Another or additional way is to keep the list of applied filters in the filtering view.

4.4.3 Summary of Framework Design

The detailed description and reasoning of the questions and alternative answers in the framework shows that the framework is built on both, data visualization theory as well as process mining visualization practices. Most of the questions have originated from the data visualization theory, but adjusted to the specifics of the process mining. Some missing aspects from data visualization theory have been added where process visualization practices have proved a common use of those missing aspects. For example, the common use of legends and labels, which is not covered in the base data visualization theory.

5. Validation

The framework is developed for a practical use in a design process of process mining diagrams. Validation has to assess its suitability to such context. A real-life case study is conducted to reveal the strengths and weaknesses of the framework as well as give a direction for further development. The case study methodology and design are explained in section 5.1 and section 5.2 describes the results of the validation process.

5.1 Methodology and Design of a Case Study

A case study is a form of research, where phenomena is explored within its real-life context [64]. It is suitable for answering "why" and "how" questions, particularly in cases where context can provide insightful information and the research requires an observational approach [65]. The case study method is usually applied to research a topic of interest, but it has been also used for validation purposes [66]. It is chosen as a validation method for the framework, because this method allows observing the framework in the context it is developed for. The effect the framework has on its context determines its general applicability in practice as well as specific benefits and disadvantages.

5.1.1 Identification of the Case

The design of the case study depends on pre-set boundaries – an identification of the unit of the analysis (the case) and specific research questions [64].

This case study is conducted to gather data that would help to assess the framework. The ambition of the framework is to improve the process of visualizing process mining diagrams, specifically process maps. In order to achieve that, it must be applicable in the context of visualizing process mining diagrams. Firstly, it has to be understandable; secondly, relevant in the given context; and thirdly, complete in order to help to create thought-through visualizations. In addition, it has to be easy to use for the target audience to ensure the balance between the time and effort the users spend on the framework and the benefits they gain. Therefore, the assessment of the framework means evaluating its understandability, relevance, completeness and usefulness. The case study helps to collect data that would help to assess those specific points.

As the framework is primarily built for developers of process mining tools, their interaction with the framework and their subjective opinions are a good source for gathering data. The case study is interested in pinpointing the aspects of the framework that have a positive impact, as well as elements that require further development or discarding. In broader terms, the case study is conducted to capture the effect of the framework on the process mining visualization development context. Therefore, the unit of the analysis is defined as follows:
C1: The effect of the framework on data visualization design tasks executed by developers of process mining tools.

The effect is observed through the lens of the following research questions:
C1-RQ1: How is the framework understandable/unclear for the users?
C1-RQ2: How is the framework relevant/irrelevant for the process of visualizing process mining diagrams?
C1-RQ3: Which aspects of the framework are complete/incomplete?

C1-RQ4: How easy is it to use the framework?

The research questions aim to target the understandability (C1-RQ1), relevance (C1-RQ2), completeness (C1-RQ3) and usefulness (C1-RQ4) of the framework.

5.1.2 Context and Participants of the Case Study

The case study is set in the context of a project for visualizing data from a European queuing management system. A group of developers are building a process mining tool that would help to translate the data into insightful information to improve and innovate the queuing process. The process mining techniques they are developing the tool for, are process discovery, performance analysis, performance predictions and deviance mining. The framework is used to help to visualize the process diagram of the tool the team is developing.

5.1.3 Design of the Case Study

Three members of the team are included to the case study. Each participant is invited to an individual session, where they are asked to use the framework for their visualization task. A group workshop format is not chosen because each participant has an individual task that may require a different focus in the use of the framework compared to tasks of other participants. It must be kept in mind that the participants cannot be expected to come up with conclusive and complete design decisions while they use the framework, because in their design process they usually discuss and mutually agree on the decisions as a group (see section 5.1.2).

During the session, data is gathered in two ways – in an interview and a direct observation format. The session is divided to three parts. The first step is a semi-structured interview. The initial interview helps to gain a general background information about the project and participants as well as identifies the struggles the participants are facing in their current project. The exact questions are listed in Appendix IV.

The second stage of the case study is in a form of a direct observation. Firstly, the participant is asked to explain briefly the current visualization idea. The team has been working on few visualization concepts, which are used as a comparison point – the participant is later asked if and how the initial idea improved through the use of the framework. Therefore, the idea must be first addressed before the use of the framework.

Secondly, the participant is given a brief introduction on the framework, after which he/she is asked to use the framework in designing the visualization for their task. The participant can either improve the initial idea or develop a completely new idea. He/she is encouraged to comment on the aspects that are unclear, provoke new ideas or have an effect on them in some other way. The participant is also allowed to ask clarifications if something is not understandable. In this case, the observer can interact with the participant and give a brief explanation. Each such comment and/or interaction will be noted down. The participant has 30 minutes to use the framework. If the time is up, the participant is asked if he/she wishes to continue to use the framework or finish the try-out. If the participant chooses to continue, he/she is asked to notify the observer when they are finished using the framework. In any case, the second stage is over when the participant wants to finish using the framework.

The third step is a semi-structured interview that gathers the participant's opinions on the framework and the development progress of their visualization task. Firstly, the participant is asked to describe the new or improved visualization idea or the general state of the progress of the visualization if a single idea cannot be described. After that the participant is asked a set of questions about his/her opinions about the framework. The goal of the last interview is to find out which ways did the framework prove itself understandable, relevant, complete and usable and which ways it failed at doing so. Exact questions are listed in Appendix IV.

The individual workshops take place through video calls and are documented in audio recordings and notes. In addition, screenshots are taken if it is necessary to capture the visuals, such as sketches.

5.2 Results of the Case Study

The data was collected from three individual sessions. The first part of the session gave background information about the participant's relation to process mining and visualization, the project at hand and the visualization process in the project. This information is presented in section 5.2.1.

The second and third part of the session gave feedback on the understandability, relevance, completeness and usefulness. The data was extracted from the observational part and interviews according to the listed categories: section 5.2.2 discusses the understandability of the framework, section 5.2.3 relevance, section 5.2.4 completeness and section 5.2.5 usefulness. Section 5.2.6 summarizes the ideas from the feedback that could be used to develop the framework further and section 5.2.7 identifies the threats to validity.

5.2.1 Background of the Project and the Participants

Three members of the team were participating in the case study – a data scientist and two researchers. All the participants have a previous experience in process mining. The data scientist is developing a PhD thesis in process mining field and both of the process mining researchers have about 10 years of experience in conducting and participating in process mining studies. In addition, all the participants have some experience in data visualization field. The first participant has been using data visualization mostly for presentation purposes – to show the findings of the data. The second participant has become acquainted with the data visualization concepts through practice as well as theory – the lectures he holds require familiarity with the data visualization literature. The third participant has developed process mining tools that include visual presentations. None of the participants are professional visualization designers.

Each member has their own focus in the project – performance analysis, predictive analysis and deviance mining. However, visualization design in the project is a group effort. They have agreed on a common base – a process map – and the members propose additional visualization ideas from the perspective of their focus. The personal ideas go through a group discussion before they are agreed upon.

At the time when the framework was introduced to the team members, the group had come up with several visualization ideas. The ideas that included a process map were based on the most widely used existing tools, such as Disco [48], and mimicked the basic visualization of those tools. However, the team saw potential in going beyond the existing solutions. For example, they wanted to add an integrated map for comparing process variants, where each process variant is visualized with a differently colored link. In addition, they wanted to show the exact values of the waiting times, not only as a derived average, but the full distribution of the values, for example in a form of a histogram or a density plot.

The main concern for their visualization was how to show the complex reality without overwhelming the analyst. The user has to have access to the real picture of the process, but it cannot be only that, because oftentimes the reality is too complex to be comprehensible in one image. They were dealing with problems about enabling overview and drilling down to details.

Other visualization struggles that the team was facing are specific for the techniques the participants were working on. Performance analysis requires the most effective way to visualize process variants and project performance data on the diagram. The problem in visualizing deviance mining lays in the complexity of sequences – if a certain sequence characterizes a process variant, the activities in the characteristic sequence may not be directly succeeded by one another – they may or may not have various activities in-between. The sequence is not clear-cut and hence, it is difficult to summarize and visualize such sequence. Predictive mining is also struggling with the activities on the diagram – predictions are for a certain time-interval and it may happen that the activities depicted on the general map may not be executed in the future prediction.

All the participants were aware that the visualization they are working on is a complex design task. Even so, they had not considered taking input from data visualization theory. They relied on the examples of common existing tools and their previous experience on the field.

5.2.2 Understandability of the Framework

Understandability refers to the comprehension of the framework by the target users, who have reasonable knowledge of process mining. The understandability was questioned from three angles – was it clear how to use the framework, was the purpose of the framework understandable, and were the terminology and illustrations comprehensible. Notions about the understandability were gathered from the second part of the session – which questions in the framework the participants did not understand – and from the third part of the session – participants' opinions on the understandability.

Generally, the participants did not have problems with understanding how to use the framework. After the brief introduction of instructions, all the participants worked on their own with the framework. They asked various questions about the framework (see table 3), but none of the questions addressed the basics of the framework. The general elements of the framework – numbering of the questions, markings of different types of answers (radio buttons, checkboxes, etc.), illustrations and tables of strengths and weaknesses – did not cause confusion and the participants used those elements as they were supposed to.

Also, the participants did not struggle with identifying the purposes of the framework. The participants found the framework useful for tool improvement, making vague visualization ideas more concrete and using it as an inspiration point for designing new visualizations. One participant referred to it as a catalogue of tested ideas, which can be revisited several times during the design process. Another participant saw its use also in user surveys to find out the solutions that the target users would prefer. Two participants pointed out its potential to be developed into a mock-up tool, where the answers of the questions result in an example visualization.

Even though, the basic understandability of the framework was good, all the participants highlighted aspects that could improve it. During the use of the framework participants were allowed to ask about the questions that were unclear for them. Table 3 lists the exact questions that were brought up, the reason of confusion and how many participants mentioned it.

Table 3. Unclear Questions in the Framework

Question	*Reason of Confusion*	*Number of Participants*
1.1.3 How are the basic elements ordered?	Target	1
1.1.3.1 How is the sequence of the process shown?	Target	1
1.1.4 How is the diagram aligned?	Hierarchy, definition	2
1.1.4.1 How are the process diagrams faceted?	Definition	1
1.1.4.2 What is the alignment based on?	Definition, target	2
1.1.4.3 Is the layout deterministic or nondeterministic?	Definition	1
1.2.1 Which attributes are shown on the diagram?	Missing answer, definition	1
1.2.2 Which channels express the attributes?	Target	2
1.2.3 How is the data faceted on the diagram?	Definition	1
1.2.3.1 Which channels and attributes are visible in each layer?	Definition	1
1.2.4 How does the user know the meaning of channels?	Definition	1
2.1.1 What can be changed on the diagram?	Definition	1
2.1.2 How do the changes appear?	Definition	1
2.1.4.2 How does the user get feedback to the actions?	Definition	1
2.2.1 Does the diagram need panning?	Definition	2

Overall, fifteen questions were brought up during the second part of the session. Eleven questions were mentioned only by one participant, four questions were mentioned by two participants and none were mentioned by all participants. Most of the unclear questions are in the first part of the framework – eleven questions belong to the first section about encoding and the four to the second section about interaction.

There were four different reasons for confusion. The most common reason was definitions. Twelve questions and answers included terms that were unfamiliar or unclear to the participants. Another reason was an unclear target – the participant did not what is expected from him/her or what is the target of the question. For example, are the questions about what the user wishes or what is feasible. One participant found the hierarchy marking of the answers unclear (marked as “hierarchy” in the table). One participant could not find an option for a suitable answer (marked as “missing answer” in the table).

The participants’ opinions collected from the semi-structured interview provided further insights about unclear aspects of the framework.

The feedback on the vocabulary in the framework was divided. Two participants mentioned that some of the terminology was confusing, specifically in the questions they asked during the practical part of the session. One participant found the terms easy, but added that this is due to his familiarity with the literature of visualization theory. One participant suggested a glossary, where the terms could be easily looked up.

Two participants found targeting of the questions unclear – it was not easy to understand what is the question about. Both of the participants brought out two reasons for it. Firstly, the wording of the questions – it was not clear if the questions are about existing solutions or prospective preferences. For example, "How is the diagram aligned?" refers to something that already exists, while wording such as "How would you like the diagram to be aligned?" would clearly direct the designer to think about his/her ideas that do not have to be ready-made.

Another reason for the targeting confusion was the sequence of topics and transitions. The questions move from one topic to another with abrupt transitions and the user may miss the que of switching to another target of the questions. For example, question 1.1.4.1.1 is about faceting process diagrams, while 1.2.3 is also about faceting, but this time faceting data on one diagram. If the sequence and hierarchy of those questions are not carefully followed, the targets of the questions are not understandable.

Illustrations were generally found helpful in understanding the questions and alternative answers. One specific illustration was not understandable for one participant (illustration next to the question 1.2.3.1).

One participant mentioned that it was easy to understand the framework because of his experience with Disco and was not sure if people with less experience with existing tools would find it as understandable. One participant thought that examples from real tools could improve the understandability.

5.2.3 Relevance of the Framework

The relevance of the framework means how well does it help with the visualization tasks for process mining diagrams. The following aspects were considered: firstly, the relevance of the identified purposes, secondly, the relevance of the named benefits and thirdly, if the participants thought the framework was relevant enough to recommend it to a colleague.

The purpose of the framework was easy to pinpoint for the participants. All the mentioned purposes were relevant for process mining visualization tasks – tool improvement, a point of inspiration and a guidance for making ideas more concrete.

Overall, the main value of the framework was found in giving new ideas and helping to enrich or modify existing ones. All the participants mentioned that they got new suggestions for the visualizations in their project. One participant found the framework valuable for general purposes, but not for seeking answers for very narrow tasks, such as how to visualize deviances.

All the participants would recommend the framework for colleagues, who are struggling with visualization tasks for process mining. One participant thought it would be useful also for other types of data visualization tasks.

5.2.4 Completeness of the Framework

The section about the completeness pinpoints the aspects that are missing or excessive in the framework. The feedback on completeness was collected in two ways – firstly, which parts of the framework need to be added, modified or discarded; and secondly, how complete is the visualization idea after the use of the framework.

Most of suggestions for adding and changing the framework stem from the issues of understandability. For instance, examples of real tools and a glossary of definitions were suggested to make the framework more clear. Also, the transitions between topics were brought out as a potential place to improve the comprehensibility of the framework. One participant suggested a solution for clarifying the targets of the questions by having less topics in the framework. For example, focusing on one of the main topics – representation of data or interactivity – and going deeper in the selected topic, while discarding the other.

Other further development opportunities were found in adding more alternatives and questions to the framework. One suggestion was to add even more alternatives to all the questions. Another participant saw a possibility to include more questions specifically about embedded data – how to visualize the data that is shown in the pop-up windows. It was also mentioned that the framework is already quite long and therefore, the additional questions and answers can make it too lengthy for the users' convenience.

All the participants saw a potential in digitalizing the framework (the participants were working with a print-out of the framework). Two participants mentioned the potential to develop these questions into a mock-up tool that could result in an example diagram based on the selected alternatives. One participant suggested to hide the positive and negative aspects in the default view and give the user an option to reveal it if necessary. This idea came about because some of the aspects seem to give a suggestion to the preferred answer – for example if one alternative has two positive aspects and the other four, it seems that the second option is better.

None of the participants thought they got a complete idea of their visualization. One participant mentioned that the new or improved ideas must be discussed with their whole team. Another participant said that this framework should be visited several times during the design process. It was also mentioned that the framework works well in the initial stages of the design process and that it is not meant for generating complete ideas.

5.2.5 Usefulness of the Framework

The usefulness in this context means the ease and convenience of use from the perspective of time and effort. It is measured in terms of how much effort the framework required from the participants in general and if the time and effort put into the framework was worth it compared to benefits it gave.

Two participants managed to go through the whole framework within 25 minutes and one participant used 45 minutes. The participant who took more time noted that the framework could take even longer time because of the listing of positive and negative aspects that are in a textual format and require careful consideration. Another participant said that the framework could be visited several times, therefore, it could take more time than half an hour.

One participant estimated the level of required focus high, while two others thought it required little effort from them. The participants said that the difficulty lays in the understandability aspects of the framework, such as terms and targets of the questions.

All the participants thought that time and effort they put into the framework was worth it. They got new and more concrete ideas for their visualization in a relatively short time. The benefits and the time spent were balanced.

5.2.6 Summary of the Feedback

Overall, the framework was found relevant for coming up with new ideas or improving and clarifying existing ideas of process mining visualizations. The basic use of the framework is easy and understandable – the participants did not struggle with the understandability of aspects, such as which questions to answer or the meaning of illustrations or tables. The time required for going through the framework on paper varies from 25 to 45 minutes, but it can be used for even longer. The time and effort spent on the framework is considered balanced in relation to the benefits it gives. The participants could easily name potential uses for the framework and would recommend it to their colleagues who struggle with visualization tasks.

Even though the general effect of the framework was positive, the feedback uncovered several ways how to improve the framework. Table 4 summarizes the specific problems and opportunities that appeared.

Table 4. Problems and Opportunities Extracted from the Feedback

Aspect	*Problem/Opportunity*	*Possible Solutions*
Understandability	Unclear terminology	Add a glossary
	Unclear targets of questions	Add transitions Modify the wording of questions
	Additional visual examples	Redesign the unclear example Add examples of existing diagrams
Completeness	Missing topics	Add technique-specific questions and answers (e.g. for deviances) Add questions about designing embedded data
	Excess topics, lack of focus	Pick one of two main topics (encoding or interaction), develop it further and discard the other
	Additional alternative answers	Add more alternative answers
Relevance	Extension of use	Develop it into a mock-up tool Develop it into a survey tool

The main concern for the further development of the framework should be the improvement of understandability – this topic got the most feedback and suggestions for changes. The main problems in understanding lay in unfamiliar terminology and targets of questions. The first could be solved with a glossary. The second could be improved by modifying the wording of questions or adding transitions, e.g. textual introductions for new topics. Also, the visualizations could be improved. For instance, examples from real tools can be included. In addition, the visualization that was unclear for one user can be redesigned.

The framework got several suggestions on completeness. The missing topics were the design of embedded data and technique-specific questions or examples, such as visualization of deviances. One participant thought that the framework would benefit from having a narrower focus – either only on the encoding or interaction. In addition, it was mentioned that additional alternative answers would be interesting for the user.

In terms of relevance, the improvement could be done in extending the tool beyond its current purpose. Some ideas that came up were developing the framework into a mock-up tool or adapting it for the use of user surveys to gather statistics about users' preferences.

5.2.6 Threats to Validity

The case study was a good fit for gathering feedback on the framework in terms of the representativeness of the participants and context, which was aligned with the primary users and context the framework is developed for. However, some aspects of the case study may impose threats to the validity.

The number of the participants was low, which makes the generalization of the feedback difficult. Even though each participant had a different focus in the project, they were developing visualizations for the same tool and had the same base diagram to refer to. For more diverse feedback, more participants should be included from various process mining projects.

There were some overlaps in the feedback, but most of the opinions and ideas for improvement were individual and unique. Such variation ties the feedback to the individual who gave it and threatens the generalizability of the feedback. Perhaps a combination of structured and semi-structured interview questions would have reduced the variation.

Finally, the sessions were led by the author of the framework that might have influenced the approach of the participants. They may have felt a stronger need to cooperate and help the researcher compared to a set-up, where a neutral third party would have conducted the sessions.

6. Conclusions

This thesis presented a visualization framework for designing process mining diagrams. The need of the framework was supported by a state of art research. The research showed a high importance of the visualizations in process mining field – most of process mining techniques use some form of visualizations. It also revealed the complexity of the designs. Every study that included a node-link diagram – the most common type of visualization –, presented it in a unique way. The third finding from the state of art research showed that regardless of the importance and complexity of the visualizations, most of the diagrams were designed without any help from visualization frameworks. Instead, the design decisions were based on a combination of logical argumentation, existing practices and domain input. Some authors used also input from visualization theory, but did that in a fragmented way. Consequently, the research supported a need for a visualization framework for process mining diagrams.

The framework aims to identify the important aspects in interactive process maps that require conscious design decisions and provide information that helps the designer to justify their choices. It is based on two cornerstones – existing process mining visualizations and data visualization theory. Majority of the topics covered in the framework were extracted from a visualization theory by Tamara Munzner [37]. In addition, the idea of presenting the topics in a form of hierarchical questions stemmed from Munzner's work [37]. However, adjustments were made to the base theory to make it relevant to process mining. Firstly, all the questions were retargeted to extract answers about the most common and complex form of charts in process mining – a node-link diagram, more specifically a process flow map. Questions were enriched with alternative answers that were relevant to process maps. The alternative answers were extracted from both, data visualization theory and existing process mining visualization practices. In addition, strengths and weaknesses of the alternative design choices were given where possible. These aspects were derived again from both, visualization theory as well as existing visualizations. Illustrations that are specific to process mining, were designed and added to the framework to increase the comprehensibility through visual examples.

The framework was tested and evaluated in a case study. It was used by three participants, who were all involved in a project for developing a process mining tool for analyzing border crossing data. The participants used the framework for 25-45 minutes and shared their opinions in a semi-structured interview. The collected data was analyzed from the perspectives of understandability, relevance, completeness and usefulness. Generally, the participants managed to use the framework without major struggles. The framework was found relevant and balanced in terms of how much effort it requires and how beneficial it is to the task at hand. However, understandability and completeness triggered several suggestions. Some of the terms and targets of the questions were confusing for the participants. Possible solutions were suggested, such as rewording the questions to clarify the targets and adding a glossary to define the terms. The main feedback on the completeness was about the missing aspects of the framework. The participants would have appreciated more alternative answers, more questions about designing embedded data and more technique-specific questions, such as questions that could help to design deviances. The main value of the framework was found in making vague ideas concrete, coming up with new ideas and improving existing ones. The extended use was seen in developing the framework into a mock-up tool and using it for surveys to find out the users' preferences for such diagrams.

The case study approach for validation was a good fit because the framework was tested out by the users and in the context the framework is developed for. However, the number of the participants was small and all the participants were part of the same project, which makes the generalizability of the feedback difficult. Another aspect that reduces the generalizability is the high variance of the feedback – most of the opinions were unique and depended on the individual who gave it. In further development of the framework, more feedback should be gathered from a higher number of participants in a variety of projects and some structured interview questions should be added to be able to make general conclusions out of the feedback. Nevertheless, the feedback that was gathered in the context of this thesis provides already valuable insights and ideas how to improve the framework on a cosmetic level as well as in general directions.

In addition to the ideas from the feedback, the framework can be developed further by expanding the defined scope of the framework. The framework can be extended to other types of charts besides process diagrams, such as various dashboards used in process mining tools. Also, it can be developed further to help with the full user interface design, not only the design of interactive process maps. Current version is tailored for static logs, which can be extended also to dynamic logs. In addition, the extension of the framework can be explored in terms of developing the diagrams for different types of screens besides laptops/desktops, for example taking in consideration characteristics of touchscreen design. The format of the framework can be improved by making it digital and taking advantage of the possibilities of interactive design, such as options to insert the answers interactively and produce automated reports of the collected data.

7. References

[1] IBM, “10 Key Marketing Trends for 2017,” 2016.

[2] A. Bolt, M. de Leoni, and W. M. P. van der Aalst, “A Visual Approach to Spot Statistically-Significant Differences in Event Logs Based on Process Metrics,” *Adv. Inf. Syst. Eng. (CAISE 2016), B. Ser. Lect. Notes Comput. Sci.*, vol. 9694, pp. 151–166, 2016.

[3] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, “Exploring Processes and Deviations,” *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 202, pp. 304–316, 2015.

[4] T. Gschwandtner, “Visual Analytics Meets Process Mining: Challenges and Opportunities,” *Data-Driven Process Discov. Anal. SIMPDA 2015, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 244, pp. 142–154, 2017.

[5] S. Bachhofner, I. Kis, C. Di Ciccio, and J. Mendling, “Towards a multi-parametric visualisation approach for business process analytics,” *Adv. Inf. Syst. Eng. (CAISE 2017), B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 286, pp. 85–91, 2017.

[6] M. Hipp, A. Strauss, B. Michelberger, B. Mutschler, and M. Reichert, “Enabling a User-Friendly Visualization of Business Process Models,” *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 202, pp. 395–406, 2015.

[7] M. Gall, G. Wallner, S. Kriglstein, and S. Rinderle-Ma, “A study of different visualizations for visualizing differences in process models,” *Adv. Concept. Model. B. Ser. Lect. Notes Comput. Sci.*, vol. 9382, pp. 99–108, 2015.

[8] T. Munzner, “A Nested Model for Visualization Design and Validation,” *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 921–928, 2009.

[9] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design: Beyond Human - Computer Interaction*. West Sussex, UK: John Wiley & Sons Ltd, 2011.

[10] R. Unger and C. Chandler, *A Project Guide to UX Design: For user experience designers in the field or in the making*, 2nd ed. Berkeley: New Riders, 2012.

[11] K. Oruste, “Process Mining in Industry,” University of Tartu, 2017.

[12] N. Gupta, K. Anand, and A. Sureka, “Pariket: Mining Business Process Logs for Root Cause Analysis of Anomalous Incidents,” *Databases Networked Inf. Syst. DNIS 2015, B. Ser. Lect. Notes Comput. Sci.*, vol. 8999, p. 244/263, 2015.

[13] The Editors of Encyclopaedia Britannica, “Cartography,” *Encyclopaedia Britannica*. Encyclopædia Britannica, inc., 2017.

[14] M. Friendly, M. Sigal, and D. Harnanansingh, “The Milestones Project: A Database for the History of Data Visualization,” in *Visible Numbers: Essays on the History of Statistical Graphics*, M. A. Kimball and C. Kostelnick, Eds. New York, USA: Routledge, 2013, pp. 219–234.

[15] B. Kitchenham, “Procedures for Performing Systematic Reviews,” Keele, UK, 2004.

[16] M. Gall, G. Wallner, S. Kriglstein, and S. Rinderle-Ma, “Differencegraph - A ProM plugin for calculating and visualizing differences between processes,” *CEUR Workshop Proc.*, vol. 1418, pp. 65–69, 2015.

[17] C. Cordes, T. Vogelsang, and H.-J. Appelrath, “A Generic Approach for Calculating and Visualizing Differences Between Process Models in Multidimensional Process Mining,” *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 202, pp. 383–394, 2015.

[18] M. T. Wynn *et al.*, “ProcessProfiler3D: A visualization framework for log-based process performance comparison,” *Decis. Support Syst.*, vol. 100, no. SI, Special Issue, pp. 93–108, 2017.

[19] K. Slaninova, D. Vymetal, and J. Martinovic, “Analysis of Event Logs: Behavioral

Graphs," *Web Inf. Syst. Eng. - WISE 2014 Work. B. Ser. Lect. Notes Comput. Sci.*, vol. 9051, pp. 42–56, 2015.

[20] K. Slaninova, J. Martinovic, P. Drazdilova, and V. Snashel, "From Moodle Log File to the Students Network," *Int. Jt. Conf. SOCO "13-CISIS"13-ICEUTE'13, B. Ser. Adv. Intell. Syst. Comput.*, vol. 239, pp. 641–650, 2014.

[21] A. Jalali, "Supporting Social Network Analysis Using Chord Diagram in Process Mining," *Perspect. Bus. Informatics Res. BIR 2016, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 261, pp. 16–32, 2016.

[22] P. Kecman and R. M. P. Goverde, "Process mining of train describer event data and automatic conflict identification," *WIT Trans. Built Environ.*, vol. 127, pp. 227–238, 2012.

[23] K. Jorbina *et al.*, "Nirdizati: A web-based tool for predictive process monitoring," *CEUR Workshop Proc.*, vol. 1920, 2017.

[24] R. P. J. C. Bose and W. M. P. van der Aalst, "Discovering Signature Patterns from Event Logs," *Proc. 2013 IEEE Symp. Comput. Intell. Data Mining, CIDM 2013 - 2013 IEEE Symp. Ser. Comput. Intell. SCI 2013*, pp. 111–118, 2013.

[25] P. M. Dixit, H. S. G. Caballero, A. Corvo, B. F. A. Hompes, J. C. A. M. Buijs, and W. M. P. van der Aalst, "Enabling Interactive Process Analysis with Process Mining and Visual Analytics," *Proc. 20th Int. Jt. Conf. Biomed. Eng. Syst. Technol.*, vol. 5, pp. 573–584, 2017.

[26] A. Burattin, M. Cimitile, and F. M. Maggi, "Lights, Camera, Action! Business Process Movies for Online Process Discovery," *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 202, pp. 408–419, 2015.

[27] F. Mannhardt, M. de Leoni, and H. A. Reijers, "The multi-perspective process explorer," *CEUR Workshop Proc.*, vol. 1418, pp. 130–134, 2015.

[28] T. Vogelsang and H.-J. Appelrath, "Multidimensional process mining with PMCube explorer," *CEUR Workshop Proc.*, vol. 1418, p. 90/94, 2015.

[29] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Process and deviation exploration with inductive visual miner," *CEUR Workshop Proc.*, vol. 1295, pp. 46–50, 2014.

[30] R. C. Basole, H. Park, M. Gupta, M. L. Braunstein, D. H. Chau, and M. Thompson, "A Visual Analytics Approach to Understanding Care Process Variation and Conformance," *ACM Int. Conf. Proceeding Ser.*, vol. 6, 2015.

[31] B. F. van Dongen and A. Adriansyah, "Process mining: Fuzzy clustering and performance visualization," *Bus. Process Manag. Work. BPM 2009, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 43, pp. 158–169, 2010.

[32] A. Pini, R. Brown, and M. T. Wynn, "Process visualization techniques for multi-perspective process comparisons," *Asia Pacific Bus. Process Manag. AP-BPM 2015, B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 219, pp. 183–197, 2015.

[33] M. de Leoni, S. Suriadi, A. H. M. ter Hofstede, and W. M. P. van der Aalst, "Turning event logs into process movies: animating what has really happened," *Softw. Syst. Model.*, vol. 15, no. 3, pp. 707–732, 2016.

[34] J. Stolfa, S. Stolfa, M. Kopka, and V. Snashel, "Adaptation of Turtle Graphics Method for Visualization of the Process Execution," *Afro-European Conf. Ind. Adv. AECIA2014 B. Ser. Adv. Intell. Syst. Comput.*, vol. 334, pp. 327–334, 2015.

[35] J. Gulden and S. Attfield, "Business Process Models for Visually Navigating Process Execution Data," *SAP Bus. Process Manag. Work. B. Ser. Lect. Notes Bus. Inf. Process.*, vol. 256, pp. 583–594, 2016.

[36] W. Lucas, J. Xu, and T. Babaian, "Visualizing ERP Usage Logs in Real Time," *ICEIS Proc. 15th Int. Conf. Enterp. Inf. Syst.*, vol. 3, pp. 83–90, 2013.

[37] T. Munzner, *Visualization Analysis and Design*. Boca Raton: AK Peters/CRC Press, 2014.

[38] K. Hornbæk, B. B. Bederson, and C. Plaisant, "Navigation Patterns and Usability of Zoomable User Interfaces with and without an Overview," *ACM Trans. Comput. Interact.*, vol. 9, no. 4, pp. 362–389, 2002.

[39] W. W. Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*, 2nd ed. New Jersey: John Wiley & Sons Ltd, 2011.

[40] S. Wexler, J. Shaffer, and A. Cotgreave, *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenario*. John Wiley & Sons Ltd, 2017.

[41] S. Few, *Information Dashboard Design: The Effective Visual Communication of Data*, 1st ed. O'Reilly Media, Inc., 2006.

[42] R. Damelio, *The Basics of Process Mapping*, 2nd ed. Productivity Press, 2011.

[43] S. Carpendale, "Evaluating Information Visualizations," *Lect. Notes Comput. Sci. Inf. Vis. Human-Centered Issues Perspect.*, vol. 4950, pp. 19–45, 2008.

[44] H. Lam, E. Bertini, and P. Isenberg, "Empirical Studies in Information Visualization: Seven Scenarios," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 9, pp. 1520–1536, 2012.

[45] L. Wilkinson, *Grammar of graphics*. New York: Springer, 1999.

[46] E. Tufte, *The Visual Display of Quantitative Information*. Cheshire: Graphics Press, 1983.

[47] C. Ware, *Information Visualization: Perception for Design*. Hampshire: Morgan Kaufman, 1999.

[48] Fluxicon Process Laboratories, "Disco." 2012.

[49] Celonis SE, "Celonis." Munich, 2016.

[50] Object Management Group, "Business Process Model and Notation (BPMN)," 1997. [Online]. Available: http://www.bpmn.org/. [Accessed: 26-Mar-2018].

[51] D. Clarck, "Content Management and the Separation of Presentation and Content," *Tech. Commun. Q.*, vol. 17, no. 1, pp. 35–60, 2007.

[52] L. Colligan, J. E. Anderson, H. W. W. Potts, and J. Berman, "Does the process map influence the outcome of quality improvement work? A comparison of a sequential flow diagram and a hierarchical task analysis diagram," *BMC Health Serv. Res.*, vol. 10, no. 7, 2010.

[53] S. K. Card and J. Mackinlay, "The structure of the information visualization design space," *Proc. 1997 IEEE Symp. Inf. Vis.*, 2002.

[54] S. S. Stevens, *Psychophysics*. New York, USA: John Wiley & Sons Ltd, 1975.

[55] W. S. Cleveland and R. McGill, "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 531–554, 1984.

[56] J. Heer and M. Bostock, "Crowdsourcing graphical perception: using mechanical turk to assess visualization design," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 203–212, 2010.

[57] R. Sieber, S. Wiesmann, and C. Schmid, "Smart Legend - Smart Atlas!," in *Proceedings of the 22nd International Cartographic Conference*, 2005.

[58] J. Heer and G. G. Robertson, "Animated Transitions in Statistical Data Graphics," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1240–1247, 2007.

[59] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," *IEEE Symp. Vis. Lang. 1996. Proc.*, pp. 337–343, 2002.

[60] C. Ahlberg, C. Williamson, and B. Shneiderman, "Dynamic queries for information exploration: an implementation and evaluation," *CHI '92 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 619–626, 1992.

[61] R. Bade, S. Schlechtweg, and S. Miksch, “Connecting time-oriented data and information to a coherent interactive visualization,” *CHI ’04 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 105–112, 2004.

[62] M. Harrower and B. Sheesley, “Designing Better Map Interfaces: A Framework for Panning and Zooming,” *Trans. GIS*, vol. 9, no. 2, pp. 77–89, 2005.

[63] B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 4th ed. Boston: Pearson Addison Wesley, 2004.

[64] P. Baxter and S. Jack, “Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers,” *Qual. Rep.*, vol. 13, no. 4, pp. 544–559, 2008.

[65] R. K. Yin, *Case study research and Applications: Design and methods*, 6th ed. Los Angeles: SAGE publications, 2018.

[66] F. Milani, M. Dumas, R. Matulevičius, and N. Ahmed, “Modelling Families of Business Process Variants: A Decomposition Driven Method,” *Inf. Syst.*, vol. 56, pp. 55–72, 2016.

Appendix

I List of Research Papers Included to the State of Art Study

Author(s)	Title	Year	Source
Wynn, Moe T. Poppe, Erik Xu, Jingxin ter Hofstede, Arthur H.M. Brown, Ross Pini, Azzurra van der Aalst, Wil M.P.	ProcessProfiler3D: A visualisation framework for log-based process performance comparison	2017	Decision Support Systems
Dixit, Prabhakar M. Caballero, Humberto Simon Garcia Corvo, Alberto Hompes, Bart F. A. Buijs, J. C. A. M. van der Aalst, Wil M.P.	Enabling Interactive Process Analysis with Process Mining and Visual Analytics	2017	Proceedings of the 20th International Joint Conference on Biomedical Engineering Systems and Technologies
Gschwandtner, Theresia	Visual Analytics Meets Process Mining: Challenges and Opportunities	2017	Data-Driven Process Discovery and Analysis, SIMPDA 2015, Book Series: Lecture Notes in Business Information Processing
Bachhofner, Stefan Kis, Isabella Di Ciccio, Claudio Mendling, Jan	Towards a Multi-parametric Visualisation Approach for Business Process Analytics	2017	Advanced Information Systems Engineering (CAISE 2017), Book Series: Lecture Notes In Business Information Processing
Jorbina, Kerwin Rozumny, Andrii Verenich, Ilya Di Fancescomarino, Chiara Dumas, Marlon Ghidini, Chiara Maggi, Fabrizio Maria La Rosa, Marcello Raboczi, Simon	Nirdizati: A Web-based Tool for Predictive Process Monitoring	2017	CEUR Workshop Proceedings
de Leoni, Massimilano Suriadi, Suriadi ter Hofstede, A.H.M. van der Aalst, Wil M.P.	Turning event logs into process movies: animating what has really happened	2016	Software and Systems Modeling
Gulden, Jens Attfield, Simon	Business Process Models for Visually Navigating Process Execution Data	2016	SAP Business Process Managment Workshops, Book Series: Lecture Notes in Business Information Processing
Bolt, Alfredo de Leoni, Massimilano van der Aalst, Wil M.P.	A Visual Approach to Spot Statistically-Significant Differences in Event Logs Based on Process Metrics	2016	Advanced Information Systems Engineering (CAISE 2016), Book Series: Lecture Notes In Computer Science
Jalali, Amin	Supporting Social Network Analysis Using Chord Diagram in Process Mining	2016	Perspectives in Business Informatics Research, BIR 2016, Book Series: Lecture Notes in Business Information Processing
Cordes, Carsten Vogelsang, Thomas Appelrath, Hans-Juergen	A Generic Approach for Calculating and Visualizing Differences Between Process Models in Multidimensional Process Mining	2015	Business Process Management Workshops, BPM 2014, Book Series: Lecture Notes in Business Information Processing

Burattin, Andrea Cimitile, Marta Maggi, Fabrizio Maria	Lights, Camera, Action! Business Process Movies for Online Process Discovery	2015	Business Process Management Workshops, BPM 2014, Book Series: Lecture Notes in Business Information Processing
Stolfa, Jakub Stolfa, Svatopluk Kopka, Martin Snashel, Vaclav	Adaptation of Turtle Graphics Method for Visualization of the Process Execution	2015	Afro-European Conference for Industrial Advancement, AECIA2014 Book Series: Advances in Intelligent Systems and Computing
Basole, Rahul C. Park, Hyunwoo Gupta, Mayank Braunstein, Mark L. Chau, Duen Horng Thompson, Michael	A Visual Analytics Approach to Understanding Care Process Variation and Conformance	2015	ACM International Conference Proceeding Series
Pini, Azzurra Brown, Ross Wynn, Moe T.	Process Visualization Techniques for Multi-perspective Process Comparisons	2015	Asia Pacific Business Process Management, AP-BPM 2015, Book Series: Lecture Notes in Business Information Processing
Mannhardt, Felix de Leoni, Massimilano Reijers, Hajo A.	The Multi-perspective Process Explorer	2015	CEUR Workshop Proceedings
Leemans, Sander J.J. Fahland, Dirk van der Aalst, Wil M.P.	Exploring Processes and Deviations	2015	Business Process Management Workshops, BPM 2014, Book Series: Lecture Notes in Business Information Processing
Gupta, Nisha Anand, Kritika Sureka, Ashish	Pariket: Mining Business Process Logs for Root Cause Analysis of Anomalous Incidents	2015	Databases in Networked Information Systems, DNIS 2015, Book Series: Lecture Notes in Computer Science
Vogelsang, Thomas Appelrath, Hans-Juergen	Multidimensional Process Mining with PMCube Explorer	2015	CEUR Workshop Proceedings
Gall, Manuel Wallner, Guenter Kriglstein, Simone Rinderle-Ma, Stefanie	Differencegraph - A ProM Plugin for Calculating and Visualizing Differences between Processes	2015	CEUR Workshop Proceedings
Hipp, Markus Strauss, Achim Michelberger, Bernd Mutschler, Bela Reichert, Manfred	Enabling a User-Friendly Visualization of Business Process Models	2015	Business Process Management Workshops, BPM 2014, Book Series: Lecture Notes in Business Information Processing
Slaninova, Katerina Vymetal, Dominik Martinovic, Jan	Analysis of Event Logs: Behavioral Graphs	2015	Web Information Systems Engineering - WISE 2014 Workshops, Book Series: Lecture Notes in Computer Science
Gall, Manuel Wallner, Guenter Kriglstein, Simone Rinderle-Ma, Stefanie	A Study of Different Visualizations for Visualizing Differences in Process Models	2015	Advances in Conceptual Modeling, Book Series: Lecture Notes in Computer Science
Slaninova, Katerina Martinovic, Jan Drazdilova, Pavla Snashel, Vaclav	From Moodle Log File to the Students Network	2014	International Joint Conference SOCO '13-CISIS'13-ICEUTE'13, Book Series: Advances in Intelligent Systems and Computing
Leemans, Sander J.J. Fahland, Dirk van der Aalst, Wil M.P.	Process and Deviation Exploration with Inductive visual Miner	2014	CEUR Workshop Proceedings

Lucas, Wendy Xu, Jennifer Babaian, Tamara	Visualizing ERP Usage Logs in Real Time	2013	ICEIS: Proceedings of the 15th International Conference on Enterprise Information Systems
Bose, R.P. Jagadeesh Chandra van der Aalst, Wil M.P.	Discovering Signature Patterns from Event Logs	2013	Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SCI 2013
Kecman, Pavle Goverde, Rob M. P.	Process mining of train describer event data and automatic conflict identification	2012	WIT Transactions on The Built Environment
van Dongen, Boudewijn F. Adriansyah, Arya	Process Mining: Fuzzy Clustering and Performance Visualization	2010	Business Process Management Workshops, BPM 2009, Book Series: Lecture Notes in Business Information Processing

II Instructions of the Framework

This version of the framework is meant to be used in the design process of process diagrams.

Before using the framework, the designer should know the purpose of the diagram as well as attributes that are planned to be visualized.

The framework is presented in an hierarchical structure, where high-level questions are divided to sub-questions. The user should answer only the last level sub-questions – the questions, which are accompanied with possible answers.

The answers are presented either as radio-buttons (select one), select-boxes (select one or several) or blank spaces to fill in. In addition, strengths and weaknesses of the answers are listed where possible. The strengths and weaknesses are in a table format, where strengths are on the left (with bullet-points marked with “+”) and weaknesses on the right (with bullet-points marked with “-“).

Question 1.2.2 “Which channels express the attributes?” is exceptional. It is not answered by alternative options. Instead, visual channels for ordinal and categorical attributes are listed in order of effectiveness. In addition, there is a list of common practices of the application of the visual channels on process maps. The user can use those lists as a point of inspiration.

The two first questions are answered for the user, because the current version of the framework is designed according to those pre-selected choices – the framework helps to make design-decisions about node-link diagrams (1.1.1), where nodes and links are separated from one another (1.1.2).

It is recommended to follow the proposed sequence of the framework, but it is not required.

The user does not have to answer all the questions. The questions that are not relevant to the task at hand should be left unanswered.

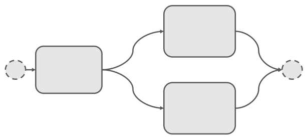
The output of the framework is a collection of design decisions that need to be developed further into a diagram by the user. The output of the framework is not a ready-made solution for the diagram.

III Framework

1. How to encode data?

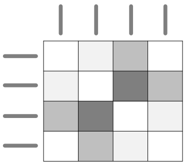
1.1 How to arrange data?

1.1.1 What is the base diagram?



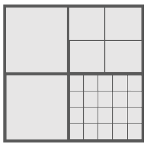
✓ Node-link diagram

+ used for the task of understanding the topology of a process; + used for the task of discovering hierarchy of processes (tree diagrams); + intuitive.	- not easily scalable; - space-consuming; - complex diagrams impose a cognitive overload; - occlusion.



○ Adjacency matrix

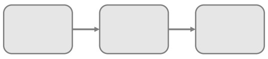
+ scalable; + used for the task of identifying activities and estimating number of links;	- unfamiliar for most users; - not possible to find multiple-link paths – not useable for topology tasks.



○ Enclosure

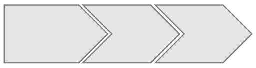
+ scalable; + used for the task of discovering hierarchy of processes; + intuitive.	- not possible to detect a sequence – not useable for topology tasks.

1.1.2 What are the basic elements of the diagram?



✓ Separated nodes and links

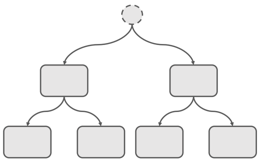
+ visualizes complex process flows, including rework; + allows drilling down to details; + better scalability than the merged version.	- requires more space than the merged version.



○ Merged nodes and links

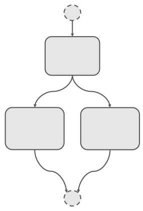
+ used for high-level process flows; + little cognitive load on the user – less elements than separated version; + compact.	- not easily scalable; - usually shows only one direction of the flow.

1.1.3 How are the basic elements ordered?



○ Hierarchical

+ used for the task of discovering hierarchy of processes.	- does not show the relative timing or sequence of the activities.

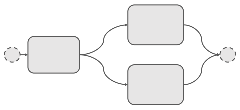


○ Sequential

+ shows relative timing of activities; + commonly used in process mining.	- detailed and coarse parts of the flow are mixed.

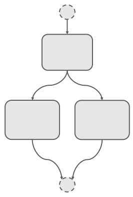
1.1.3.1 How is the sequence of the process shown?

□ Orientation of the diagram:



○ From left to right

+ intuitive in English environment – the same direction as reading text.	- difficult to scroll with a mouse.

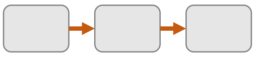


○ From up to down

+ easy to scroll.	

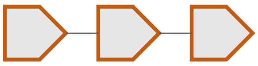
□ Directional shapes of elements:

□ Links shaped as arrows



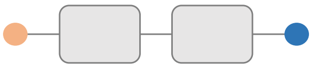
+ space-saving, while still noticeable.	

□ Nodes shaped as arrows



+ larger and more noticeable than links with arrows.	- shape channel of nodes cannot be used for anything else.

□ Other: …



□ Start and end nodes:

□ Encoding of the start node (color, shape, etc): …

□ Encoding of the end node (color, shape, etc): …

□ Other: …

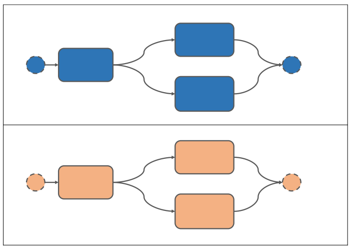
1.1.4 How is the diagram aligned?

1.1.4.1 How many processes are shown?

○ One

○ Many

1.1.4.1.1 How are the process diagrams faceted?



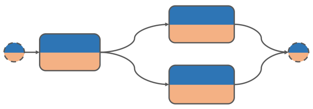
□ Juxtaposed:

+ topology of each separate process is easy to understand.	- more cognitive load on the user than in superimposed layers when used for comparison as the eyes have to travel from one diagram to another to spot the differences.

○ Vertical

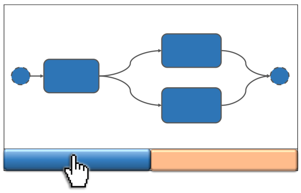
○ Horizontal

○ Matrix



□ Superimposed layers

+ easy to use for comparison purposes.	- requires attentive design of highlighting differences and other metrics; - difficult to understand topology of each separate process.



□ Separate views

+ used for faceting alternative visualizations of the same process model.	- not recommended for comparison purposes as it imposes a great cognitive load on the user memory.

□ Other: …

1.1.4.2 What is the alignment based on?

□ Best fit of proximity

+ space-saving; + easy to compute.	- sometimes proximity carries a meaning due to random chance, sometimes it is arbitrary and can lead to false conclusions.

□ Semantic meaning: …

+ uses space to represent another dimension of data.	- adds to visual clutter, especially if grouping elements, such as containment marks are included; - space-consuming; - computationally demanding.

□ Other: …

1.1.4.3 Is the layout deterministic or nondeterministic?

○ Deterministic

+ easy to reference elements based on their location.	- computationally demanding.

○ Nondeterministic

+ computationally less demanding than deterministic layout.	- the user must familiarize himself/herself with the layout after every loading.

1.2 How to map data?

1.2.1 Which attributes are shown on the diagram?

□ Categorical: …

□ Ordered: …

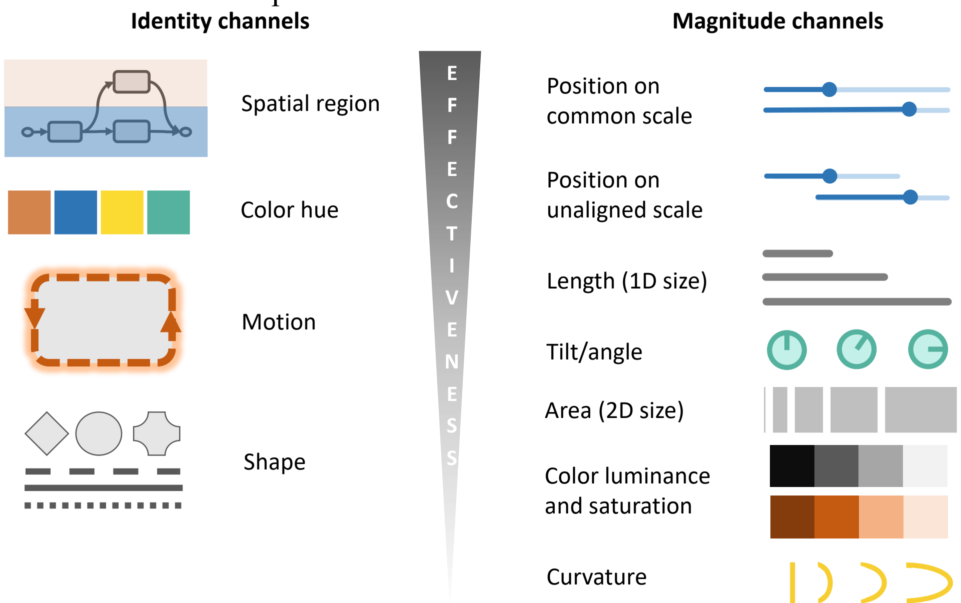
1.2.1.1 What is the direction of ordering?

○ Sequential

○ Diverging

○ Cyclic

1.2.2 Which channels express the attributes?



*The most important attributes should be shown with the most effective channels (on top).
**Equally important attributes can be expressed with the same channel and the data can be faceted into exclusive layers the user can choose between.

□ Identity channels: …

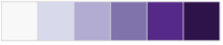
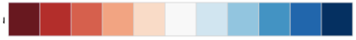
Common practices:

Shape	The most commonly used visual channel for categorical attributes on process diagrams because this channel is available without constraining the use of other channels. Attribute level can be communicated as follows: shape of nodes (circle, square, etc) ; shape of edge (continuous, dashed, etc); a symbol placed on a node or an edge.
Spatial region	The use of spatial region is limited due to the sequential quality of the process – ordering is already communicating the sequence

	dimension. The rest of the spatial region can be mapped in following ways (look also alignment section 1.1.4.2): vertical alignment of nodes (for horizontally oriented diagrams); horizontal alignment of nodes (for vertically oriented diagrams); adding containment marks; placing connected activities in close proximity, such as parallel activities.
Color hue *with any use of color it is important to make sure that it is visible for color blind users: http://www.color-blindness.com/coblis-color-blindness-simulator/	Color hue is often used to highlight the matches or mismatches in the process flows, when two or more processes are compared. It is a pop-out for the user to immediately identify issues. Color hue is also used in combination with color saturation for ordinal variables, when more than one performance metric is assigned to color saturation channel, e.g. shades of blue on the nodes express processing time, while shades of orange express throughput.
Motion	Used to visualize individual process instances in an animated layer (look section 1.2.3.2). It is a very strong, but underexplored channel in data visualization, which makes it prompt to misuse.

□ Magnitude channels: …

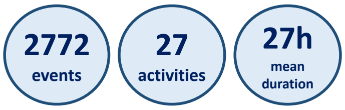
Common practices:

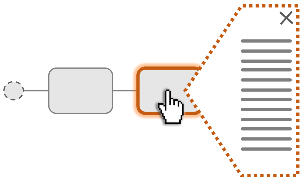
Color saturation and luminance *with the use of color saturation and luminance it is important to ensure the visibility of other elements, such as labels.	Saturation and luminance are commonly used on nodes, expressing data about activities – the darker the shade, the higher value. Color coding on nodes is stronger than on links, because it's a larger area (visible also when zoomed out). Encoding can be: sequential  diverging 
Area	Area channel is often used on links – the thicker the line, the higher the value. Nodes can be enriched by area marks when layering other types

	of diagrams on the nodes, such as pie charts.
Length	Length of links and/or nodes can show waiting and processing times. This approach offers a strong pop-out of outliers (long waiting or processing times), but it requires a lot of screen space and may not be useful for exploring the topology of the process as the diagram becomes too stretched out to get an overview.
Position on common scale/ unaligned scale	Some visualizations have dashboard diagrams (bar charts, line charts, etc) integrated into the process flow diagram to compare performance of process cohorts or activities. This can be done by placing charts on top of or next to the nodes.
Other channels	Other channels are less commonly used. Some additional channels that are not listed have also been used for showing magnitude, such as levels of blur and transparency.

□ Textual sets: …

Common practices:





Process overview statistics	Process overview statistics, such as average process time or total throughput, are usually shown in a separate area of the view, not layered on top of the diagram.
Element statistics	Various statistics of activities, such as total throughput and throughput of unique instances, are usually marked as labels and/or embedded into the diagram elements (look sections 1.2.3 and 1.2.4 for embedding and labelling).

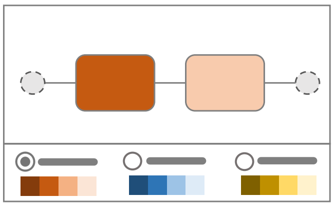
1.2.3 How is the data faceted on the diagram?

□ Superimposed layers: …

1.2.3.1 Which channels and attributes are visible in each layer?

□ All layers: …

□ Layer 1: …, layer 2: …, …, layer n: …



1.2.3.2 Are there animated layers?

1.2.3.2.1 Which elements are shown with movement?

□ Process instance path

□ Process instance status

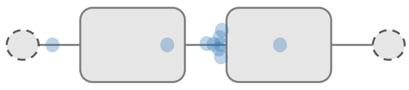
□ Other: …

1.2.3.2.2 How are the animated elements mapped?

- □ Shape: …
- □ Color: …
- □ Size: …
- □ Motion: …
- □ Other: …

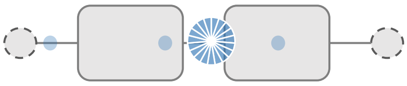
1.2.3.2.3 How to solve occlusion?

- □ Transparency



+ the user can distinct separate instances.	- not possible to estimate the number of instances after the opacity level is 100% due to overlaps; - color conflicts of overlapping items, when color coding is used.

- □ Merging moving items



+ better scalability than transparency.	- the user cannot easily distinct separate instances.

- □ Other: …

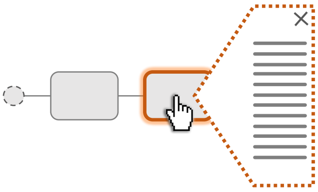
1.2.3.3 Can the user see the diagram without layers?

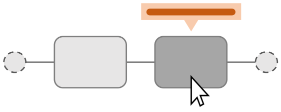
- ○ Yes

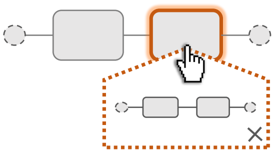
+ lessens visual distraction for topology-specific tasks.	- additional layer choice for the user – adds to the complexity of the diagram.

- ○ No

+ less complex set of choices.	- visual distraction for tasks that require analyzing the topology of the process







□ Embedded data: …

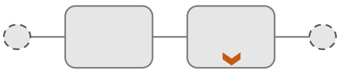
1.2.3.4 What is embedded?

- □ Attribute values
- □ Labels
- □ Sub-processes
- □ Other: …

1.2.3.5 Where is data embedded?

☐ Nodes
☐ Links
☐ Other: …

1.2.3.6 Is there an indicator showing the embedding point?



○ Yes

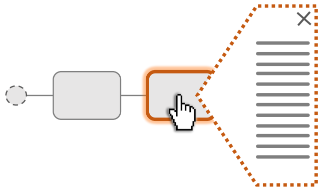
+ gives a hint of embedding to the user.	- additional elements add to the visual complexity of the process diagram.

☐ Shape: …
☐ Color: …
☐ Other: …

○ No

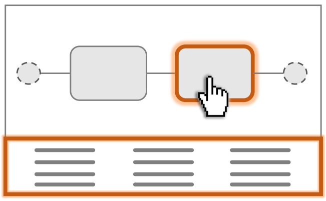
+ less complex diagram.	- the user has to discover the embedded data by experimenting.

1.2.3.7 Where does the embedded data appear?



○ On the diagram

+ element and embedded data are close – easy for eyes to track.	- pop-up windows occlude parts of the base diagram.



○ Off the diagram

+ the full process is in the view when the details are shown.	- space-consuming.

2772
events
27
activities
27h
mean
duration

☐ Off the diagram: …

+ more data can be encoded into one view.	- additional sections in the view take space from the main diagram.

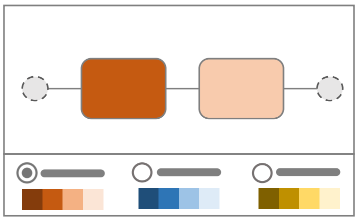
1.2.4 How does the user know the meaning of channels?

☐ Legend:

1.2.4.1 Which channels and values are shown on the legend?

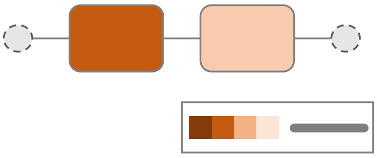
☐ Channels: …
☐ Values: …

1.2.4.2 Is legend separate or integrated into the control panel?



□ Integrated into the control panel

+ space-saving; + faster to use than separate area version – selecting and understanding the encoding is done as one action.	- more difficult to identify info than in a separate area version as the legend is mixed with control panel widgets.



□ Separate area

+ easy to use – a conventional way.	- space-consuming; - scatters user's focus between diagram, control panel and legend.

1.2.4.2.1 Is the legend dynamic or static?

○ Dynamic – includes only encoding of the selected layer

+ space-saving; + faster to identify encodings of interest than in a static version.	- works against visual memory – the user needs to understand the legend again every time it changes.

○ Static – same legend for all the layers

+ only one layout of the legend supports user's visual memory; + gives an overview of all attributes.	- space-consuming; - difficult to identify info of interest amongst many encodings.

□ Labels:

1.2.4.3 Which labels are visible?

□ All the time: …

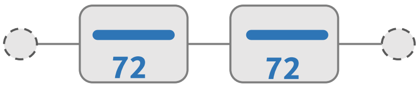
□ Layer 1: … , layer 2: … , …, layer n: …

□ Embedded (hover, click): …

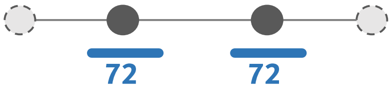
□ Other: …

1.2.4.4 Where are the labels placed?

□ On nodes: …

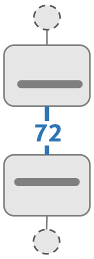


+ clearly understandable which node the label belongs to.	- the node must fit the text of the label – constrains the size and shape channel of the nodes; - the label must be visible – constrains the color channels of the nodes.



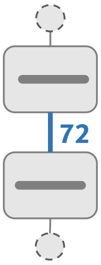
□ Next to nodes: …

+ does not constrain the visual channels of the nodes.	- in complex processes difficult to match the label with the node; - additional elements add to the visual complexity of the process diagram.



□ On links: …

+ clearly understandable which link the label belongs to.	- the labels need additional background encoding to be visible, which occludes the links.



□ Next to links: …

+ does not occlude the links.	- in complex processes difficult to match the label with the link; - additional elements add to the visual complexity of the diagram.

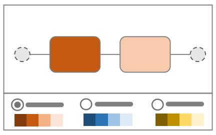
□ Other: …

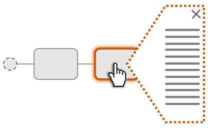
1.2.4.5 How to guarantee the readability of labels?

□ Color is matched with other colors on the diagram
□ Readable size
□ Semantic zooming (look zooming section)
□ Magnified when hovered
□ Other: …

2. How to design interaction?

2.1 How can the user change the visualization?

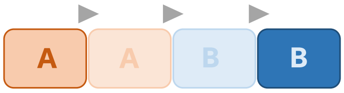




2.1.1 What can be changed on the diagram?

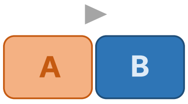
- □ Layers:
  - □ Data: …
  - □ Encoding: …
- □ Embedded data:
  - □ Data: …
  - □ Encoding : …
- □ Other: …

2.1.2 How do the changes appear?



- □ Animated transitions: …

+ keeps the connection between changed elements; + guides the focus of the user if only few elements change.	- confuses the focus of the user when many elements change; - may lead to false conclusions if the animation does not follow semantics of the data.



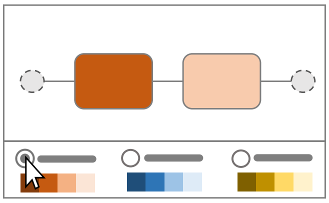
- □ Jump cuts: …

+ quick.	- the connection between changed elements is weak.

2.1.3 What is the default appearance?

- □ Basic elements: nodes, links, …
- □ Layer: …
- □ Embedded data: …
- □ Orientation and alignment: …
- □ Other: …

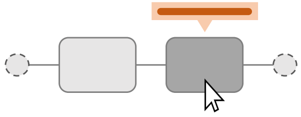
2.1.4 How can the changes be triggered?



2.1.4.1 Where can the user trigger the changes?

- □ Control panel: …

+ gives an overview which changes can be triggered; + helps to keep track on the applied changes.	- space-consuming.



- □ On the visualization: …

+ space-saving.	- triggering the changes discovered by experimenting.

2.1.4.1.1 Which actions trigger changes on the diagram?

- □ Hover: …
- □ Click: …
- □ Double click: …
- □ Drag: …
- □ Scroll: …
- □ Touchpad gestures: …
- □ Other: …

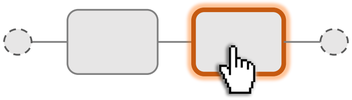
□ Keyboard shortcuts: …

+ space-saving.	- triggering the changes discovered by experimenting.

□ Other:…

2.1.4.2 How does the user get feedback to the actions?

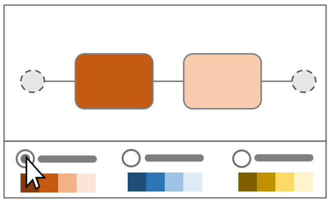
□ Highlight: …



+ helps the user to evaluate if their selection matches with their intention; + used if several elements can be selected or if the element requires deselecting; + used to link data in various places on the view.	- additional elements add to the visual complexity of the diagram; - can collide with existing encoding.

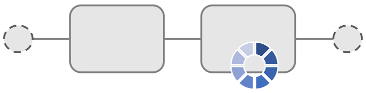
□ Color: …
□ Shape: …
□ Motion: …
□ Other: …

□ Immediate change: …



+ quick if only few configurations need to be changed.	- slow if there are several configurations to be changed as every selection makes the diagram load a new version.

□ Progress indicator: …



+ used if the change takes longer than user would expect.	

□ Other: …

2.1.4.3 How can the user undo the change?

□ Deselect: …
□ Select something else: …
□ Back button: …
□ Close button: …
□ Click elsewhere: …
□ Other: …

2.2 How can the user reduce data?

2.2.1 Does the diagram need panning?

2.2.1.1 How far can the user pan?

□ Default: …
□ Up-down: …
□ Left-right: …

2.2.1.2 Which manipulation actions are for panning?

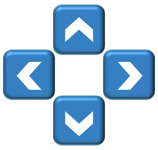
□ Scroll
□ Touchpad gestures: …
□ Keyboard arrows
□ Pinch and drag
□ Other: …

2.2.1.3 Which control elements are for panning?

□ Scrollbars

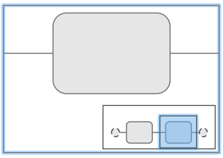
+ compact;
+ intuitive;
+ allow quick panning.	

□ Move buttons



+ compact.	- only slow (step-by-step) panning.

□ Overview-detail pane



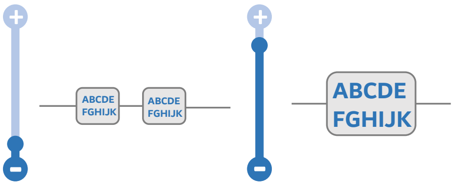
+ aids navigation in complex diagrams; + intuitive; + allows quick panning.	- space-consuming; - requires abstraction design in the overview panel.

□ Other: …

2.2.2 Does the diagram need zooming?

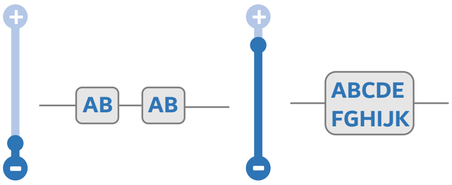
2.2.2.1 What type of zooming?

○ Geometric



+ intuitive.	- labels and visual channels lose readability when zoomed out.

○ Semantic



+ all the important elements of the diagram are visible when the diagram is zoomed out.	- additional design for elements on each level of zoom; - difficult to find a general way to shorten the activity names or other textual elements.

2.2.2.2 How close or far can the user zoom?

□ Default: …
□ The closest: …
□ The furthest: …

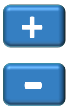
2.2.2.3 Which manipulation actions are used for zooming?

□ Scroll
□ Double click
□ Touchpad gestures: …
□ Keyboard shortcuts: …
□ Other: …

2.2.2.4 Which control elements are for zooming?

□ Slider

+ compact;
+ intuitive;
+ allow quick zooming.	



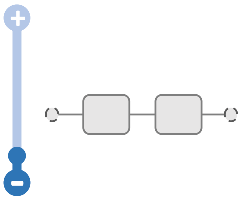
□ Zoom buttons

+ compact.	- only slow (step-by-step) zooming.

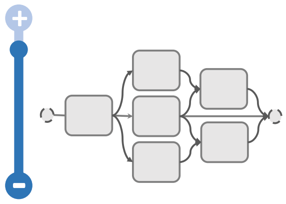
□ Other: …

2.2.3 Does the diagram need abstracting?

2.2.3.1 What type of abstraction?



□ Number of paths
□ Number of activities
□ Other: …



2.2.3.2 How simple or complex can the diagram be?

□ Default: …
□ Minimum number of nodes and links: …
□ Maximum number of nodes and links: …

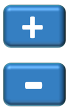
2.2.3.3 Which manipulation actions are used for abstracting?

□ Touchpad gestures: …
□ Keyboard shortcuts: …
□ Other: …

2.2.3.4 Which control elements are for abstracting?

□ Slider

+ compact;
+ intuitive;
+ allow quick abstracting.	



□ Abstraction buttons

+ compact.	- only slow (step-by-step) abstracting.

□ Other: …

2.2.4 Does the diagram need filtering?

2.2.4.1 Which filters can the user apply?

□ Attributes: …
□ Values: …

2.2.4.2 How many filters can the user apply?

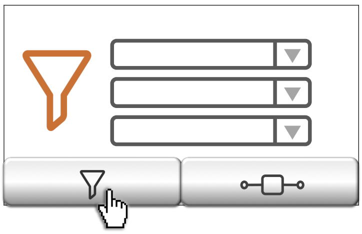
○ One

+ easy to keep track on the filters.	- does not support complex analytical tasks.

○ Many

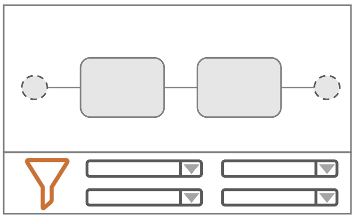
+ allows filtering for complex analytical tasks.	- requires visual aid for remembering applied filters; - computationally more complex.

2.2.4.3 Where can the user apply filters?



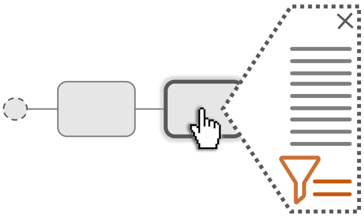
□ Separate filter view

+ used for advanced filtering as it allows enough space for all possible filtering options.	- user has to navigate to another view to apply filters; - user has to switch from process layout of the items to list layout.



□ Control panel for filtering on the diagram view

+ user does not have to navigate between views; + user can see both, process layout as well as list layout of items.	- space-consuming; - requires a concise composition of complex filters.

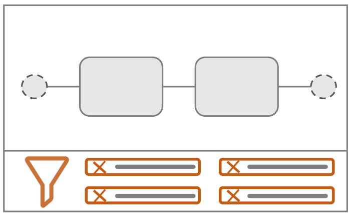


□ Shortcuts on the diagram

+ user does not have to navigate between views or control panel and diagram; + user does not have to transition from process layout to list layout of items.	- not obvious, where and how to filter as user has to find the filtering shortcuts by experimenting; - does not allow to apply complex multi-level filters.

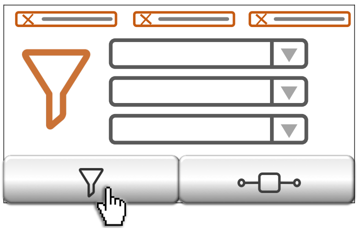
□ Other: …

2.2.4.4 How can the user keep track on the applied filters?



□ Overview on the main view

+ helps user to keep track on applied filters without any additional navigation; + user does not have to remember applied filters when using diagram view.	- space-consuming; - requires a concise composition.



□ On the filter view

+ space-saving; + applied filters do not have to be summarized concisely, but can be shown in full complexity.	- user has to navigate to another view to see the filters; - user has to remember the filters when using diagram view.

□ Other: …

IV Interview Questions

Questions for interview 1:

Please tell me a little bit about your professional background and your previous experience in process mining.

What is the project you are developing now? What is your role in it?

How is the process of visualizing the algorithms you are developing? Have you done visualizations in other projects? How has the process been conducted there?

What are the struggles in the current visualization process? What have been struggles in other projects?

Are you aware of any visualization frameworks or theory that could help to improve the process? Why have you used/not used those frameworks?

Questions for interview 2:

Understandability:
Was it clear how to use the framework? What was unclear?
Were the terms and illustrations understandable? What was confusing?
What do you think is the purpose of this framework?

Relevance:
Did the framework fulfil its purpose in this workshop?
How did it help the visualization process?
Would you recommend the framework to your colleagues?

Completeness:
What would you change or take out of the framework? What was missing from the framework?
How far did you get with your visualization? Which aspects do you still have to work on?

Usefulness:
Do you think it was easy to use the framework or did you have to put a lot of effort into using it?
Did the time spent match with the benefits you got from the framework? (Was the time you spent on the framework worth it?)

IV Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Marit Sirgmets**,
(author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

A Visualization Framework for Designing Process Mining Diagrams,
(title of thesis)

supervised by Fredrik Payman Milani, Taivo Pungas,
(supervisor's name)

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **08.05.2018**