UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

VINICIUS STEIN DANI

Visual Feedback about Problems in Business Process Models: Systematic Literature Review, Survey, Case Studies and Recommendations

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Prof. Dr. Lucinéia Heloisa Thom
Coadvisor: Prof. Dr. Carla Maria Dal Sasso Freitas

Porto Alegre
August 2019

CIP — CATALOGING-IN-PUBLICATION

Dani, Vinicius Stein

Visual Feedback about Problems in Business Process Models: Systematic Literature Review, Survey, Case Studies and Recommendations / Vinicius Stein Dani. – Porto Alegre: PPGC da UFRGS, 2019.

131 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2019. Advisor: Lucinéia Heloisa Thom; Coadvisor: Carla Maria Dal Sasso Freitas.

1. Business process management. 2. Process model problems. 3. Visual feedback. 4. BPMN. I. Thom, Lucinéia Heloisa. II. Freitas, Carla Maria Dal Sasso. III. Título.

"Nobody said it was easy,
no one ever said it would be this hard."
— Chris Martin and Band

"You can edit a bad page,
but you can't edit a blank page."
— Jodi Picoult

"It's not about how hard you hit.
It's about how hard you can get hit
and keep moving forward."
— Rocky Balboa

AGRADECIMENTOS

Agradeço aos meus pais, por todo carinho e apoio incondicional. Agradeço também à CAPES e ao PPGC-UFRGS por terem confiado a mim valiosos recursos para que eu pudesse desempenhar este trabalho.

Agradeço às minhas orientadoras pela dedicação e aos demais professores do INF-UFRGS dos quais tive a oportunidade de ser aluno. Agradeço muito aos queridos colegas do Laboratório “doscientas quinze” e especialmente aos amigos Francine Ianiski, Gustavo Guerra, Ivan Friess e Leonardo Seivald pelas parcerias. Também agradeço às amigas Liza Lunardi e Carol Kaiser pela convivência diária no 1908D.

Como sempre costumo dizer: embora eu seja contra religião, acredito que há uma força divina usualmente chamada de “Deus” e a Ele(a) (seja quem e/ou o que for) dedico os meu mais profundos agradecimentos. Principalmente pela perseverança que Ele me deu para progredir no mestrado diante das dificuldades que se apresentaram e que poderiam ter me feito desistir. Por fim, agradeço, também, aos demais familiares, amigos, professores e colegas os quais me apoiaram nos momentos difíceis mas de inestimável crescimento pessoal e profissional.

ABSTRACT

Business process modeling is an essential task in business process management. Process models that are comprehensively understood by business stakeholders allow organizations to profit from this field. However, when not correctly modeled, process models may hamper businesses profitability. In this work, we explored and reported what is being investigated in the topic *visualization of business process models*, since visualization is known as improving perception and comprehension of structures and patterns in datasets. We performed a systematic literature review (SLR) through which we selected and analyzed 46 papers out of 1686 studies. Based on the findings of our SLR, we conducted a survey with 57 participants, and developed two case studies. From the SLR we concluded that there still are challenges to be explored regarding visual feedback about problems in process models. From the survey we gathered a list of modelers' demands regarding feedback about problems in process models. For example, modelers would like to get feedback about problem according to the modeler level of experience, and be able to activate/deactivate automatic validation. The two case studies complemented the data we gathered from the survey. In our work, we give emphasis to the Business Process Model and Notation (BPMN) because it is an ISO standard. The goal of the first case study was to investigate the behavior of a set of well-known BPMN-based modeling tools concerning the feedback provided to modelers regarding the same set of problems. In the second case study, we investigated the extent to which a set of problems still occur across a set of process models modeled by students and professionals learning about process modeling. Then, we mapped the identified demands to the way that the process modeling tools provide feedback regarding problems in process models as well as to the solutions found in the literature. Finally, from the gaps found in the mapping we propose a set of recommendations for visual feedback about problems in process models.

Keywords: Business process management. Process model problems. Visual feedback. BPMN.

***Feedback* Visual sobre Problemas em Modelos de Processos de Negócio: Revisão Sistemática da Literatura, *Survey*, Estudos de Caso e Recomendações**

RESUMO

A modelagem de processos é uma tarefa essencial no gerenciamento de processos de negócio. Modelos de processo que são compreendidos de forma abrangente pelos participantes do negócio permitem que as organizações lucrem. No entanto, quando não são corretamente modelados, os modelos de processo podem dificultar a rentabilidade das empresas. Neste trabalho, exploramos e relatamos o que está sendo investigado no tópico *visualização de modelos de processos de negócio*, uma vez que a visualização é conhecida por possibilitar a melhoria na percepção e compreensão de estruturas e padrões em conjuntos de dados. Realizamos uma revisão sistemática da literatura (RSL), através da qual selecionamos e analisamos 46 artigos de 1686 estudos. Com base nas conclusões da nossa RSL, realizamos uma *survey* com 57 participantes e desenvolvemos dois estudos de caso. A partir da RSL, concluímos que ainda há desafios a serem explorados em relação ao feedback visual sobre problemas em modelos de processo. A partir da *survey*, reunimos uma lista de demandas de modeladores referentes ao feedback sobre problemas em modelos de processo. Por exemplo, os modeladores gostariam de poder ativar/desativar a validação automática do modelo. Os dois estudos de caso complementaram os dados que coletamos da *survey*. Em nosso trabalho, demos ênfase à Business Process Model and Notation (BPMN), por ser um padrão ISO. O objetivo do primeiro estudo de caso foi investigar o comportamento de um conjunto de ferramentas de modelagem baseadas em BPMN quanto ao feedback fornecido aos modeladores em relação a um mesmo conjunto de problemas. No segundo estudo de caso, investigamos até que ponto um conjunto de problemas ainda ocorre em um conjunto de modelos de processo modelados por estudantes e profissionais que estão aprendendo sobre modelagem de processos. Em seguida, mapeamos as demandas identificadas para o modo como as ferramentas de modelagem de processos fornecem feedback sobre problemas em modelos de processos, bem como para as soluções encontradas na literatura. Finalmente, a partir das lacunas encontradas no mapeamento, propomos um conjunto de recomendações para feedback visual sobre problemas em modelos de processo.

Palavras-chave: Gerenciamento de processos de negócio. Problemas em modelos de processos. *Feedback* visual. BPMN.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

RQ	Research Question
OMG	Object Management Group
UML	Unified Modeling Language
EPC	Event-Driven Process Chain
SLR	Systematic Literature Review
BPM	Business Process Management
YAWL	Yet Another Workflow Language
BPMS	Business Process Management Suite
BPMN	Business Process Model and Notation
ISO	International Organization for Standardisation

CONTENTS

1 INTRODUCTION

Business Process Management (BPM) is a set of methods, techniques, and tools for discovering, analyzing, redesigning, executing and monitoring business processes, and has received considerable attention in recent years due to its potential to increase productivity and reduce costs (DUMAS et al., 2013; Van Der Aalst, 2013). One of the ways organizations can document their business operations and implement reproducible processes as well as continually improve them is through the use of BPM and specific notations for business process modeling. There are a variety of business process modeling notations, such as Business Process Model and Notation (BPMN) (LEOPOLD; MENDLING; GÜNTHER, 2016), Event-driven Process Chains (EPC) (KELLER; NÜTTGENS; SCHEER, 1992), Unified Modeling Language 2.0 Activity Diagrams (UML AD) (FOWLER; SCOTT, 1999), Yet Another Workflow Language (YAWL) (HOFSTEDE; AALST, 2005), Petri Nets (PETRI, 1962) and DECLARE (PESIC; SCHONENBERG; AALST, 2007). In this study, we restrain our focus mainly on BPMN, since this notation is an ISO standard[1] and an OMG specification (Object Management Group, 2015). Moreover, BPMN is adopted by several process modeling tools (MEIDAN et al., 2017).

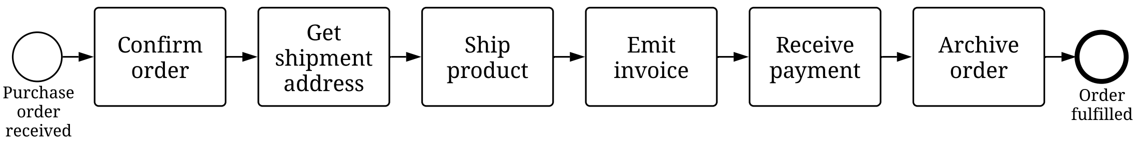
A business process is a collection of activities, events, and decision-making steps, which comprises different resources with the aim of bringing value to the stakeholder (DUMAS et al., 2013). Business processes play an important role in organizations (ROY et al., 2014). Employees from different business and technical departments, not necessarily advanced modelers, are more often involved with process modeling task nowadays (BECKER; ROSEMANN; Von Uthmann, 2000). Such task is known as being challenging to manage (MENDLING; REIJERS; Van Der Aalst, 2009), generally because of the modeling notation's complexity caused by its variety of elements and semantics (LEOPOLD; MENDLING; GÜNTHER, 2016). Beyond that, a business process model supports the understanding of an organization's business processes (RITTGEN, 2010). Designing a process model to represent a real-world process appropriately relies on the modeler expertise or the advice of an experienced modeler.

The business process modeling task aims at supporting the definition and representation of business processes through the identification of a set of activities capable of representing the real world functional behavior of these processes, taking into consideration all the elements of the organization that are involved in the process (e.g., departments,

[1] ISO/IEC 19510:2013: http://www.omg.org/spec/BPMN/ISO/19510/PDF

resources). Through a process model, an organization can reduce communication inconsistencies (BECKER; ROSEMANN; Von Uthmann, 2000). Figure 1.1 exemplifies a simple process model of an “order fulfillment” process, where a set of activities need to be performed in a certain order so the “order is fullfiled” after the process ends. In this example, the process starts after the purchase order is received. The order is confirmed, the shipment address is obtained, the product is sent, and the invoice is emitted; after that, the payment is received and the order archived. The order is finally fulfilled.

Figure 1.1: Simple process model example of an “order fullfilment” process, adapted from (DUMAS et al., 2013).



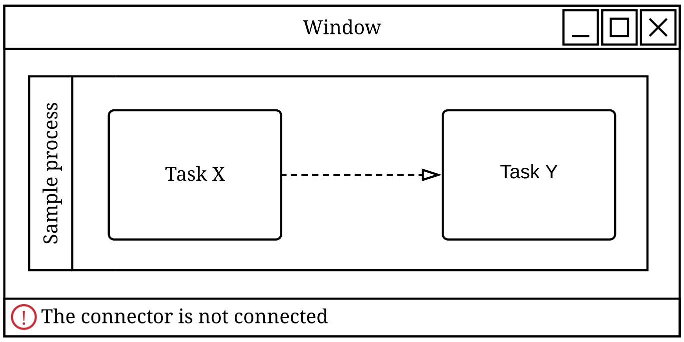
When correctly implemented, process models can generate significant savings for the industry (HAMMER, 2010). On the other hand, modeling problems may generate process execution errors in a production environment, creating extra costs for the organization (GEIGER et al., 2017). According to Goldberg Júnior et al. (JÚNIOR et al., 2018), modelers with less involvement with the process modeling task frequently commit at least one mistake regarding understandability of the process model. Since the process participants and stakeholders do not necessarily hold expertise in process modeling and business processes are often very complex, it is a challenge to find a user-friendly and easy to understand layout of the process (i.e., the different manners the various elements of a model may be distributed in the canvas) (RINDERLE et al., 2006).

Process models can be enriched visually in a variety of ways (ROSA et al., 2011b; ROSA et al., 2011a; YOUSFI et al., 2016), for example, through “omission” of a subset of elements to target on a specific process model part; or, through “graphical highlight” to visually emphasize specific aspects of process model elements. However, few works suggest new approaches to the graphical representation of process modeling issues (SMUTS; BURGER; SCHOLTZ, 2014). In our study, we propose recommendations to fulfill modelers demands regarding feedback about problems in process models.

1.1 Motivation

Process modeling is not a trivial task and the analysis of modeling problems has not received proper attention (ROY et al., 2014). BPMN offers an extensive variety of graphic elements for process modeling, with different options for representing the same process semantics (LEOPOLD; MENDLING; GÜNTHER, 2016). Although process modeling and automation tools help users by detecting some modeling errors, many errors are still not detected (e.g., an initial event is not modeled, join gateways are not used after the use of precedent split gateways, among others) depending on the tool (GEIGER et al., 2017). In addition, errors are generally indicated through non-instructional text messages. Figure 1.2 presents an example of a modeling error message displayed by the Bizagi Modeler tool[2] (version 2.9.0.4): *The Connector is not connected*, referring to the attempt of using a message flow instead of a sequence flow. If there are more connectors in a larger process model, it will not be possible for the modeler to know exactly which connector is problematic. A message flow is the representation of the flow of messages between pools/lanes in BPMN, while the sequence flow represents the order of execution of the activities and events in a process model.

Figure 1.2: Visual feedback indicating a modeling problem as presented by the process modeling tool Bizagi Modeler. The message says *The Connector is not connected*, referring to an attempt of using a message flow instead of a sequence flow.



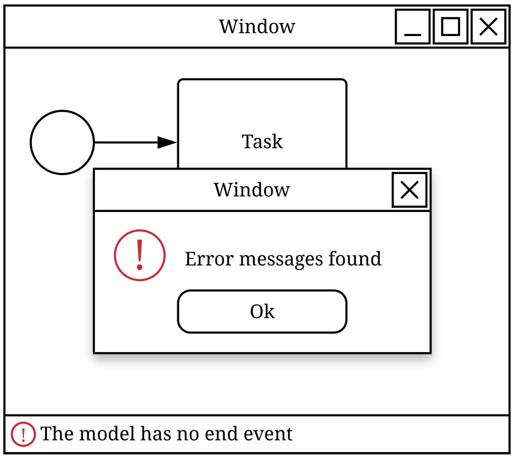
Rozman et al. (ROZMAN; POLANCIC; HORVAT, 2008) identified and analyzed problems in a set of process models modeled by students and the most common ones were defined as *anti-patterns*. More recent studies show that many of these anti-patterns still appear within process models (VIDACIC; STRAHONJA, 2014; SUCHENIA; LIGEZA, 2015; "SUCHENIA et al., 2017). Moreover, the same problem within a BPMN-based process model is presented to the modeler in different ways across a set of well-known

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

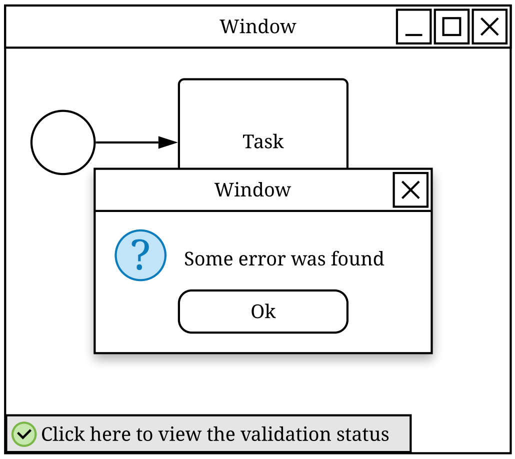
BPMN-based process modeling tools, even though BPMN is a standard (DIAS, 2018). Figure 1.3 represents how four different well-known BPMN-based tools feedback modelers regarding the same process model problem. We use these set of tools in our study.

Figure 1.3: Visual feeback provided by four different well-known BPMN-based tools about the same process model problem. The problem pointed by the tools is the "*Missing end event*" in the model. In this problem, the user did not include the end event of the process. The four tools used are (a) Bizagi [3], (b) Bonita [4], Camunda [5], (c) and (d) Signavio [6].

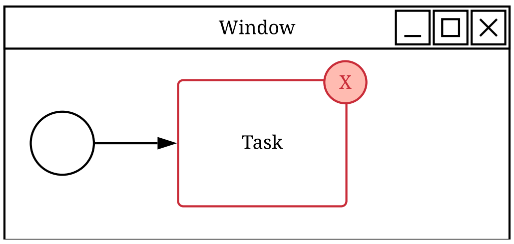
(a) Bizagi v2.9.0.4



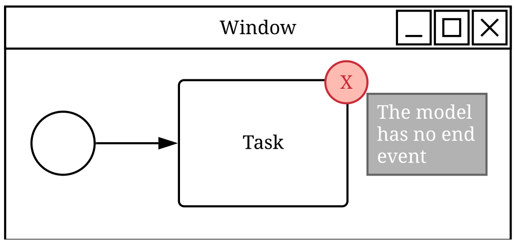
(b) Bonita v7.5.4



(c) Camunda v2.0.3



(d) Signavio v12.6.0



Finally, based on BPM experts experience, it has been noted that these problems still appear frequently among novice undergraduate academic, graduate and postgraduate students, and analysts learning to model using this standard. On the other hand, observing information visualization techniques, there are several techniques for representing data through visual components and attributes. Such techniques aim at supporting users in comprehending data for better performing their tasks (CARD; MACKINLAY; SHNEIDERMAN, 1999; SHNEIDERMAN, 1996). Tufte (TUFTE, 1990) says a visualization should not be judged by the amount of information it displays but how easy it is to understand the information it conveys. This suggests that using information visualization techniques might improve the business process modeling task.

1.2 Objectives and Contributions

Based both on our first findings in the literature and analysis of business process modeling tools and discussion with experts, we started our research with the aim of proposing a set of recommendations regarding the visualization of problems occurring in a business process model that is coherent with the demands of the modelers. To help us achieving this goal, we defined two main objectives: (i) to identify what has been investigated in the topic visualization of business process models; and, (ii) to identify what are the modelers' demands regarding feedback about problems in process models.

Initially, regarding the first objective, we conducted a systematic literature review (SLR) on visualization of business process models. We aimed at identifying gaps to be explored and confirm if there still are challenges for studying feedback about problems in process models. The results of our SLR corroborated the need of further exploration of feedback about problems in business process modeling. Then, towards our second objective, we performed a survey to identify process modelers' demands concerning feedback about problems in process models. Also, we wanted to measure modelers' satisfaction and learnability regarding the feedback provided by the process modeling tools they use. From that, we also learned that there are research opportunities for improvements in this field.

After conducting the survey, our findings showed us that the modelers are, in general, not satisfied, nor they think they can learn from the feedback provided by the modeling tools they use. Then, we performed two case studies. In the first case study, we investigated the behavior of a set of well-known BPMN-based modeling tools concerning the feedback provided to modelers regarding the same set of problems. In the second one, we investigated the extent to which a set of problems still occur across a set of process models modeled by students and professionals learning about process modeling. Results from both case studies showed that the BPMN-based modeling tools, in general, do not give feedback to modelers in the same way, even when concerning to the same problem; and, the set of problems analyzed within the process models still often occur. These results reinforce the need to advance towards giving better feedback about problems during the process modeling task, feedback more coherent to the modelers' demands.

[3]https://www.bizagi.com/
[4]https://www.bonitasoft.com/
[5]https://www.camunda.com/
[6]https://www.signavio.com/

The contributions of this work are:

- A systematic literature review on the visualization of business process model, providing a state-of-the-art report about the visualization techniques used in business process modeling.
- A list of process modelers' demands of feedback about problems in process models;
- A mapping of the process modelers' demands previously identified to what features the process modeling tools provide, and to the solutions the literature describes, regarding feedback about problems in process models;
- A set of recommendations, based on process modelers' demands, about how tools should give feedback to modelers regarding problems occurring during the business process modeling task.

As a secondary contribution, we provide the questionnaire and interview application process that we used to conduct our survey, to allow future researchers to update our findings.

1.3 Text organization

This dissertation is structured as follows. In Chapter 2, we present the necessary fundamentals for this study, while Chapter 3 describes the systematic literature review about visualization of business process models. Chapter 4 reports the survey we performed with business process modelers, and in Chapter 5, we present two case studies we conducted, and a mapping of the extent to which the modelers demands are being fulfilled by the current solutions either published or available as tools. Our recommendations to represent feedback about problems within process models is presented in Chapter 6. Finally, the last chapter contains our conclusions and suggestions for future work.

2 FUNDAMENTALS AND RELATED WORKS

The necessary fundamentals to understand this work is provided in the Section 2.1. Initially, we present the definition of BPM and BPMN; then, we describe the visualization analysis framework we adopted to analyze the visualization approaches found in the selected papers in our SLR. Finally, we discuss the related works to our study in Section 2.2.

2.1 Fundamentals

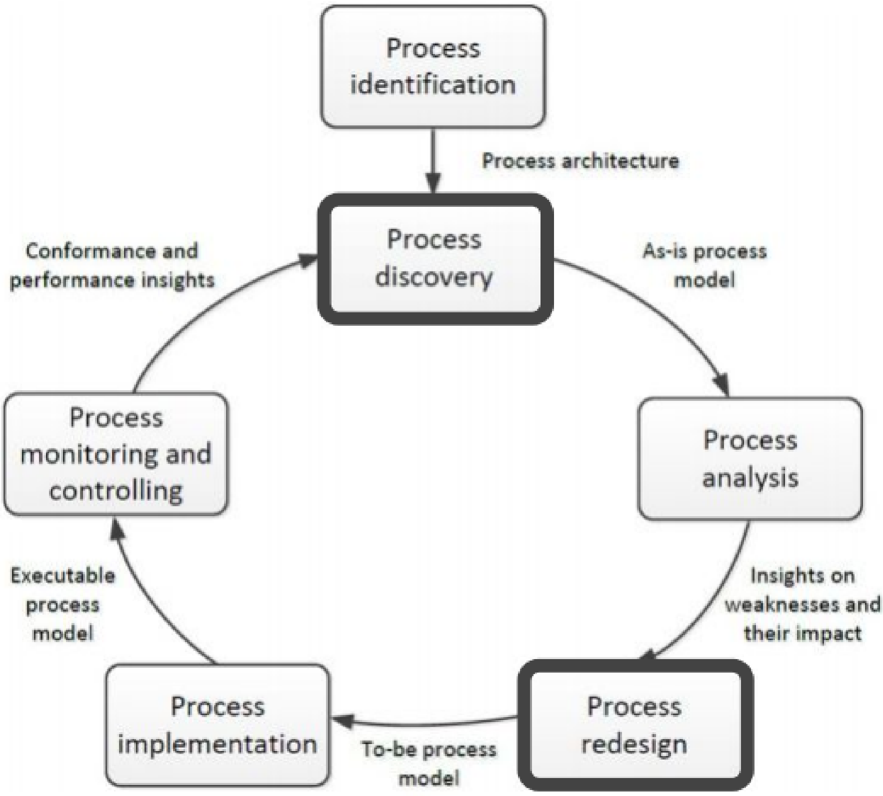
In the following subsections, we present the necessary fundamentals to a proper understanding of our study. We start by introducing BPM, followed by BPMN. Thereafter, we present the visualization analysis framework we utilized.

2.1.1 Business Process Management

BPM is composed of a collection of methods and tools to handle the tasks of modeling, managing and analyzing business processes. A business process is a set of collaborative and dynamically related activities, events, persons, hardware, software, and decision points, with the main objective of delivering value to an organization's customer through a service or a product (WESKE, 2012). When a business process becomes too complex, it can be decomposed into smaller processes, called *sub-processes*, which consist of a subset of elements comprising the process. The most popular process modeling techniques support this concept of sub-process, including BPMN (REIJERS; MENDLING, 2008). The business process modeling task is the process of drawing business processes in a graphical workflow view, aiming at representing the current organization's processes (also known as "as is" processes) to further analyze and improve, achieving new versions of the processes (also known as "to be" processes), which may be implemented and monitored (DUMAS et al., 2013).

Organizations can implement reproducible processes, manage and continually improve them through the use of BPM and the adoption of BPMN to the process modeling task, following the BPM life-cycle (Figure 2.1) proposed by Dumas et al. (DUMAS et al., 2013). The BPM life-cycle consists of six phases, being two of them most directly related to the task of process modeling itself: *process discovery*, where the current state of

Figure 2.1: BPM Life-cycle as proposed by Dumas et al. (DUMAS et al., 2013). The highlighted phases are the ones we focused in our work.



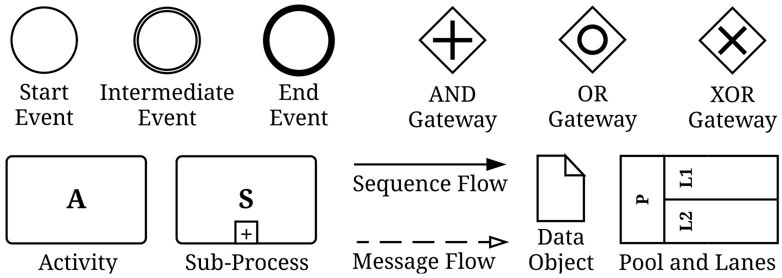
each process is documented in the form of “as is” business process models; and *process redesign*, where the “to be” process models are generated, considering improvement points identified by the analyst.

2.1.1.1 Business process model and notation

The BPMN is the notation often used in the process modeling task (e.g., on the stages of process discovery and process redesign). Initially published in 2004 by the Business Process Management Initiative (BPMI) and maintained by the Object Management Group (OMG) since 2006, BPMN aims at providing an easy-to-understand notation to all business users (e.g., analysts and technical representatives). Approximately 73.22% of the Business Process Management Suites (BPMS) analyzed by (MEIDAN et al., 2017; SARAEIAN; SHIRAZI; MOTAMENI, 2017) enables the automation of business processes modeled with BPMN. BPMSs are tools that support the application of BPM in business environments allowing the automation of business processes and the management of the BPM life-cycle (MEIDAN et al., 2017).

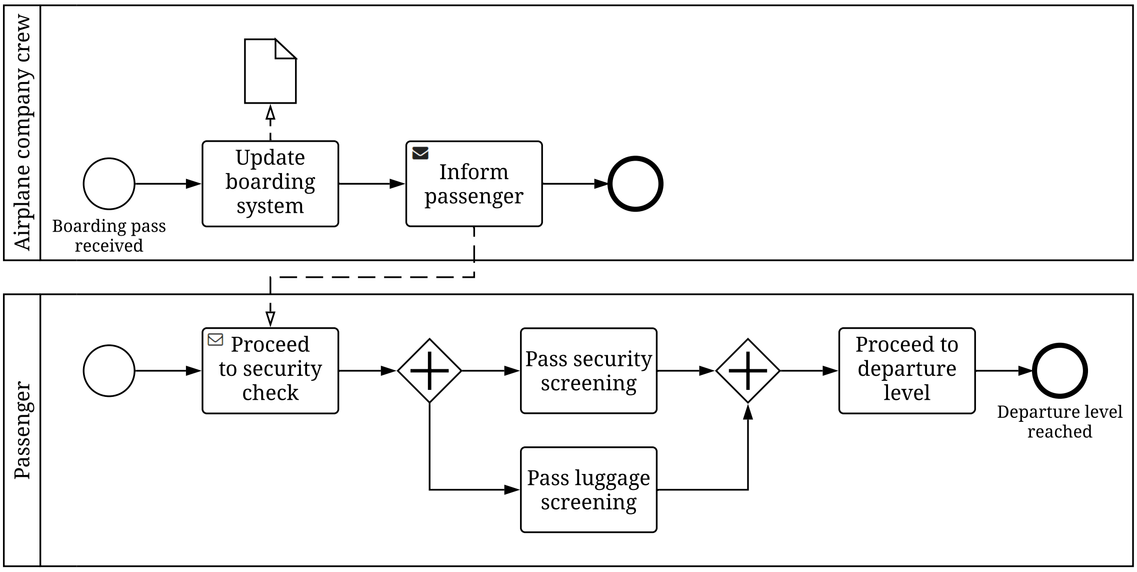
To represent a process, BPMN provides a variety of elements with different purposes (Object Management Group, 2015). The basic BPMN modeling elements are shown

Figure 2.2: BPMN basic elements.



in Figure 2.2. An usage example of the core BPMN elements is shown in Figure 2.3.

Figure 2.3: BPMN usage example.



In BPMN, *events* represent actions that require no duration to be performed. Events are basically of three types, based on when they affect the process flow: start, intermediate or end events. In the process model represented by Figure 2.3, the start event is "Boarding pass received". *Activities*, when seen as a single unit of work, are called "tasks". When a process is too complex (e.g., is composed by more than 50 elements (MENDLING; REIJERS; Van Der Aalst, 2009)), subsets of its elements may be grouped up to comprise sub-processes within the main process. The element "Update boarding system document" (see Figure 2.3) is an example of an activity. In this process model example, the activity "Pass security screening" could be, for example, further detailed and become a sub-process to express more steps comprising the activity.

Gateways are used to split/join the performed actions flow within the process. Also called "decision points", the gateways may be of three types: AND, for concurrency; OR, for inclusive choices; and XOR, for exclusive choices. A gateway is used, for example, when both the passenger and the luggage pass through screening activities (see Figure 2.3). *Sequence flows* are used to link two elements and handle the order through which a process will be executed; and, *message flows* are used to display the flow of messages between two participants. In Figure 2.3, after the “Update boarding system document” activity, a sequence flow is used to guide the process to its next activity which is, in this case, “Proceed to security check”; while the usage of a message flow is when a message is sent from the airplane company crew to the passenger.

Data objects display how data is required or produced by activities. In Figure 2.3, a document is updated by the airplane company crew after they received the boarding pass. *Pools* group together elements of an organization while lanes divide a pool into different organization’s resources (e.g., departments, participants). In our example, both airplane company crew and passenger are pools with one lane each.

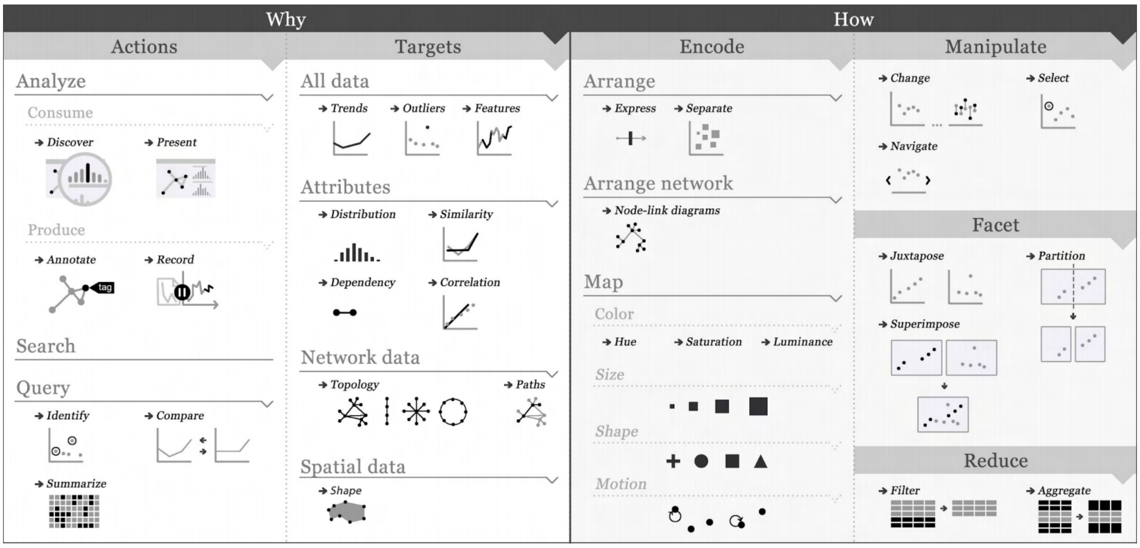
2.1.2 Visualization Analysis Framework

Computer-based visualizations may be obtained through different forms and a variety of techniques and methods (MUNZNER, 2014). Such extensive visualization idiom space (i.e., diverse visualization possibilities to represent similar data) hampers the analysis task of visualization tools in terms of how visualizations are created and how they implement interaction with users.

Munzner (MUNZNER, 2014) proposed a framework that helps researchers to structure the diversity of visualization tools according to abstract elements that can represent generically what each tool is intending to deliver. Therefore, such a framework supports researchers in comparing different visualizations according to their characteristics, guiding the analysis of visualizations usage by means of three questions: (i) **what** is the data that the user is willing to see; (ii) **why** the user intends to use the visualization; and (iii) **how** the visualization is constructed in terms of design choices. Each question tuple "what-why-how" has a corresponding data-task-idiom answer tuple, and the choices in each of these questions are independent of each other. Herein we focus in the "why" and "how" aspects (Figure 2.4) because the "what" part corresponds to the data structure that naturally represents business process models.

To summarize, the framework supports researchers in analyzing visualizations from an abstract point-of-view instead of a domain-specific one, which eases the comparison of different visualizations. According to Munzner (MUNZNER, 2014), when visualizations are analyzed and compared from a domain-specific point-of-view, they appear to be different, which is misleading from a visualization analysis perspective, since there are similarities among different visualizations when they are considered as abstract elements. She also says “the visualization analyst might decide to use additional terms to completely and precisely describe the user’s goals”. So, the framework is composed of, but not limited to, a small set of words to describe the goals of people using a visualization tool and how the idiom of this tool supports people’s goals.

Figure 2.4: “Why” and “How” aspects of the visualization analysis framework proposed by Munzner (MUNZNER, 2014).



In the subsections 2.1.2.1, 2.1.2.2 and 2.1.2.3, we describe the terms associated with each aspect of the framework based on (MUNZNER, 2014).

2.1.2.1 What

In the analysis framework, the answer to “What is the data that user sees?” can be one or more datasets from four possible types: *fields, tables, geometry*, and *networks*. In the context of this work, the data corresponds to business process models that correspond to a dataset type *networks*, or graphs, which are used to represent relationships (links) between items (nodes) (MUNZNER, 2014). In BPMN, for example, a node may represent an activity, and a link, a control flow.

2.1.2.2 Why

To describe “Why the user intends to use the visualization?”, the framework defines *actions* and *targets*. Actions represent possible user goals when using a visualization tool and can be of three types: analyze, search, and query. Each type of action represents different cases that are described as follows.

Firstly, "analyze" may be of two different types: *consume* or *produce* information. The *consume* type is the most common use case and corresponds to the consumption of information already generated. It is divided into three cases: *present*, when the visualization is used to communicate anything already understood by the viewer; *discover*, when the user wants to acquire new knowledge; and, *enjoy*, when the user is driven mostly by curiosity and not by a need. The *produce* type refers to visualizations that enable the user to generate new information. It can be divided into two cases: *annotate*, when the user is allowed to add graphical or textual annotations to visualization elements that already are present in the visualization; and, *record*, when the visualization provides a manner to persist its elements as screen-shots, interaction logs or annotations made by the user.

Secondly, "search" can be of four different types, according to whether the user previously knows (or not) about the target location and identity. These types are: *look up*, when a user knows both location and identity of what he is looking for; *locate*, when a user knows the identity of what he is looking for but does not know its location; *browse*, when a user knows the location of what he is looking for but does not know its identity. For example, when looking for a range of possible items, the user may know where this type of item is but does not know exactly which is the item he is looking for; and, *explore*, when the user does not know the location nor the identity of what he is looking for. It is important to highlight that a visualization may comprise any combination of search types at the same time. For example, the user may see a BPMN model and look up for a specific activity and, at the same time, the user may see the same model and browse for an event.

Finally, "query" can be of three different types: *identify*, when the user identifies a single target among others, the visualization tool returns the target's characteristics; *compare*, differently from identify, refers to multiple targets and allows the user to compare characteristics of these targets; and, *summarize*, that refers to all possible targets within the dataset, and the user obtains an overview of the dataset.

Regarding targets, i.e., the thing that the user presents, looks up or identify, there are four kinds of abstract targets: (i) all data, which refer to what user may retrieve from the dataset as a whole. When targeting all data, the user may find trends, outliers, and features. A *trend* is a behavior that exposes, for example, increases and peaks in a dataset. *Outliers* are data that overstep or stand out in any manner from the rest of the dataset. *Features* are any particular structure of interest in a visualization; (ii) attributes, which are specific properties encoded visually wherein the user may show interest for an individual value (finding *extremes* or *distribution* of values for an attribute) or for multiple attributes (finding *dependencies*, *correlations* and *similarities* between the attributes); (iii) network data, through which the user may find relationships between nodes and links, understand the *network topology* and the existing *paths* between the network's nodes; and, (iv) spatial data, which refers to the visualization of geometric *shapes* and its understanding and comparison.

An important statement from Munzner (MUNZNER, 2014) concerning this framework is that "why a visualization is used doesn't dictate how it is designed".

2.1.2.3 How

"How the visualization is constructed in terms of design choices?" can be answered using a set of options that represent visual forms and/or interaction features. The options are: *encode*, *manipulate*, *facet* and *reduce*. Encoding data within a view can be achieved through different choices for arranging and mapping data. When arranging data, the view may *express* data position distribution over an axis, and *separate* data into regions which have, in their turn, positions distributed along the spatial plane. When mapping data, the visualization designer has different choices such as color, size, shape, and motion. The color space is defined by *hue* (pure color without white and black), *saturation* (amount of white mixed with the pure color) and *lightness* (amount of black mixed with a color). *Size* may be seen from three perspectives: length, which is a one-dimensional size that may be both in width or in height of any given element; area, which is a two-dimensional size; and, volume, a three-dimensional size. *Shape* may be represented by any drawable form using points and lines. *Motion* is represented by the movement of any visual element from one spatial position to another.

Manipulating a view may be performed in three different ways: change, select, and navigate. *Change* refers to any action that makes the way the dataset is being visualized to shift to another way (e.g., switching from a list view to a chart view; or merely switching between different chart views). *Select* refers to the possibility of the user to point out elements of interest. *Navigate* enables the user to, for example, move a large business process model to different directions within the viewport and, thus, visualize a complex dataset that may not fit into the limited screen.

Faceting (or splitting) data over multiple views offers three choices: juxtapose, partition, and superimpose. *Juxtapose* multiple views is when the same data is shown across multiple views, in a coordinated manner, and under different perspectives. *Partition* is when each view, for example, disposed side by side, is composed of a dataset, and represents different data. *Superimpose* is when different views of data are disposed over each other as different layers. Reducing data comprises three design choices: filter, aggregate, and embed. *Filter* refers to the removal of visual elements from the visualization. *Aggregate* refers to group elements that together represent a unique element. *Embed* refers to presenting a selected subset of the data within the same view, where the whole data is presented. For example, enabling a user to select a BPMN collapsed sub-process to display to the user as a tooltip with the sub-process expanded.

2.2 Related Works

Although herein we present the first systematic literature review on the use of visualization in business process modeling, we found some works that identified mechanisms and visual representations used either for reducing the perceived complexity of business process models or serving as components of visual embellishment of such models.

La Rosa et al. (ROSA et al., 2011b; ROSA et al., 2011a) explore mechanisms to reduce the perceived complexity of process models through visual representations of the model. In their work, they identify and present sets of patterns that generalize existing mechanisms with the aim of simplifying the representation of process models. These patterns were gathered from a review of the BPM literature and existing or proposed standards by OMG and W3C, for example, followed by a survey of the identified patterns by BPM experts. For each identified pattern the authors found more than five languages, research approaches or tools which use them. Some examples of the patterns collected are: “enclosure highlight”, which aims to visually enhance a set of model elements based on properties shared among the elements; “pictorial annotation”, aiming at adding, for example, domain-specific information to the model (e.g., indicate criticality through annotating a task with an exclamation mark); “naming guidance”, in order to transmit domain-specific information through nomenclature conventions; “merging”, with the purpose of consolidating a family of variants of process models into a single reference model, without redundancies; and, “extension”, aiming at making a model more straightforward to understand for a specific audience by extending a modeling language to adapt it to a given application domain.

Another related work is by Aysolmaz and Reijers (AYSOLMAZ; REIJERS, 2016), where eight possible components of visual embellishment of process models are identified.

According to the authors, these components are still to be developed and exploited to reinvigorate process models visually. Examples of such components are: "usage of narration and on-screen text", to integrate narration and on-screen text using animation and visualization techniques, and "embedding process perspectives", to integrate different perspectives to a process model also with the use of animation and visualization techniques.

The main differences between these works and ours are that they are not based on a systematic review of the literature and have a different focus. In the case of the works by Rosa et al. (ROSA et al., 2011b; ROSA et al., 2011a), the focus is on presenting an assessment of existing languages and tools regarding the identified patterns. For example, they show that for "pictorial annotation", tools like JDeveloper and Protos automatically assign icons and images to elements of process models, but do not allow customization. As for the work by Aysolmaz and Reijers (AYSOLMAZ; REIJERS, 2016), the proposed categorization is about components to be explored, and not about what is being investigated concerning the visualization of process models. Moreover, the proposal of these possible components for process model embellishment is not backed up by works from others.

2.3 Final comments

In this chapter, we presented the fundamentals necessary to the understanding of our work. We described the most important concepts: BPM, BPMN, and the adopted visualization analysis framework. BPM and BPMN permeated the entire work, and the visualization analysis framework was used more on the analysis of the studies gathered during our SLR. Finally, we end the chapter with a brief review of the works most related to ours.

3 VISUALIZATION OF BUSINESS PROCESS MODELS

This chapter presents the SLR that was performed to identify studies that have been published regarding visualization of business process models in the last ten years (i.e., between January 2009 and December 2018), since BPMN 2.0 had its beta version release on 2009. Our goal with this SLR was to identify and report existing open research questions to be further explored in the field. Among these studies, we wanted to find out how many were addressing visualization of problems in process models, and to which extent they were doing so.

To the best of our knowledge, no SLR nor systematic mapping has been done to investigate this research topic. Although we know that a broad SLR could start with surveying papers on visualization of conceptual models because there is a need on that (GULDEN; REIJERS, 2015), we restrain our focus mainly, but not exclusively, on articles that based their proposals on BPMN, since this notation is an ISO standard[1] and an OMG specification (Object Management Group, 2015).

The review provides a state-of-the-art report on the use of visualization techniques in business process modeling using BPMN. To do this, we analyzed several studies from two perspectives: (i) the first perspective is based on the observation of the similarities between the selected papers regarding their main scope. We aim at answering what the studies are proposing in the topic of visualization of business process models; (ii) the analysis from the second perspective intended to find the relationships among the papers according to the visualization analysis framework proposed by Munzner (MUNZNER, 2014). We targeted answering questions like why the users use the visualizations and how the information about the process models is encoded.

This chapter reproduces the article we published recently (DANI; FREITAS; THOM, 2019). In Section 3.1 we describe the methodology adopted to conduct the SLR, while in Section 3.2 we propose the classification of the studies into six categories according to the analysis of their main scope. Section 3.3 presents the high-level visualization analysis of the studies, whereas Section 3.4 analyze and discuss the results.

[1]ISO/IEC 19510:2013: http://www.omg.org/spec/BPMN/ISO/19510/PDF

3.1 Methodology

A SLR aims at summarizing the topic being studied and identifying the existence of gaps in current research to position new research activities. We conducted our systematic literature review following Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007) to summarize the research on visualization applied to business process modeling aiming at identifying, selecting, evaluating and interpreting the works we considered relevant in this topic. The gaps identified and the report of the analysis of our results are discussed in Section 3.4.

Before starting our systematic literature review, we conducted a preliminary research which provided us with a variety of papers exploring visualization of business process models. After that, we decided to investigate what is being studied and developed in the topic "visualization of business process models" from a wider point of view. Based on this prior research we directed our work.

A systematic literature review is a process composed of a sequence of phases: planning, where the review protocol is defined; execution, where the selection of studies and data extraction are performed; and publishing, where the results of the analysis phase are reported.

Although it is not possible to avoid publication bias, the definition of a review protocol, before the collection of the candidate papers, allows reducing the probability of generating a biased result (KITCHENHAM; CHARTERS, 2007). The review protocol is composed by the research questions, the definition of the studies selection process, the search string and search sources, the exclusion and inclusion criteria (EC and IC, respectively) and, finally, how data will be extracted and synthesized. We present details of our review protocol in the following subsections.

3.1.1 Research questions

The research question (RQ) is the most important element and drive the entire systematic review. Based on the RQ, the other components of the review protocol are generated, i.e., the search string, inclusion and exclusion criteria, and data extraction strategy, so the RQ may be properly answered (KITCHENHAM; CHARTERS, 2007).

Since the main goal of our systematic review is to identify what has been published about visualization of business process models, that goal defined our primary RQ to guide the entire research process. Thus, our RQ1 has been set as follows:

- **RQ1 (primary)**: What is being investigated in the topic visualization of business process models?

To help to obtain data and to summarize different aspects of the topic being studied as well as to identify gaps in current research, we defined four secondary RQs. Those RQs allow identifying what is being investigated specifically, which aspect is missing in the current set of publications, the frequency with which the topic has been addressed in publications, and who are the main authors publishing about the subject. The secondary RQs also guided the setting of some of the exclusion and inclusion criteria as we will see in the next sections.

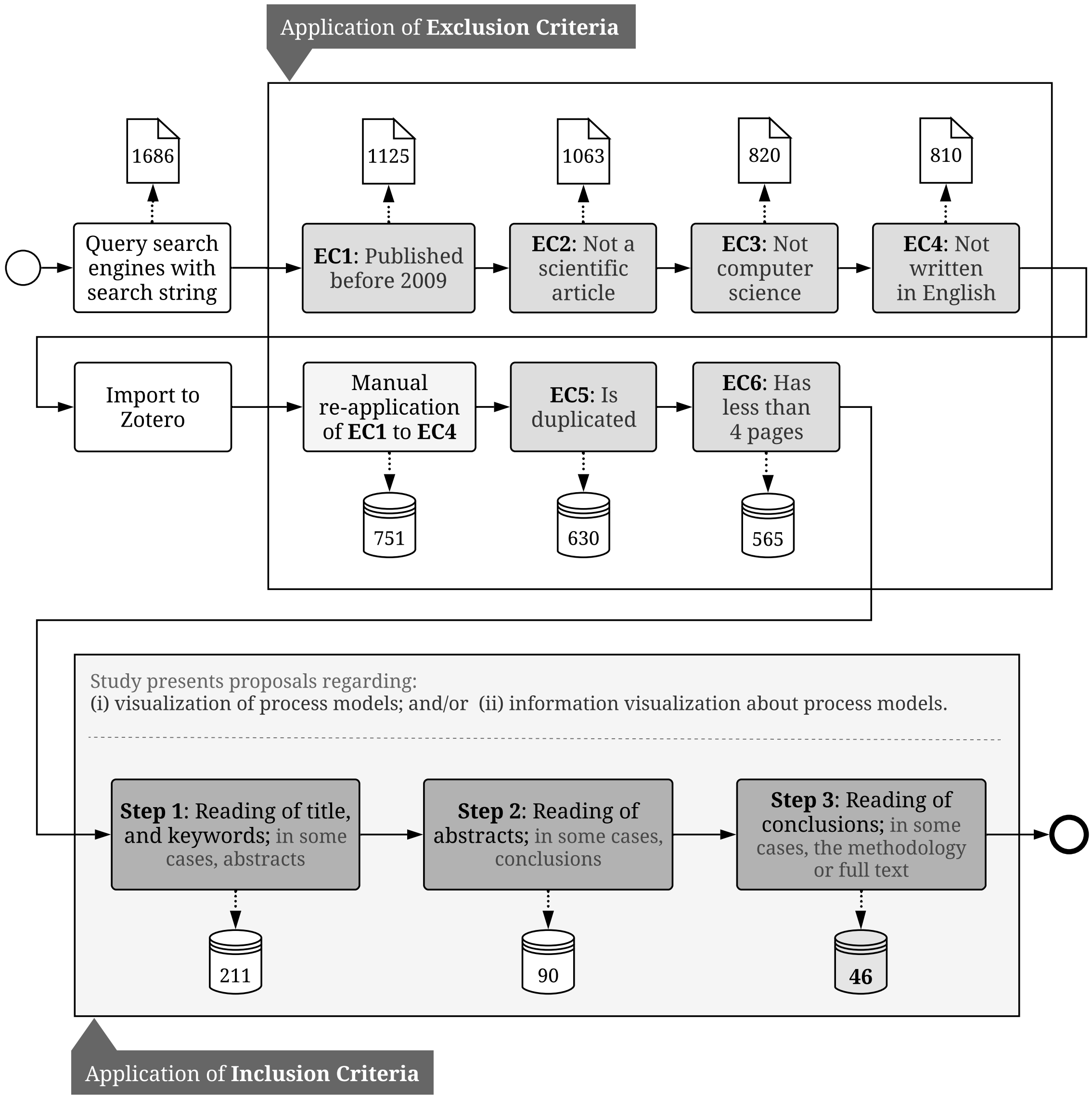
- **RQ2**: Are the studies concerned with improving the understandability of process models?
- **RQ3**: Are there open problems for further research on this topic?
- **RQ4**: How active is the research on this topic since 2009?
- **RQ5**: Who is leading research on this topic?

3.1.2 Overview of the studies selection process

The studies selection process is the most important stage in the execution of a systematic literature review (KITCHENHAM; CHARTERS, 2007) and was carried out in a set of phases. Each phase and the respective amount of selected papers can be observed in Figure 3.1.

Initially, we applied the search strings to the search sources, without using any filters. Then, filters were used within the search engines, whenever possible, to restrain search results based on EC1 to EC4. After that, papers were imported to Zotero, EC1 to EC4 were manually reapplied, and the other ECs and the ICs were applied. The application of each step of the study selection process will be explained in more detail in Subsections 3.1.3, 3.1.4, and 3.1.5.

Figure 3.1: Study selection process and the amount of articles obtained after each phase.



3.1.3 Search sources and search string

According to the York University Centre for Reviews and Dissemination (CDR) Database of Abstracts of Reviews of Effects (DARE) criteria[2], as cited by Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007), a literature search is likely to cover all relevant studies when searching is performed in 4 or more digital libraries. Then, through the analysis of other systematic literature review's in the area of BPM (Moreno-Montes De Oca et al., 2015; DIKICI; TURETKEN; DEMIRORS, 2017), and our preliminary research, we identified 5 relevant sources to be used in our literature review: ACM Digital Library[3],

[2]http://www.york.ac.uk/inst/crd/crddatabase.htm#DARE
[3]http://dl.acm.org/

IEEE Xplore[4], SpringerLink[5], Science Direct[6], and Scopus[7]. We considered including Web of Science also, but it would return a subset of the papers retrieved within the chosen digital libraries.

After choosing the search sources, based on our previous results and on the RQs alongside discussions with BPM experts holding several years of academic and professional experience in the BPM discipline, we defined the search fields and search string as follows:

- **Search fields:** Title, abstract, and keywords;
- **Search string:** (bpmn OR "process model" OR "process modeling") AND (visualization OR understandability).

We justify our search string as follows. Initially, we wanted to focus our research only on studies based on BPMN, since BPMN is an ISO Standard broadly used in industry. However, the resulting set was too limited. Thus, based on our preliminary research and discussions with BPM experts, we adopted "bpmn OR" to be part of our search string. Moreover, we identified articles dealing with understandability of process models through visualization. Therefore, to have a more inclusive result space, we chose to use "visualization OR understandability". It is worthwhile to comment that we also considered the idea of using "comprehension" as a search term. However, since in a pre-analysis phase such a term retrieved fewer papers than "understandability", we opted for using the latter. Moreover, we were focusing specifically on the understandability influenced by visualization, and therefore, we discarded articles aiming exclusively at the understandability of process models.

Due to differences among each search engine, we adapted the search string to conform with the format and limitations of each digital library. For example, to apply the search string in ACM Digital Library, we used: acmdlTitle:(+(bpmn "process model" "process modeling") +(visualization understandability)) OR recordAbstract:(+(bpmn "process model" "process modeling") +(visualization understandability)) OR keywords.author.keyword:(+(bpmn "process model" "process modeling") +(visualization understandability))"; while in Scopus, we simply used: ((bpmn OR "process model" OR "process modeling") AND (visualization OR understandability)).

To confirm that the search string was returning the desired coverage of papers, we performed an iterative process using different versions of the search string within each

[4]http://ieeexplore.ieee.org/
[5]http://link.springer.com/
[6]http://www.sciencedirect.com/
[7]http://www.scopus.com/

search source, and the top relevant papers returned were compared to the results from the top relevant papers returned by the specified search string. The papers that fell outside the intersection between the results of the variant search strings and the specified search string were analyzed based on title, abstract, and keywords (and, in some cases, the conclusions and even other sections of the text). We observed that most of them were not relevant to our systematic literature review, which made us confident about the specified search string.

3.1.4 Exclusion criteria

The exclusion criteria (EC) were used to filter the papers obtained from the search sources based on their format and publication details. Thus, the following criteria were defined:

- **EC1**: published before 2009;
- **EC2**: not a scientific article;
- **EC3**: not computer science;
- **EC4**: not written in English;
- **EC5**: duplicated;
- **EC6**: having less than 4 pages.

The year of 2009 was chosen as the starting point of this literature review because the first beta version of the BPMN 2.0 was released in that year (OMG (Object Management Group), 2009). However, intending to minimize the probability of leaving significant contributions regarding visualization of process models behind, after the application of the whole selection process, we performed a final selection step over papers published before 2009, which we describe in Section 3.1.5. Criteria EC2, EC3, and EC4 were defined to restrain the initial set of papers to those that were within the desired scope of our survey, while EC6 guaranteed a certain measure of quality.

Criteria EC1 to EC4 were applied, where possible, directly through the search engines. The results were then imported into Zotero[8], through which EC5 and EC6 criteria were manually applied. After that, also manually, we reapplied criteria EC1 to EC4, in order to prevent any unwanted articles from being among the ones selected, and to remove the ones that could not be directly removed by the search engines. The total amount of papers obtained after applying the ECs is presented in Figure 3.1.

[8]http://www.zotero.org/

The studies resulting from the application of the ECs were imported into Mendeley[9] for reading (since this program has desktop, web, and mobile versions) and preparing for the subsequent application of the inclusion criteria and data extraction.

3.1.5 Inclusion criteria

The inclusion criteria (IC) were applied for selecting the papers that had their content related to the topic of this systematic literature review. They are as follows:

- **IC1**: The paper presents proposals regarding visualization of process models;
- **IC2**: The paper presents a proposal to visualize information about process models.

The method used to apply the ICs was performing the following steps, the result of a step being the entry for the next one:

- **IC Step 1** The set of articles resulting from the application of the ECs was analyzed regarding their titles and keywords and, in some cases, their abstracts;
- **IC Step 2** The abstracts of the selected articles were analyzed and, in some cases, also the conclusions; and
- **IC Step 3** The conclusion of each article was analyzed and, in some cases, also the methodology section or, whenever necessary, the full text, to confirm the relationship between the article and the main scope of the systematic literature review.

After these steps, the paper was selected to be fully read if it fitted at least one of the ICs. The total amount of papers obtained in each step is presented in Figure 3.1. The whole selection process resulted in a set of 46 papers to go through the data extraction phase and compose the resulting set to be analyzed for producing the state-of-art report.

Nonetheless, we wanted to guarantee that no significant contribution to the visualization of process models was left behind due to having been published before 2009. Then, we performed a final selection step as follows. We applied our search string in the search engines to look for studies published before 2009. We compared the results of this search with the set of papers referenced by our 46 selected studies. The intersection of the two sets of papers resulted in 13 studies to which we applied our exclusion and inclusion criteria. As a result, only three papers emerged as possible candidates to be included in our systematic literature review (PHILIPPI; HILL, 2007; JABLONSKI; GOETZ, 2008;

---

[9] http://www.mendeley.com/

SADIQ; GOVERNATORI; NAIMIRI, 2007), which were cited by 4, 6, and 6 papers of the 46-papers set, respectively. After analyzing these three papers, we were sure that they would not contribute significantly to the results of our state-of-the-art report, and thus we decided to maintain the 46 studies that resulted from the selection process (Fig.3.1) to be submitted to the data extraction process.

Table 3.1: Data extraction form. In this table, the field "Value" contains an explanation about the expected value, whenever it is necessary.

Data item Value	**RQ**
Identifier Integer	
Title Name of the article	
Author Set of names of the authors	RQ5
Publication year Calendar year	RQ4
Item type Binary Conference/Journal	RQ4
Main scope Text	RQ1 and RQ3
Category Name of the category assigned to the paper as presented in Section 3.1.6	RQ1 and RQ3
Based on BPMN Binary Yes/No	RQ1
Evaluation or validationn Binary Yes/No, the study performed evaluation or validation of its proposal	RQ1 and RQ3
Evaluation or validation with users Binary Yes/No	RQ1 and RQ3
Raises hypothesis Binary Yes/No	RQ1 and RQ3
Significant statistically Binary Yes/No	RQ1 and RQ3
Focuses on collections of process models Binary Yes/No	RQ1 and RQ3
Understandability Binary Yes/No, the study explicitly pursue improvement in the understandability of process model	RQ2

3.1.6 Data extraction

To extract data from our final set of articles and support answering the research questions, we developed a template based on Petersen et al. (PETERSEN; VAKKALANKA;

KUZNIARZ, 2015). This template consists of a three-column table, where each line is a data extraction tuple (Data item, Value, RQ).

Table 3.1 presents the form used to extract data from each article. “Data item” is the data to be extracted; “Value” holds the result from the extraction, and “RQ” identifies the research question that motivated the need for extracting the respective “Data item”. As can be seen in Table 3.1, apart from the articles’ basic information, we recorded if the paper was based on BPMN, contained results from evaluation and/or validation, raised hypothesis, presented statistically significant results, aimed at improving understandability of process models, and dealt with collections of process models.

After reading each selected study, we built Table 3.2 as follows: whenever the article mentions a specific “Data item” with a binary “Value”, for example, the article mentions the use of “BPMN”, the corresponding cell received an “x” mark. Otherwise, it was left blank.

The same approach was used to fill the data extraction table related to the visualization analysis framework (Tables A.1 and A.2). The difference is that each column in these tables (i.e., “Data item”) represents an element of Munzner’s visualization analysis framework (MUNZNER, 2014), and all columns together are used to answer the same research questions RQ1 and RQ3.

Moreover, we observed the frequency of keywords in the papers, to identify which were the most used ones among the selected papers. To perform this task, we extracted the keywords from the selected papers and manually removed the ones considered too generic (e.g., design, software, application) or that appeared only once. Then, we combined the ones that made sense to be combined (e.g., process models with process model, visualizations with visualization, and so on). The keywords “visualization” and “process model” are the most recurrent ones, appearing 39 and 33 times, respectively.

The choice of configuration of the reviewer team for the data extraction activity was based on Kitchenham (KITCHENHAM; CHARTERS, 2007), and we considered the following aspects: the number of available reviewers, the number of selected studies during the selection process and the time available to conclude the systematic literature review. Thus, we chose to use the configuration “one reviewer and one evaluator”, which says that the reviewer is responsible for the data extraction from all studies, and the evaluator is responsible for the data extraction of a random sample of the studies. Then, the data extracted by both are confronted with the purpose of identifying divergences. Whenever necessary, the reviewer may act as the evaluator.

Table 3.2: Studies included and the main data items extracted. A cell marked with “x” indicates that the paper includes information related to the corresponding data item. Totals and percentages for each column are presented; the dark bar represents the number of papers.

Article	BPMN	Evaluation or validation	Evaluation or validation with users	Raises hypothesis	Statistically significant	Focuses on collections of process models	Understandability
(JÚNIOR et al., 2018)	x					x	
(ECKLEDER et al., 2009)							x
(REIJERS et al., 2011)		x	x	x	x		x
(KUMMER; RECKER; MENDLING, 2016)	x	x	x	x	x		x
(EMENS; VANDERFEESTEN; REIJERS, 2016)		x	x				x
(HIPP; MUTSCHLER; REICHERT, 2012)						x	x
(STORCH; LAUE; GRUHN, 2013)		x				x	
(HIPP et al., 2014)		x	x		x	x	
(IVANCHIKJ; FERME; PAUTASSO, 2015)	x					x	
(GULDEN; ATTFIELD, 2016)							
(AWAD; WESKE, 2010)	x						x
(LAUE; AWAD, 2011)	x						x
(WITT et al., 2015)							x
(CORRADINI et al., 2017)	x	x					x
(ONGGO; KARPAT, 2011)	x						x
(MUELLER-WICKOP et al., 2011)	x						x
(JOSCHKO; WIDOK; PAGE, 2013)	x						
(LEITNER et al., 2013)	x	x	x				x
(KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013)	x	x	x				x
(HIPP et al., 2015)	x	x	x		x		x

Table 3.2 continued from previous page

Article	BPMN	Evaluation or validation	Evaluation or validation with users	Raises hypothesis	Statistically significant	Focuses on collections of process models	Understandability
(KOSCHMIDER; FIGL; SCHOKNECHT, 2016)							x
(MERINO et al., 2016)	x						x
(SALNITRI; DALPIAZ; GIORGINI, 2017)	x	x	x	x			x
(EFFINGER; SPIELMANN, 2010)	x	x					x
(STROPPI; CHIOTTI; VILLARREAL, 2011)	x	x					
(KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011)		x	x	x			
(KRIGLSTEIN; RINDERLE-MA, 2012)							
(KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012)		x					
(KOLB; REICHERT, 2013)		x					x
(KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013)							
(KRIGLSTEIN; RINDERLE-MA, 2013)						x	
(FIGL; KOSCHMIDER; KRIGLSTEIN, 2013)			x		x		
(REICHERT, 2013)	x	x					x
(CORDES; VOGELGESANG; APPELRATH, 2015)							
(GALL et al., 2015)			x		x		
(PERALTA et al., 2015)	x	x					
(PINI; BROWN; WYNN, 2015)		x	x			x	
(GUO; BROWN; RASMUSSEN, 2012)							
(HOLZMÜLLER-LAUE et al., 2013)	x						
(KATHLEEN; ROSS; KRIGLSTEIN, 2014)		x					
(JOŠT et al., 2017)	x	x	x	x	x		x
(POLDERDIJK et al., 2018)	x	x	x				

Table 3.2 continued from previous page

Article	BPMN	Evaluation or validation	Evaluation or validation with users	Raises hypothesis	Statistically significant	Focuses on collections of process models	Understandability
(CORRADINI et al., 2018)	x						
(KRENN; KEPLER, 2018)		x	x	x			x
(CABALLERO et al., 2018)		x					
(OBERHAUSER; POGOLSKI; MATIC, 2018)	x	x	x				x
Total	23	24	16	6	7	7	23
%	50.00	52.17	34.78	13.04	15.22	15.22	50.00

3.2 Studies Classification

We classified the 46 selected studies into six categories, after observing the similarities among the main scope of the proposals they present regarding the visual representation of process models. With this categorization, we aimed at answering “What” the studies are reporting regarding visualization of business process models. The main scope is one of the data items extracted from each article, as seen in Table 3.1. The distribution of the studies per category is presented in Table 3.3, and the following sections detail each one of the defined categories.

3.2.1 Augmentation of existing process modeling language elements

This category includes 71.74% of the selected studies. They propose various ways to improve elements of a process modeling language, by augmenting their semantics. Many

Table 3.3: Detailed distribution of articles per category over the 46 studies selected to be fully read.

Category Articles	**Total** (%)
Augmentation of existing process modeling language elements (ECKLEDER et al., 2009; REIJERS et al., 2011; KUMMER; RECKER; MENDLING, 2016; EMENS; VANDERFEESTEN; REIJERS, 2016; GULDEN; ATTFIELD, 2016; AWAD; WESKE, 2010; LAUE; AWAD, 2011; WITT et al., 2015; CORRADINI et al., 2017; MUELLER-WICKOP et al., 2011; LEITNER et al., 2013; KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013; HIPP et al., 2015; KOSCHMIDER; FIGL; SCHOKNECHT, 2016; SALNITRI; DALPIAZ; GIORGINI, 2017; STROPPI; CHIOTTI; VILLARREAL, 2011; KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011; KRIGLSTEIN; RINDERLE-MA, 2012; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; KOLB; REICHERT, 2013; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; FIGL; KOSCHMIDER; KRIGLSTEIN, 2013; REICHERT, 2013; CORDES; VOGELGESANG; APPELRATH, 2015; GALL et al., 2015; PERALTA et al., 2015; PINI; BROWN; WYNN, 2015; HOLZMÜLLER-LAUE et al., 2013; KATHLEEN; ROSS; KRIGLSTEIN, 2014; JOŠT et al., 2017; POLDERDIJK et al., 2018; CORRADINI et al., 2018; CABALLERO et al., 2018)	33 (71.74%)
Creation of new process modeling language elements (ONGGO; KARPAT, 2011; JOSCHKO; WIDOK; PAGE, 2013; HIPP et al., 2015; MERINO et al., 2016)	4 (8.70%)
Exploration of the 3D space for process modeling (HIPP et al., 2015; EFFINGER; SPIELMANN, 2010; GUO; BROWN; RASMUSSEN, 2012; OBERHAUSER; POGOLSKI; MATIC, 2018)	4 (8.70%)
Information visualization about process models (JÚNIOR et al., 2018; EMENS; VANDERFEESTEN; REIJERS, 2016; HIPP; MUTSCHLER; REICHERT, 2012; STORCH; LAUE; GRUHN, 2013; HIPP et al., 2014; IVANCHIKJ; FERME; PAUTASSO, 2015; GULDEN; ATTFIELD, 2016; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; KRIGLSTEIN; RINDERLE-MA, 2013; REICHERT, 2013; PERALTA et al., 2015; PINI; BROWN; WYNN, 2015; HOLZMÜLLER-LAUE et al., 2013; CORRADINI et al., 2018)	15 (32.61%)
Visual feedback concerning problems detected in process models (AWAD; WESKE, 2010; LAUE; AWAD, 2011; WITT et al., 2015; CORRADINI et al., 2017)	4 (8.70%)
Support for different perspectives of a process model (EMENS; VANDERFEESTEN; REIJERS, 2016; HIPP; MUTSCHLER; REICHERT, 2012; HIPP et al., 2014; GULDEN; ATTFIELD, 2016; KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013; EFFINGER; SPIELMANN, 2010; STROPPI; CHIOTTI; VILLARREAL, 2011; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; KOLB; REICHERT, 2013; REICHERT, 2013; PERALTA et al., 2015; PINI; BROWN; WYNN, 2015; GUO; BROWN; RASMUSSEN, 2012; KRENN; KEPLER, 2018)	14 (30.43%)

studies explore highlighting of elements through the use of different colors or transparency of certain fragments of the process model to enable users to comprehend the model. Some authors (KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011; KRIGLSTEIN; RINDERLE-MA, 2012; CORDES; VOGELGESANG; APPELRATH, 2015; CORRADINI et al., 2018; CABALLERO et al., 2018) propose the coloring of modeling language elements and its control flows to highlight changes and facilitate the identification of differences and similarities in business process models, while other ones (ECKLEDER et al., 2009; REIJERS et al., 2011) highlight matching operators (i.e., split ANDs with their respective join ANDs). Emens et al. (EMENS; VANDERFEESTEN; REIJERS, 2016) and Jošt et al. (JOŠT et al., 2017) propose making transparent the portions of a process model that are not reachable from the activity being executed. Another interesting study framed in this category is the work by Kriglstein et al. (KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013): among other propositions, they increase the thickness of the control flow lines to improve the perception of the process model's paths that are being most executed.

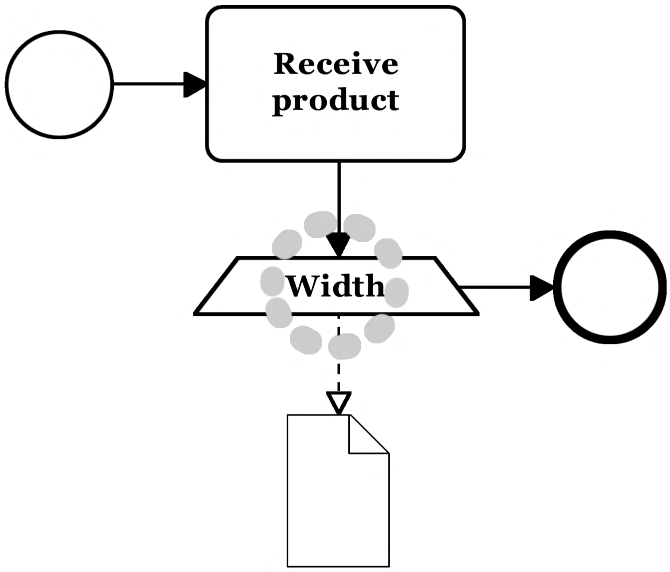
Figure 3.2: Example of annotation on modeling language elements, based on the proposal of Leitner et al. (LEITNER et al., 2013): the augmentation of the process model element is made through the superposition of an icon (indicated by the dotted circle) to an activity and, in this case, represents that the augmented activity involves some kind of access permission.

Other studies (MUELLER-WICKOP et al., 2011; LEITNER et al., 2013; KATHLEEN; ROSS; KRIGLSTEIN, 2014; SALNITRI; DALPIAZ; GIORGINI, 2017; POLDERDIJK et al., 2018) explore the use of different elements such as icons, text or images to improve the way a process model element represents its information or to represent domain-specific aspects. For example, Figure 3.2 shows the use of a key locker icon to represent that an activity demands some level of permission to be executed in the security domain (LEITNER et al., 2013). Salnitri et al. (SALNITRI; DALPIAZ; GIORGINI, 2017) also attach icons to existing process model's elements to represent security aspects. Mueller-Wickop et al. (MUELLER-WICKOP et al., 2011) attach textual information to represent financial auditing aspects, while Kathleen et al. (KATHLEEN; ROSS; KRIGLSTEIN, 2014) use images attached to activities of a process model, aiming at improving its expressiveness by picturing what each activity's task is.

3.2.2 Creation of new process modeling language elements

This category aggregates 8.70% of the selected papers, which propose different approaches to extend a process modeling language by means of adding new types of elements, with different behavior than the ones already existing in the current languages to which the new elements are being proposed. For example, Joschko et al. (JOSCHKO; WIDOK; PAGE, 2013) propose a wind farm signal event that sends specific malfunctioning signals, and a gateway capable of influencing the path through which a process model would execute, according to weather information received from a proprietary provider.

Figure 3.3: Example of new process modeling language element, based on the proposal of Merino et al. (MERINO et al., 2016): the new process modeling language element (indicated by the dotted circle) measures the width of a received product and stores this information within an object data.

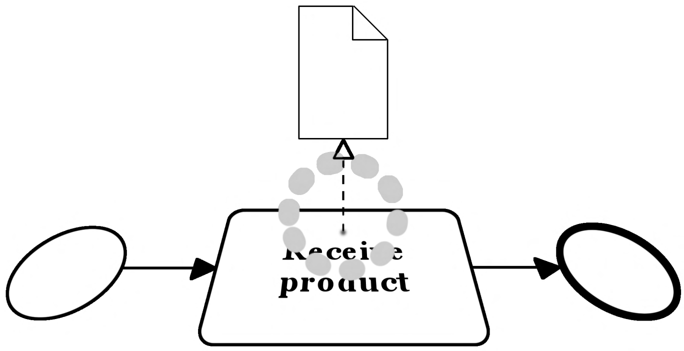


More recently, Merino et al. (MERINO et al., 2016) propose an element that is capable of measuring characteristics of a product whenever it is necessary within a process model workflow (Figure 3.3). They also propose another element capable of detecting if quality aspects about that product and its measurements are being met in any other desired process model's workflow point.

3.2.3 Exploration of the 3D space for process modeling

Usually business process models are two-dimensional (2D) representations, so modeling tasks take place in a 2D plane. However, a small number of studies (4 out of 46 papers) propose the use of a higher dimensional space to draw the process model (Figure 3.4). Hipp et al. (HIPP et al., 2015) introduced the BPMN3D visualization concept, where the object data is represented on a plane in 3D, while the rest of the process model diagram is drawn on the 2D plane. To enhance the communication among stakeholders and business analysts, Guo et al. (GUO; BROWN; RASMUSSEN, 2012) proposed a 3D simulation representing the behavior of what a process model's activity task should be (e.g., a process model's activity task named "Receive product" would be represented by a 3D simulation where a product is being received).

Figure 3.4: Example of a process model represented in 3D space (HIPP et al., 2015): the object data is draw into the third dimension (highlighted through a dotted circle) while the rest of the process model diagram is represented on the 2D plane.



In Effinger and Spielmann (EFFINGER; SPIELMANN, 2010) proposal, the process model lanes (generally representing users' roles within an organization) are represented in different layers spread one above the other across the 3D space. Each layer holds the process model lanes' elements which it represents, in a 2D plane. The control flow across the process models' elements is represented in 3D, just as each lane. In this approach, when the process model is seeing from above, it seems as it was modeled in a 2D plane. But when the process model is seen from another point of view in the 3D space, it is depicted from a resources' perspective, where it is possible to see the activities and events belonging to each process model lanes separately.

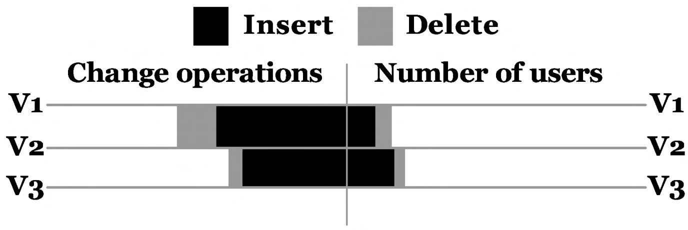
A very recent work by Oberhauser et al. (OBERHAUSER; POGOLSKI; MATIC, 2018) proposed a solution for representing BPMN models in virtual reality, including navigation, interaction and annotation features. The authors report findings from an empirical study for evaluating effectiveness, efficiency, and intuitiveness of the 3D representation compared to other model representations.

3.2.4 Information visualization about process models

This category corresponds to 32.61% of the selected papers, and is composed by studies proposing different approaches to represent information about a process model or a collection of process models. For example, Figure 3.5 shows how Kriglstein and

Rinderle-Ma (KRIGLSTEIN; RINDERLE-MA, 2013) use a horizontal stacked bar chart to display, at the same time, the total number of operations on a process model (e.g., adding or removing a process model activity) for each process model's version, and the number of users that performed these operations. The central, reference line facilitates comparison of variables among versions.

Figure 3.5: Example of information visualization technique ((KRIGLSTEIN; RINDERLE-MA, 2013)): a horizontal stacked bar chart, where each bar represents a process model's version, with the left side showing the number of insertions and deletions of elements for building that version, while the right side presents the number of users that performed the change operations.



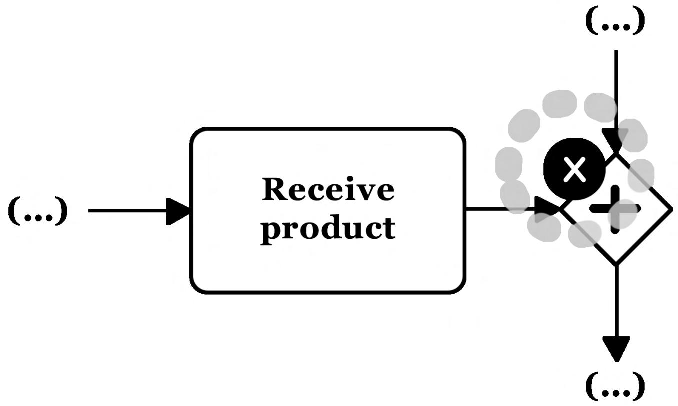
Pini et al. (PINI; BROWN; WYNN, 2015) propose different approaches to present information about a process model execution log. All their approaches suggest the use of graphics right above (or below) the activities in the process model, so data about the execution of each activity can be seen right within the process model itself. One of the proposals still explores the use of a horizontal stacked bar. But, differently from Kriglstein and Rinderle-Ma (KRIGLSTEIN; RINDERLE-MA, 2013), they use only one horizontal stacked bar per activity, which is divided into three parts, each one representing a shift of the day (i.e., morning, afternoon, night). The size of the parts may vary according to the whole amount of time that the activity took to execute in each shift of the day.

3.2.5 Visual feedback concerning problems detected in process models

Although feedback about problems in process models are important during modeling tasks, only 8.70% of the selected studies propose some kind of graphical representation for improving the perception of issues. We found studies employing different approaches to present feedback to the modeler about any type of problem within a given process model or a collection of process models. Figure 3.6 shows a proposal by Laue and Awad (LAUE; AWAD, 2011), in which they attach a graphic symbol, a white "x" surrounded by a red circle (indicated by the dotted circle in the figure), to the process model's element that generated the process modeling problem. Besides that, a textual description about the problem is triggered by a mouse over action.

The process model's elements that are part of the detected process modeling problems are highlighted in red in other proposals (AWAD; WESKE, 2010; CORRADINI et al., 2017). Moreover, Corradini et al. (CORRADINI et al., 2017) present a list of guidelines infringed in a process model and allows the user to select each of them separately to see the process model's fragment that is related to the guideline's infringement highlighted in red.

Figure 3.6: Example of visual feedback (LAUE; AWAD, 2011): the visual feedback is marked through a white "x" surrounded by a black circle (indicated by the dotted circle) attached to the element that caused the process modeling problem which, in this case, could be the use of the wrong join gateway.



3.2.6 Support for different perspectives of a process model

Studies proposing different approaches to distinguish parts of a process model to different type of users and/or situations compose this category. Reichert (REICHERT, 2013) proposes different visualizations of process models depending on users' roles. He aggregates process model's elements to give an overview of the process model diagram to managers, and provides filters based on user role to allow a specific business process participant to view only elements corresponding to his/her process model's activities.

In another study (EMENS; VANDERFEESTEN; REIJERS, 2016), a control flow perspective is presented by making less opaque the process model's activities not reachable from the one currently being executed. Other two studies (PERALTA et al., 2015; PINI; BROWN; WYNN, 2015) propose a time perspective, so the user is provided with visual features to observe the processing time of each process model's activities.

3.3 Visualization Analysis

To analyze the studies from an information visualization point of view, we adopted Munzner's visualization analysis framework (MUNZNER, 2014) (refer to section 3.3). We decided to base our visualization analysis on Munzner‘s framework once it is well-accepted in the visualization community due to it provides for a high-level abstract view of the visualizations while covering all aspects that might be involved.

To answer the questions "Why the users use the visualization" and "How the visualizations are encoded", we extracted from the studies the data presented in Tables A.1 and A.2. The data items extracted correspond to *actions* and *targets* for the "Why?" aspect, and *design choices* for the "How?". As for the aspect "What?" of the framework, the majority of the studies has graphs representing the process models as their main dataset (i.e., network data according to the framework). Therefore, for space-saving purposes, we did not add this information to the visualization-related data table. Moreover, we extracted information to know which studies allow user interaction in their proposals, resulting in 65.12% of the papers (JÚNIOR et al., 2018; ECKLEDER et al., 2009; REIJERS et al., 2011; EMENS; VANDERFEESTEN; REIJERS, 2016; HIPP; MUTSCHLER; REICHERT, 2012; STORCH; LAUE; GRUHN, 2013; HIPP et al., 2014; IVANCHIKJ; FERME; PAUTASSO, 2015; GULDEN; ATTFIELD, 2016; LAUE; AWAD, 2011; WITT et al., 2015; CORRADINI et al., 2017; ONGGO; KARPAT, 2011; KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013; EFFINGER; SPIELMANN, 2010; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; KOLB; REICHERT, 2013; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; REICHERT, 2013; PERALTA et al., 2015; PINI; BROWN; WYNN, 2015; HOLZMÜLLER-LAUE et al., 2013; JOŠT et al., 2017; POLDERDIJK et al., 2018; CORRADINI et al., 2018; KRENN; KEPLER, 2018; CABALLERO et al., 2018; OBERHAUSER; POGOLSKI; MATIC, 2018).

After extracting the data for this analysis, we wanted to group the studies to get a general and abstract view of their proposals. To avoid missing any possible existing grouping or relation among studies, we decided to use the *k*-means method for clustering them based on the extracted data. Using RapidMiner Studio[10], 8.2 version, which is a data science platform that provides an integrated environment to data analysis and visualization, we experimented *k*-means with different parameters: k = [2, .., 7], max runs = [10, 300, 3000] and maximum optimization steps = [10, 100, 1000, 10000]. The analysis of the

[10]https://www.rapidminer.com/

outcomes showed that the most meaningful results were obtained using $k = 6$, max runs = 300 and maximum optimization steps = 1000. Some data extraction elements that were present in only one study, such as *Annotate*, *Record* and *Order*, were suppressed from the data extraction table, so they would not generate bias within clusters. The distribution of studies in each cluster generated by the application of the *k*-means technique is presented in Table 3.4, and two heatmaps of visualization-related data items per cluster are shown in Figures 3.7 and 3.8. Figure 3.8 was derived from the heatmap generated from RapidMiner. Figure 3.7 was built based on the incidence of the frameworks' elements in each cluster, but instead of representing the low level aspects of the framework, we grouped the elements as the framework does. We refer to "Consume" instead of "Present/Discover", "Produce" instead of "Annotate/Record", and so on (refer to Sections 2.1.2.2 and 2.1.2.3).

Table 3.4: Detailed distribution of the 43 studies per cluster after being processed using the visualization analysis framework (MUNZNER, 2014). The clustering was based on data presented in Tables A.1 and A.2.

Cluster Articles	**Total** (%)
Cluster 0 (EMENS; VANDERFEESTEN; REIJERS, 2016; WITT et al., 2015; PERALTA et al., 2015; HOLZMÜLLER-LAUE et al., 2013; JOŠT et al., 2017; CORRADINI et al., 2018; OBERHAUSER; POGOLSKI; MATIC, 2018)	7 (16.28%)
Cluster 1 (HIPP; MUTSCHLER; REICHERT, 2012; CORRADINI et al., 2017; ONGGO; KARPAT, 2011; KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013; EFFINGER; SPIELMANN, 2010; KOLB; REICHERT, 2013; REICHERT, 2013; POLDERDIJK et al., 2018; KRENN; KEPLER, 2018)	9 (20.93%)
Cluster 2 (JÚNIOR et al., 2018; STORCH; LAUE; GRUHN, 2013; IVANCHIKJ; FERME; PAUTASSO, 2015; KRIGLSTEIN; RINDERLE-MA, 2013)	4 (9.30%)
Cluster 3 (HIPP et al., 2014; GULDEN; ATTFIELD, 2016; PINI; BROWN; WYNN, 2015; CABALLERO et al., 2018)	4 (9.30%)
Cluster 4 (ECKLEDER et al., 2009; REIJERS et al., 2011; KUMMER; RECKER; MENDLING, 2016; AWAD; WESKE, 2010; LAUE; AWAD, 2011; MUELLER-WICKOP et al., 2011; JOSCHKO; WIDOK; PAGE, 2013; HIPP et al., 2015; MERINO et al., 2016; SALNITRI; DALPIAZ; GIORGINI, 2017; STROPPI; CHIOTTI; VILLARREAL, 2011; GALL et al., 2015; GUO; BROWN; RASMUSSEN, 2012; KATHLEEN; ROSS; KRIGLSTEIN, 2014)	14 (32.56%)
Cluster 5 (KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011; KRIGLSTEIN; RINDERLE-MA, 2012; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; CORDES; VOGELGESANG; APPELRATH, 2015)	5 (11.63%)

Although we had 46 selected studies, only 43 of them passed through the visualization analysis process, since 3 articles (LEITNER et al., 2013; KOSCHMIDER; FIGL;

Figure 3.7: Representation of the incidence of the data items per cluster of papers. For simplification of the heatmap, data items were grouped according to the framework. Values close to 1 represent high incidence of the respective data item in the cluster, while the ones close to 0 represent low incidence.

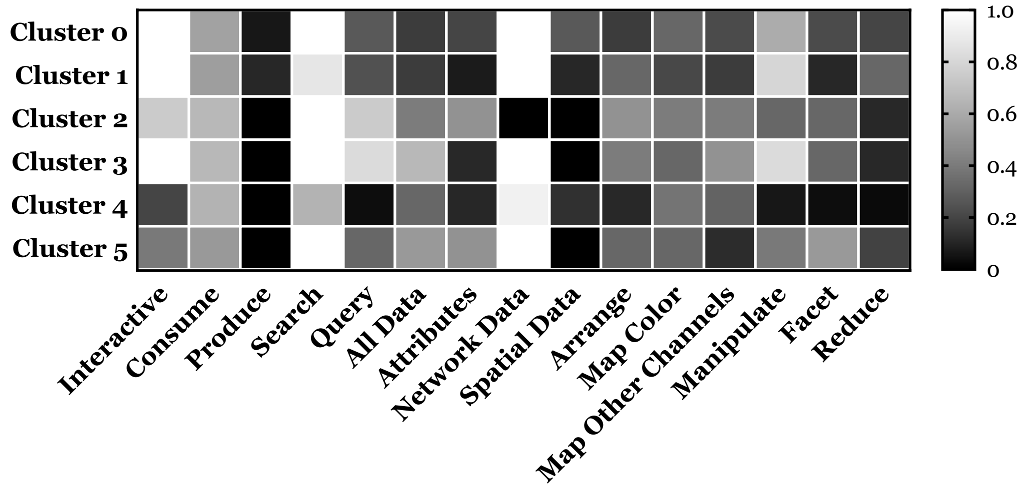
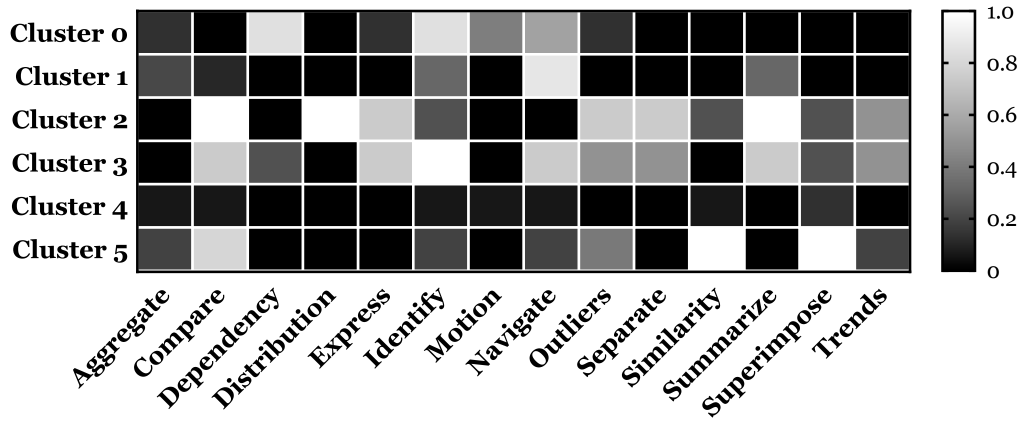


Figure 3.8: Representation of the incidence of the most representative data items characterizing *Why?* and *How?* per cluster for visualization analysis purposes. Values close to 1 represent high incidence of the respective data item in the cluster, while the ones close to 0 represent low incidence.



SCHOKNECHT, 2016; FIGL; KOSCHMIDER; KRIGLSTEIN, 2013) were mainly results from other systematic literature reviews. The difference between those reviews and ours is the scope. Our work provides an overview of how information visualization techniques have been used in business process models, while theirs focused on a specific category we have already identified as the main scope of some papers: augmentation of existing process modeling language elements. Figl et al. (FIGL; KOSCHMIDER; KRIGLSTEIN, 2013) investigate different visualization strategies for the arrangement of nodes and links in the process model diagram, while Leitner et al. (LEITNER et al., 2013) address the use of symbols attached to process model elements (in the security domain), and Koschmider et al. (KOSCHMIDER; FIGL; SCHOKNECHT, 2016) review the literature to provide an overview of the design of labels for process model element.

In the following subsections, we discuss the clustering results and describe the main characteristics of each cluster as well as the studies they grouped.

Cluster 0: characterized by “manipulate” and “query” tasks

Approaches proposed in the studies composing this cluster do not target “all data”, which means that they do not intend to support users in discovering *trends*, *outliers* or *features* within the whole dataset. As cluster 5, it is in the third position of clusters composed of papers that mostly explore “query” tasks. Articles in this cluster also place it among the three clusters to explore “manipulate” tasks the most (Figure 3.7). However, when the analysis goes to the level of the framework elements themselves, i.e., which elements are targets of the tasks, which interaction techniques are used and how feedback is provided, we notice finding *dependency* between attributes, *identify* elements with some characteristics, *navigate* in the diagram, *aggregate* elements, and use *motion* as visual representation of some feature in the diagram (Figure 3.8).

To exemplify, Peralta et al. (PERALTA et al., 2015) propose an approach which assigns, to each BPMN element, information such as time or resources needed by the element. Such information is represented through *shape* transformation of each activity according to each information desired to communicate (e.g., processing time may be represented by transforming the width from thinner to wider, as the time needed by the activity to be executed increases). This proposal allows the user to gain insights from the process, *identify outliers* and *features*.

Other studies (EMENS; VANDERFEESTEN; REIJERS, 2016; JOŠT et al., 2017) use similar ways to represent different information. They use transparency through *luminance* and *saturation*, respectively, to represent process models’ paths that are not reachable from the current *selected* activity (in the case of the latter study) or the current one being executed (in case of the former study). Both proposals allows *navigating* the process model diagram and *identifying dependencies* of the current activity, which is also present in (OBERHAUSER; POGOLSKI; MATIC, 2018). While the research by Jost et al. (JOŠT et al., 2017) indicates that their approach seems to increase the cognitive effectiveness of business process models, the prototype implemented by Emens et al. (EMENS; VANDERFEESTEN; REIJERS, 2016) was evaluated by users participating in the process, and the authors concluded that their proposal had preference over static visualizations. The main difference between both studies is that Emens et al. (EMENS; VANDERFEESTEN; REIJERS, 2016) use highlighting of parts of a process, either by blocks referring to a certain activity in focus (being currently executed) or by the role of the user (process participant) who is visualizing the process. Moreover, their study explores *motion* of the token indicating the current activity being executed.

Holzmuller-Laue et al. (HOLZMÜLLER-LAUE et al., 2013) present an interface called BPESi, composed of 3 visualization areas. The first visualization area is used for displaying the process model itself. The second employs *motion* to animate the process model execution token, which is similarly present in (CORRADINI et al., 2018). The third is used to show information regarding the activity being currently executed. This study is similar to the one by Emens et al. (EMENS; VANDERFEESTEN; REIJERS, 2016) regarding displaying the process model's execution token animation and allowing users to *identify features* of the current activity.

Finally, the proposal by Witt et al. (WITT et al., 2015) is similar to the studies by Emens et al. and Jost et al. (EMENS; VANDERFEESTEN; REIJERS, 2016; JOŠT et al., 2017) considering the use of transparency to support the *identification* of *dependencies* of a certain part of the process model. The main difference is that Witt et al.'s motivation is related to rules being violated by a certain activity of the process model. This approach provides the user with an image of the rule pattern that was found as being violated in the process model under analysis, thus allowing the user to *identify features* about the process.

Cluster 1: characterized by "manipulate" and "produce" tasks

This cluster is the one with most papers exploring "manipulate" tasks (i.e., *change* and *select* elements, and *navigate*) and "produce" information (ONGGO; KARPAT, 2011; EFFINGER; SPIELMANN, 2010) (see Figure 3.8). The use of *aggregate*, *navigate* and *summarize* is also evident among the papers as represented by the incidence index shown in Figure 3.8. For example, Polderdijk et al. (POLDERDIJK et al., 2018) report a solution that enables users to manipulate the process model defining risk characteristics to activities, so the reader may identify which tasks are safer to be performed.

Cluster 1 contains two studies that support actions to "produce" information (ONGGO; KARPAT, 2011; EFFINGER; SPIELMANN, 2010). One of them (EFFINGER; SPIELMANN, 2010) describes a tool to visualize, in different planes, the activities of different actors of a process model, allowing the visualization of the process model from the perspective of the activities of the process participants and, therefore, providing a layered view. This proposal allows the user to *change* the point of view from which he or she is viewing the process model in the three-dimensional space, and taking “snapshots” (i.e., *record* it) so, after a few “snapshots”, the user can pass through the different point of views easily, avoiding to *navigate* again from one point of view to another. The second study (ONGGO; KARPAT, 2011) proposes and implements a tool that enables the use of BPMN to represent agent-based simulation conceptual models. This tool provides modelers with *annotate features* by attaching a *shape* containing extra information about the process model activity to what it is attached, extending somehow to process model’s elements. In practice, according to the authors, this proposal can be exploited as a communication tool between simulation modelers and business users.

Other studies in this cluster (HIPP; MUTSCHLER; REICHERT, 2012; KOLB; REICHERT, 2013; REICHERT, 2013) are similar because they present information related to users’ roles and their approaches allow users to *navigate* the process model diagram. Hipp et al. (HIPP; MUTSCHLER; REICHERT, 2012) present a new concept for navigation where the main focus is to display information about collections of process models by manipulating *filters*. Their proposal allows the user to *change* the point of view from what he or she wants to inspect the process model. One point of view is to show all portions of a process model that belongs to a certain user role. Another one is a time-based view that *summarizes features* of the process model’s data (e.g., showing a wider *shape* depending on the gap between the start and end dates of the process model’s lifetime). A third one is a logic-based view, which displays the process model diagram itself. Kolb and Reichert (KOLB; REICHERT, 2013) propose and implement a proof-of-concept prototype framework, which adapts process models to each user’s perspective by *filtering* and *aggregating* elements that are not important to the role of a specific user. Managers generally need an overview of the process model, while process participants need a more detailed view, especially of the activities in which they are engaged. Finally, the main focus of the proposal presented by Reichert (REICHERT, 2013) is to view process models according to different user roles through *filtering* and *aggregating* process models’ elements. In the same sense of visualizing process models from different point of views, the study presented by Krenn (KRENN; KEPLER, 2018) enables users to visualize different aspects of process models such as interaction diagrams and function-oriented visualization.

Koschmider et al. (KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013) investigate different approaches to visually align objects and roles from the organizational context to the activities of a process model. It presents a *juxtaposed* multiple-view visualization approach combining linking and brushing techniques. In other words, when the user *selects* an element in one view, the other views are *changed* accordingly, i.e., selecting an activity from a process model displays information about the objects and roles related to that activity in another view. Statistics from their study indicated that the users easily understood the information displayed in the multiple views coordinated using linking and brushing.

Corradini et al. (CORRADINI et al., 2017) propose a tool to visualize which modeling guidelines were not satisfied by a given model. The tool does not correct any non-followed guidelines, only displays text indicating the guidelines being violated and, in some cases, the activity that generated the guideline violation is highlighted in the red. Moreover, the visualization *summarizes* all the guidelines being infringed in a list view of guidelines that shows in green the name of the guidelines that are followed and, in red, the ones infringed. The user can *change* the view to the different portions of the process model highlighted according to each one of the infringed guidelines, if there is more than one. This way the user can observe the *correlation* between the infringed guideline and the part of the process model which generated that visual feedback.

Cluster 2: characterized by “attribute targets”, and “arrange” and “query” tasks

In this cluster, most articles present features characterized as “query” elements (*compare*, *summarize*) with “attribute targets” (*distribution*, *correlation*). Moreover, it is the second cluster that has more elements classified as “arrange” (*express*, *separate*, as can be seen in Figure 3.7). Among these papers, it is more evident the presence of features like *distribution*, followed by *separate*, *express* and *compare*, targeting *summarize*, *outliers* and *trends* (Figure 3.8). All studies in this cluster explore information visualization for displaying data about collections of process models, mainly using bar charts and stacked bar charts. Three studies (JÚNIOR et al., 2018; STORCH; LAUE; GRUHN, 2013; IVANCHIKJ; FERME; PAUTASSO, 2015) propose the visualization of quantitative information of a set of process models (e.g., total number of activities within the collection), while another one (KRIGLSTEIN; RINDERLE-MA, 2013) proposes visualizing information about different versions of the same process model. Kriglstein and Rinderle-Ma (KRIGLSTEIN; RINDERLE-MA, 2013) suggest a visualization concept to compare differences between distinct versions of process models. This visualization approach is not about the process model itself, but about the characteristics of the model. It is based on a chart composed by multiple lines of stacked bar charts, where each line represents one version of the process model data, and each stack represents one type of operation that was performed in that version of the model. With that chart, the user can *discover* which version of a process model was subjected to more or fewer operations, i.e., which process version had more elements inserted or deleted. Also, the user can *search* for *correlations* among the versions of the process model. Another task that may be performed is to *compare* the displayed information between different versions of the model to *discover trends* about the operations, e.g., after which version of a process model the operations start to decline.

Two studies (JÚNIOR et al., 2018; STORCH; LAUE; GRUHN, 2013) propose different visualizations to represent information about the quality of a process model or a collection of process models. In the study presented by Storch et al. (STORCH; LAUE; GRUHN, 2013), one of the proposed visualizations is a *superimposed* bar chart that allows users to view and *compare* the number of violations per the total of elements of process models created by each type of user, for example, students or scientists. Another task that may be performed by the users is to *search* for *correlations*, *trends* and *outliers* among this data. For example, the user is capable of *changing* the chart by choosing to view only the data about models created by one or another type of user. In a recent study (JÚNIOR et al., 2018), the authors propose an interface that enables users to identify, out of a collection of process models, which are the models that do not follow a set of process modeling guidelines well known in the literature (MENDLING; REIJERS; Van Der Aalst, 2009).

Ivanchikj et al. (IVANCHIKJ; FERME; PAUTASSO, 2015) present a tool that displays different information (around 100 different metrics) about processes modeled using BPMN. This information, e.g., the total number of XOR-Split used in a collection of process models, is presented in bar charts format, supporting users to *compare* this data and finding *correlations* and *outliers*. The user can still *change* charts displaying different information about the collection of process models.

Cluster 3: characterized by “query” and “manipulate” tasks, targeting “all data”, and“map” and “arrange” design choices

This cluster stands out by being the one that mostly explore “query” elements, i.e., *identify*, *compare*, *summarize*, targeting “all data”, i.e., targeting finding *trends*, *outliers* and *features*, by means of "arrange" design choices, mainly *express* and *separate*. It is also the second cluster with more proposals related to "manipulate" elements (Figure 3.7), being more evident the use of *express* and *separate* followed by *summarize*, *outliers*, *trends*, *compare* and *identify* (Figure 3.8). Three out of four papers in this cluster (HIPP et al., 2014; GULDEN; ATTFIELD, 2016; PINI; BROWN; WYNN, 2015) present some kind of chart (e.g., bar chart, stacked bar) to represent information about a collection of process models or process models' execution logs. These charts *summarize* the datasets, which are the basis for generating the views. Cluster 3 differs from cluster 2 because it displays the process models' diagram alongside the corresponding graphics, and supports users in *identifying* process models' elements characteristics.

Two studies (HIPP et al., 2014; CABALLERO et al., 2018) present a navigation concept composed by different *juxtaposed* views. The proposal by Hipp et al. (HIPP et al., 2014) allows the user to *navigate* collections of process models (or process models per se) to view related process information. The user can *identify* characteristics of *selected* elements of interest. Moreover, using a time-based view, the user can *discover* each task execution period as well as *compare* execution periods within different tasks to *identify outliers*. In the work by Caballero et al. (CABALLERO et al., 2018), the focus is to enable users to *identify* results regarding the validation of the soundness of a process model.

Gulden and Attfield (GULDEN; ATTFIELD, 2016) provide an approach to visualize information about logs of business process models' execution. The main focus of their approach is to support users in discovering causal-temporal information related to the process or fragments of interest of the process model. Concerning the *identify/select/discover/compare* tasks, this study (GULDEN; ATTFIELD, 2016) differs from the one by Hipp et al. (HIPP et al., 2014) in the granularity of what is being compared. While Hipp et al. target the elements of user's interest, Gulden and Attfield's target are the process model's fragments of interest. These elements of interest are pointed out by the user through the *selection* of an activity, which triggers a *filtering* mechanism that provides the data related to the portion that comprehends the selected activity and its subsequent activities. The multiple-view based prototype *juxtaposes* three mini-charts horizontally in the first row, which initially *summarizes* the process model's execution data being analyzed and displayed in a fourth view that occupies the second row. When the user selects any activity, the mini-charts are updated to display information related to the respective model's fragment of user's interest.

The study by Pini et al. (PINI; BROWN; WYNN, 2015) aims to improve the way data is displayed aiming user's comparisons tasks within the process mining domain. The user is capable, for example, in one of the proposed plots, to compare behaviors of each activity over time. One of the plots that are drawn over each process model's activity presents execution data related to that activity along the day. It is composed by a bar chart that comprises *superimposed* triangular *shapes*, through which it is possible to see, for each activity, in what shift of the day that activity is executed more often. This way, users can *identify trends* and *outliers*, for example. Further, these charts use *color* to differentiate the shifts of the day.

Cluster 4: characterized by exploring design choices classified as "map" but not user interaction

This cluster is noticeably composed by studies that, in majority, do not explore user interaction. Only 3 out of 14 studies propose interactive visualizations of process models as their main scope (ECKLEDER et al., 2009; REIJERS et al., 2011; LAUE; AWAD, 2011). Also, few studies explore "query", "arrange", "manipulate" and "reduce" features, which makes sense, since this cluster comprises the papers that does not include user interaction. Moreover, it is the second cluster to less explore "attribute targets". However, as for design choices, it is one that most explores "mapping" elements, mostly *color* and *shapes*. Some studies have in common proposals to augment process models' elements through map encoding (KUMMER; RECKER; MENDLING, 2016; AWAD; WESKE, 2010; LAUE; AWAD, 2011; MUELLER-WICKOP et al., 2011; GALL et al., 2015; KATHLEEN; ROSS; KRIGLSTEIN, 2014).

Regarding "map" as design choice, a number of studies explore only color mapping (ECKLEDER et al., 2009; REIJERS et al., 2011; KUMMER; RECKER; MENDLING, 2016; AWAD; WESKE, 2010; MUELLER-WICKOP et al., 2011). For example, Eckleder et al. (ECKLEDER et al., 2009) and Reijers et al. (REIJERS et al., 2011) propose the coloring of split gateways and their respective joins. Both studies are from the same group: in the first one (ECKLEDER et al., 2009), the authors implemented an algorithm to perform the mapping and argued that one limitation of their approach refers to the limited number of colors a human being can differentiate. Reijers et al. (REIJERS et al., 2011) describe an experiment with users to investigate hypotheses, one of them being "The use of colors to highlight matching operator transitions will have a significant, positive impact on understanding accuracy", which was supported by the experiment's result with statistical significance.

Kummer et al. (KUMMER; RECKER; MENDLING, 2016) discuss that different sets of colors accepted by different cultures make a difference when used to highlight parts of a model for using this model as a communication means. One of their hypotheses was that models with colored elements would be easier to understand than the ones with no coloring, to members of the Confucian culture. They performed experiments with users (holding same level of familiarity with BPMN process modeling) and found the hypothesis was statistically supported.

Mueller-Wickop et al. (MUELLER-WICKOP et al., 2011) use colors to distinguish, for example, different event types (e.g., a green event means a financial value entry) in the context of accounting information systems. According to the authors, for auditors understand financial entries flow in this context, they should be provided with a process-oriented view of these entries.

Other studies explore only shape mapping (JOSCHKO; WIDOK; PAGE, 2013; MERINO et al., 2016; STROPPI; CHIOTTI; VILLARREAL, 2011). For example, Stroppi et al. (STROPPI; CHIOTTI; VILLARREAL, 2011) propose a BPMN extension to provide a better understanding of requirements from the resource perspective. Their extension is composed by a squared shape containing relevant textual information from the resource point of view (e.g., privileges the resource has to execute the activity) attached through a line to any process model activity.

Two articles (JOSCHKO; WIDOK; PAGE, 2013; MERINO et al., 2016) propose new elements to increment process models represented in BPMN and explore shapes for different symbols with distinct semantics. In the first one (JOSCHKO; WIDOK; PAGE, 2013), the new elements intend to smooth control operations on wind farms domain. An example of a proposed element is the “wind farm signal event”, represented by a symbol that is a windmill in a circle, which could be used to send malfunction signals. On the other hand, Merino et al. (MERINO et al., 2016) propose new elements to convey more information about activities in a workflow. They aim at improving machine-understandability of process models for real-time monitoring in manual services contexts. One of these new elements is called “advanced decision point”, represented by an interrogation point in a circle. It is capable of identifying if a particular condition is met to decide to which branch move during process execution (e.g., if the process is of a tire to be calibrated, the decision point may enter a loop to inflate tire while the value of calibration is lower than the desired value).

Color and shape mapping are addressed by other authors (LAUE; AWAD, 2011; SALNITRI; DALPIAZ; GIORGINI, 2017; GALL et al., 2015; GUO; BROWN; RASMUSSEN, 2012; KATHLEEN; ROSS; KRIGLSTEIN, 2014). Salnitri et al. (SALNITRI; DALPIAZ; GIORGINI, 2017) describe a security-oriented BPMN extension. They aim at representing real-world requirements of the security domain through a set of proposed symbols that can be attached to BPMN elements. One of these symbols represents the availability of an element as a circular clock shape with the number “24” in the center. Another symbol indicates that a data object can only be accessed by authorized personnel, being visually represented by a circular red wax seal. Both symbols are surrounded by a thick circular border colored in orange. Under different configurations of models presented during an experiment, the users found the models more understandable with the proposed symbols than without them.

The study by Gall et al. (GALL et al., 2015) investigate the use of symbols such as green-check marks and red-subtraction marks to represent, respectively, elements that were added to and removed from different versions of a process model. Laue and Awad (LAUE; AWAD, 2011) focused on sequence flow errors and fragments of the model that can make it difficult to understand. To highlight the elements that are responsible for any identified error, they suggest attaching a white “x” in a red circle mark to the element. Then, when the user hovers the mouse over this mark, a textual message about the error should be exhibited.

Although this cluster is characterized by focusing on color and shape mappings, we also observed other features being used. In the article by Laue and Awad (LAUE; AWAD, 2011), the query action *identify* is present when the user hovers the mark and can read the characteristics of the element it represents. Reijers et al. (REIJERS et al., 2011) use the manipulate action *change* when the user is enabled to activate and deactivate the coloring of matching operators.

Cluster 5: characterized by “faceting”, “attribute targets”, and “all data targets”

The studies in cluster 5 are the ones that most explore “faceting” elements, i.e., *juxtapose*, *partition*, *superimpose*). As those in cluster 2, they are directed to “attribute targets”, although also targeting “all data”, as can be observed in Figure 3.7. In this cluster, it is more evident the use of *similarity* when dealing with attribute targets, *superimpose* as facet, *compare* (as query action) and *outliers*, for "all data targets" (see

Figure 3.8). Moreover, all studies in this cluster, as seen in Table 3.4, explore the visualization of differences between processes models. Their main focus is to make users able to *compare* process models identifying *similarities* and differences among them. In some studies (KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011; CORDES; VOGELGESANG; APPELRATH, 2015), when an element is removed, its control flow is shown in different colors (orange (KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011) or red (CORDES; VOGELGESANG; APPELRATH, 2015)), while the added elements are displayed in green (in both); and, the changed elements are displayed in yellow (CORDES; VOGELGESANG; APPELRATH, 2015).

Kriglstein and Rinderle-Ma (KRIGLSTEIN; RINDERLE-MA, 2012) conduct a systematic literature review about how visualization is used to show differences among process models, and performed a survey to identify the expectations of users regarding this. Among the findings of their study, the authors identified that the tools only highlight changes between processes but do not allow users to trace the changes across the processes' different versions.

Kabicher-Fuchs et al. (KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012) presented an approach for visualizing the differences among versions of a process model. In their paper, they explored many of the visualization data analysis elements (Tables A.1 and A.2), such as *change*, by allowing the user to change the way a process model is presented. In their proposal, at first, the process model is displayed as is, i.e., with no visual augmentation. When the user *selects* a month of the year, in the timeline view, the differences between process model versions can be seen through the *superposition* of the different versions of the process model with *color* changes applied to activities and control flows (e.g., red corresponding to removed elements and green to added elements). This timeline view is a Gantt chart, presenting all process models' lifetime durations across the months. They adopted multiple views, which are *juxtaposed* and coordinated.

Another study (KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013) proposes highlighting not only the differences but also similarities between process models aiming to support both the comparison of two process models as well as the comparison of instance traffic between two process models at different moments in time.

Finally, other articles (KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013) also present proposals to enable users to find *outliers*. While in Kabicher-Fuchs et al.'s proposal (KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012) the timeline view allows observing which are the months when a specific process model changed the most, the work by Krigelstein et al. (KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013) identify which are the paths that are most executed along the process.

3.4 Results and Discussion

In this section, we discuss our findings by answering the research questions we addressed in our systematic literature review.

3.4.1 Research question 1 (primary): what is being investigated in the topic visualization of business process models

The main research question, RQ1, is the most generic one. The objective of this review was to identify, broadly, what is being investigated regarding visualization of process models, avoiding restrictions to this answer as much as possible.

After the data extraction, on the one hand, we classified the selected studies into six categories, as shown in Section 3.1.2, so we could have an overview of the literature from the main scope of the papers' proposals. We found out that the articles deal with (i) augmentation of existing process modeling elements, (ii) creation of new elements, (iii) exploration of the 3D space, (iv) suggestions of different ways for visualizing process models information and (v) process modeling mistakes, and (vi) visualization of process models from distinct point of views. While most studies focus on the visualization of the process model diagram itself, 16.28% explore both information visualization and process model diagram in the same view, and 9.30% of them only explore information visualization about process models.

On the other hand, after the data extraction related to the visualization analysis framework, the absolute majority of the studies, specifically 93.02% of them, provide ways to the users "to consume" information rather than "to produce". A little more than half of the papers, more precisely 65.12% of them, provide some kind of user interaction within their proposals, either by selecting elements or navigating through process models, for example. Surprisingly, still, almost 35% (34.88%) of the studies are static visualizations.

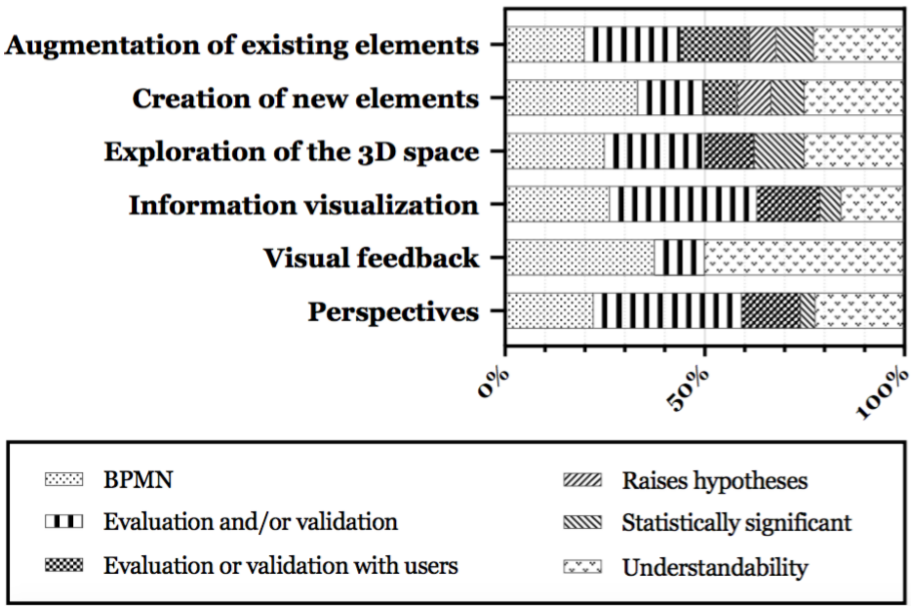
To avoid missing any possible existing grouping or relation among the studies analyzed under Munzner's visualization analysis framework, they were clustered using the

k-means algorithm. We obtained six clusters differing from each other in different aspects, although one of them (identified as Cluster 0) is more neutral facing the others (see Figure 3.7), with its set of papers not showing any particular feature that might differentiate them from the others.

All clusters contain studies that allow users to "search" for information at some extent, only Cluster 2 showing no papers that handle "network data". So, Cluster 2's articles address mostly information visualization features. Cluster 4 is the largest one and mainly constituted of studies that do not explore user interaction in their design choices.

Considering the 46 selected studies, exactly half of them are based on BPMN and 52.17% present evaluation or validation of their proposals, while in 34.78% such evaluation or validation involves experiments with users. 13.04% of the articles investigate hypotheses, and 15.22% present results that are considered statistically significant. Exactly half of the studies aims at improving the understandability of process models explicitly. It is noteworthy that only 19.56% of the papers describe an online tool, out of which one is not accessible anymore. Also noteworthy is the fact that, according to Google Scholar[11], only 28.26% of the studies we surveyed here, are referenced by 10 or more other papers.

Figure 3.9: Data extraction distribution over each category.



[11] https://scholar.google.com/

3.4.2 Research question 2: are the studies concerned with improving the understandability of process models

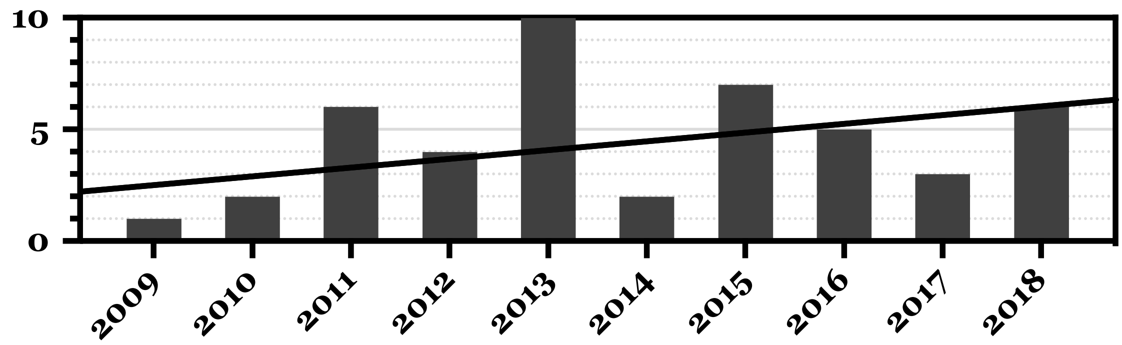
RQ2 focus on identifying the papers that aim at improving the understandability of process models. Since there are many studies (ECKLEDER et al., 2009; HIPP; MUTSCHLER; REICHERT, 2012; AWAD; WESKE, 2010; ONGGO; KARPAT, 2011), just to cite a few, addressing this issue, we hypothesized that the majority of the selected papers would focus on this problem. However, the analysis of the frequency of the keywords showed that not all selected papers have that goal.

After performing the data extraction process (Table 3.2), we could not confirm our hypothesis since only half of the studies are concerned with improving the understandability of process models by proposing different ways of visualizing the process models or visualizing information about process models. Although there are studies like (STROPPI; CHIOTTI; VILLARREAL, 2011) that are not explicitly concerned with that issue, they are concerned with, for example, improving communication between different parties involved with the process modeling (e.g., process participants, process owners, process analysts). Approximately 89.00% of the papers in Cluster 1, 64.28% of the studies in Cluster 4, and 57.14% in Cluster 0 explicitly mention being concerned with improving the understandability of process models. This scope was not found in papers belonging to Clusters 2, 3 and 5.

3.4.3 Research question 3: are there open problems for further research on this topic

We answer this research question from two perspectives: the studies' main scope and the visualization analysis. Regarding the studies' main scope, Figure 3.9 presents

Figure 3.10: Total articles published per year (2009-2018).



the recurrence of data-extraction per category. As can be seen, the "visual feedback" category may pose challenges for further research since it is the one that most lacks studies presenting evaluation or validation of their proposals, especially involving experiments with users. This category does not show any paper providing results with statistical significance. Also, 3 out of 4 of the papers in this category include user interaction in their design choices. Since half of the studies base their proposals on BPMN and, being BPMN an ISO standard, this also could indicate space for further research on visualization of process models modeled with BPMN.

From the high-level visualization analysis point of view, only two studies present proposals that enable users "to produce" information. This is an important characteristic yet to be explored in future research. Furthermore, although Clusters 2 and 3 are composed of studies proposing visualization of data about process models or collections of process models, only 30.43% of all studies indeed present proposals regarding such features. So, information visualization applied to process models or collections of process models could also be further explored, mainly to improve user interaction and the understanding of process models diagrams. Finally, few studies explore 3D representations for displaying process models and related information, this finding corroborating what is stated by Oberhauser et al. (OBERHAUSER; POGOLSKI; MATIC, 2018).

3.4.4 Research question 4: how active is the research on this topic since 2009

Research question 4 is related to the year and source (journal or conference) of publication of each paper. We grouped the selected articles per year to investigate if the literature in this area tends to growth or decrease in the coming years. As can be seen in Figure 3.10, linear regression shows that the number of articles per year is increasing, i.e., there is a tendency of more publications about visualization of process models in the next years. Table 3.5 provides the distribution of papers per year, and Table 3.6 presents articles per source. One can notice that 54.35% of the articles were published in journals and 45.65% in conferences, in the last ten years (i.e., the period between January 2009 and December 2018).

Table 3.5: Distribution of articles per year (2009-2018).

Year	Articles	Total (%)
2009	(ECKLEDER et al., 2009)	1 (2.17%)
2010	(AWAD; WESKE, 2010; EFFINGER; SPIELMANN, 2010)	2 (4.35%)
2011	(REIJERS et al., 2011; LAUE; AWAD, 2011; ONGGO; KARPAT, 2011; MUELLER-WICKOP et al., 2011; STROPPI; CHIOTTI; VILLARREAL, 2011; KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011)	6 (13.04%)
2012	(HIPP; MUTSCHLER; REICHERT, 2012; KRIGLSTEIN; RINDERLE-MA, 2012; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; GUO; BROWN; RASMUSSEN, 2012)	4 (8.70%)
2013	(STORCH; LAUE; GRUHN, 2013; JOSCHKO; WIDOK; PAGE, 2013; LEITNER et al., 2013; KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013; KOLB; REICHERT, 2013; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; KRIGLSTEIN; RINDERLE-MA, 2013; FIGL; KOSCHMIDER; KRIGLSTEIN, 2013; REICHERT, 2013; HOLZMÜLLER-LAUE et al., 2013)	10 (21.74%)
2014	(HIPP et al., 2014; KATHLEEN; ROSS; KRIGLSTEIN, 2014)	2 (4.35%)
2015	(IVANCHIKJ; FERME; PAUTASSO, 2015; WITT et al., 2015; HIPP et al., 2015; CORDES; VOGELGESANG; APPELRATH, 2015; GALL et al., 2015; PERALTA et al., 2015; PINI; BROWN; WYNN, 2015)	7 (15.22%)
2016	(KUMMER; RECKER; MENDLING, 2016; EMENS; VANDERFEESTEN; REIJERS, 2016; GULDEN; ATTFIELD, 2016; KOSCHMIDER; FIGL; SCHOKNECHT, 2016; MERINO et al., 2016)	5 (10.87%)
2017	(CORRADINI et al., 2017; SALNITRI; DALPIAZ; GIORGINI, 2017; JOŠT et al., 2017)	3 (6.52%)
2018	(JÚNIOR et al., 2018; POLDERDIJK et al., 2018; CORRADINI et al., 2018; KRENN; KEPLER, 2018; CABALLERO et al., 2018; OBERHAUSER; POGOLSKI; MATIC, 2018)	6 (13.04%)

Table 3.6: Distribution of articles per publication type (2009-2018).

Type	Articles	Total (%)
Conference paper	(JÚNIOR et al., 2018; ECKLEDER et al., 2009; STORCH; LAUE; GRUHN, 2013; HIPP et al., 2014; IVANCHIKJ; FERME; PAUTASSO, 2015; ONGGO; KARPAT, 2011; MUELLER-WICKOP et al., 2011; JOSCHKO; WIDOK; PAGE, 2013; EFFINGER; SPIELMANN, 2010; STROPPI; CHIOTTI; VILLARREAL, 2011; KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011; KRIGLSTEIN; RINDERLE-MA, 2012; KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012; FIGL; KOSCHMIDER; KRIGLSTEIN, 2013; GUO; BROWN; RASMUSSEN, 2012; KATHLEEN; ROSS; KRIGLSTEIN, 2014; POLDERDIJK et al., 2018; CORRADINI et al., 2018; KRENN; KEPLER, 2018; CABALLERO et al., 2018; OBERHAUSER; POGOLSKI; MATIC, 2018)	21 (45.65%)
Journal	(REIJERS et al., 2011; KUMMER; RECKER; MENDLING, 2016; EMENS; VANDERFEESTEN; REIJERS, 2016; HIPP; MUTSCHLER; REICHERT, 2012; GULDEN; ATTFIELD, 2016; AWAD; WESKE, 2010; LAUE; AWAD, 2011; WITT et al., 2015; CORRADINI et al., 2017; LEITNER et al., 2013; KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013; HIPP et al., 2015; KOSCHMIDER; FIGL; SCHOKNECHT, 2016; MERINO et al., 2016; SALNITRI; DALPIAZ; GIORGINI, 2017; KOLB; REICHERT, 2013; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; KRIGLSTEIN; RINDERLE-MA, 2013; REICHERT, 2013; CORDES; VOGELGESANG; APPELRATH, 2015; GALL et al., 2015; PERALTA et al., 2015; PINI; BROWN; WYNN, 2015; HOLZMÜLLER-LAUE et al., 2013; JOŠT et al., 2017)	25 (54.35%)

3.4.5 Research question 5: who is leading research on this topic

RQ5 is concerned with the most recurrent authors, both in the selected studies and in publications cited by these studies. The search sources and the research assistant applications we used, provided us with a variety of data about the articles, and we extracted the authors and title of each one of the 46 selected papers. We also obtained the authors and titles of around 1450 papers referenced by the selected articles.

All of the most recurrent authors of the selected studies (see Table 3.7) appear among the most recurrent referenced authors in these papers. A possible conclusion about

Table 3.7: Distribution of articles per most recurrent author among the selected papers (2009-2018).

Author	**Article**	**Affiliation**	**Total (%)**
Ralf Laue	(LAUE; AWAD, 2011)	Univ. of Leipzig, Computer Science Faculty, Germany	2 (4.35%)
	(STORCH; LAUE; GRUHN, 2013)	Univ. of Applied Sciences of Zwickau, Dept. of Info. Science, Germany	
Jan Mendling	(ECKLEDER et al., 2009)	Humboldt-Universität zu Berlin, Germany	3 (6.52%)
	(REIJERS et al., 2011)	Humboldt-Universität zu Berlin, Germany	
	(KUMMER; RECKER; MENDLING, 2016)	Wirtschaftsuniversität Wien, Austria	
Hajo A. Reijers	(ECKLEDER et al., 2009)	Eindhoven Univ. of Technology, Netherlands	3 (6.52%)
	(REIJERS et al., 2011)	Eindhoven Univ. of Technology, Netherlands	
	(EMENS; VANDERFEESTEN; REIJERS, 2016)	Eindhoven Univ. of Technology, Netherlands	
Manfred Reichert	(HIPP; MUTSCHLER; REICHERT, 2012)	Univ. of Ulm, Institute of DB and Information Systems, Germany	5 (10.87%)
	(KOLB; REICHERT, 2013)	Univ. of Ulm, Institute of DB and Information Systems, Germany	
	(REICHERT, 2013)	Univ. of Ulm, Institute of DB and Information Systems, Germany	
	(HIPP et al., 2014)	Univ. of Ulm, Institute of DB and Information Systems, Germany	
	(HIPP et al., 2015)	Univ. of Ulm, Institute of DB and Information Systems, Germany	
Stefanie Rinderle-Ma	(KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011)	Univ. of Vienna, Faculty of Computer Science, Austria	5 (10.87%)
	(KRIGLSTEIN; RINDERLE-MA, 2012)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(KRIGLSTEIN; RINDERLE-MA, 2013)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(GALL et al., 2015)	Univ. of Vienna, Faculty of Computer Science, Austria	
Simone Kriglstein	(KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011)	Univ. of Vienna, Faculty of Computer Science, Austria	9 (19.56%)
	(KRIGLSTEIN; RINDERLE-MA, 2012)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013)	Univ. of Vienna, Faculty of Computer Science, Austria	

Table 3.7 continued from previous page

Author	Article	Affiliation	Total (%)
	(KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(KRIGLSTEIN; RINDERLE-MA, 2013)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(FIGL; KOSCHMIDER; KRIGLSTEIN, 2013)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(KATHLEEN; ROSS; KRIGLSTEIN, 2014)	Univ. of Vienna, Faculty of Computer Science, Austria	
	(GALL et al., 2015)	Univ. of Vienna, Inst. for Design and Assessment of Technology, Austria	

this finding is that we could gather the most relevant papers in the research field, which makes us confident about the results obtained with our systematic literature review. The most recurrent authors within the selected studies are Simone Kriglstein, authoring 9 papers, which represent 19.56% of our selected papers; Stefanie Rinderle-Ma and Manfred Reichert, authoring 5 articles each; Jan Mendling, Hajo A. Reijers, and other 4 authors, authoring 3; Ralf Laue, and other 10 authors, authoring 2 articles. The most recurrent referenced author is Jan Mendling, with 110 citations distributed among his 59 papers.

Also, Table 3.7 allows observing the cooperation among authors. For example, Simone Kriglstein and Stefanie Rinderle-Ma were the two authors that most cooperated, having worked together in 5 papers (KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011; KRIGLSTEIN; RINDERLE-MA, 2012; KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013; KRIGLSTEIN; RINDERLE-MA, 2013; GALL et al., 2015), while Jan Mendling and Hajo A. Reijers co-authored 2 papers (ECKLEDER et al., 2009; REIJERS et al., 2011).

3.5 Final Comments

We conducted a systematic literature review on the visualization of business process models, considering papers published in the last ten years (from January 2009 to December 2018). After the application of the exclusion and inclusion criteria on a set of 1686 papers, 46 studies were selected to be fully read and pass through the data extraction process.

Based on the data extracted from the 46 selected papers, we analyzed them from two point of views. The first one enabled us to group them into six categories, according to their main scope: augmentation of existing process modeling language elements, creation of new process modeling language elements, exploration of the 3D space for process modeling, information visualization about process models, visual feedback concerning problems detected in process models, and support for different perspectives of a process model. Then, we identified which are the main areas that are being explored regarding visualization of business process models. We concluded that the categories less explored and which could present research challenges for further exploration are “visual feedback” (concerning problems detected in process models) and “information visualization” (about process models) since the papers addressing these aspects present no or few results from evaluation or validation of their proposals.

From the second point of view, we analyzed the selected studies based on a visualization analysis and then we obtained a high-level abstract view of the studies' proposals. After that, we identified open problems concerning the approaches presented in the articles, such as few studies exploring user interaction and, mainly, few proposals allowing users to produce information from process models. It might be interesting, for example, to explore how to enable users to annotate process models with their own domain-specific (or subject specific) information for further reuse during the modeling task.

Among the selected studies, 52.17% of the papers performed evaluation or validation, out of which only 30.43% conducted tests with users. Moreover, although some papers propose generic approaches theoretically easy to adapt to specific modeling languages, another interesting finding is that half of the selected studies base their approaches on BPMN. From 2014 to 2017, there are 23 studies within the selected papers, roughly half of them have based their approaches on BPMN. We understand this aspect as an open opportunity too. Since BPMN is an ISO standard and there are many tools based on BPMN, there should be more research intending to improve the knowledge about this standard.

As a limitation of our systematic literature review, we understand that the data extraction was constrained to some extent, mainly regarding the understandability aspect: if a study did not express to be aiming at improving the understandability of the process model explicitly it was not marked as addressing this aspect during the data extraction process, which produced Table 3.2. One can also consider as a limitation the fact that current business process modeling tools, which also provide visualization features, were not included in this work since they are not described in papers that passed through our selection process.

Besides identifying the current research concerning to visualization of business process models, our motivation with this review was also to support and inspire researchers for further work aiming at bringing forward the field of business process model visualization, to have the advantages of information visualization helping the tasks of business process modeling and management.

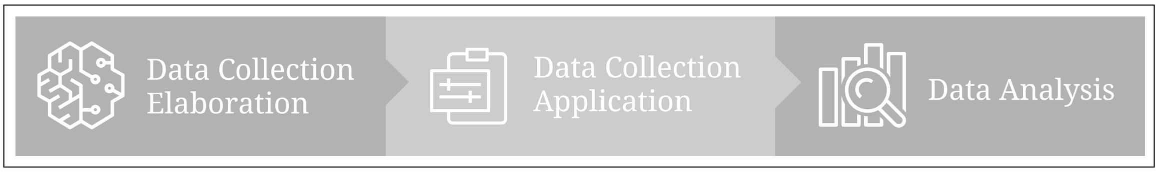
4 FEEDBACK ABOUT PROBLEMS DURING THE BUSINESS PROCESS MODELING TASK

According to Kriglstein et al. (KRIGLSTEIN; RINDERLE-MA, 2012), the main objective when performing a survey is to get a better view of users' perspectives regarding a certain situation. In this chapter, we present the survey we conducted to identify two important issues of our study:

- the demands of process modelers' in relation to visual feedback about problems during the business process modeling task; and,
- the satisfaction and learnability experienced by process modelers regarding the way that the modeling tools provide feedback about problems during the business process modeling task.

We decided to conduct this survey based on our findings from the SLR, in which we identified that there are challenges to be explored regarding visual feedback about problems in process models. Based on Fowler (FOWLER, 2007) and Bryman (BRYMAN, 2013), we designed our survey as composed of three main stages (Figure 4.1): *data collection elaboration*, to define how the data will be collected; *data collection application*, to define how the gathering of the data will be conducted; and *data analysis and reporting*, where we analyze and report the data collected.

Figure 4.1: Overview of the survey stages.



4.1 Research Questions

To guide our survey and help addressing the two issues mentioned above, we conceived and focused on answering the following research questions:

RQ$_{4.1}$ *What are the modelers' demands regarding feedback about problems in process model?*

RQ$_{4.2}$ *How is the satisfaction experienced by the modelers with the way the modeling tools they use give feedback about problems in process model?*

RQ$_{4.3}$ *How is the learnability experienced by the modelers considering the way the modeling tools they use give feedback about problems in process model?*

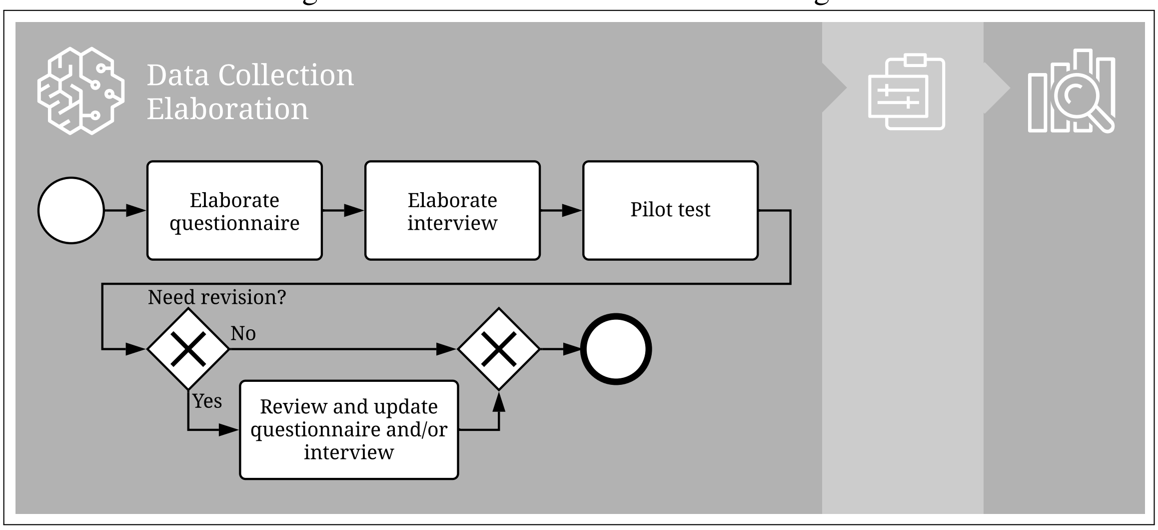
4.2 Data Collection Elaboration

According to Fowler (FOWLER, 2007), it is important to define the characteristics of the population we want to analyze, so the researcher can construct a data gathering tool including questions or answering options that allow filtering respondents. In our case, the population to be studied is comprised of academics or professionals, or both, with any level of experience in business process modeling task.

Based on the survey research questions, we decided to use a questionnaire and an interview to gather the data we need. We used a questionnaire to obtain answers from a sample of volunteers with some level of experience in business process modeling task (e.g., students, researchers, professors, analysts, etc.). Then, we used the results from interviews with business process management experts (e.g., researchers, professors, and analysts in the field of BPM), to verify if we could confirm and complement the data we gathered with the questionnaire.

The main stages of our data collection elaboration process (Fig. 4.2) are: questionnaire elaboration, interview elaboration, and pilot testing. After the pilot testing, we reviewed and updated the questionnaire and the interview.

Figure 4.2: Data collection elaboration stages.



4.2.1 Questionnaire elaboration

We defined the questionnaire structure and questions based on (BARBOSA; SILVA, 2010; ROBSON; MCCARTAN, 2016; BABBIE, 2015). Since we wanted to analyze data from participants with some level of experience with the process modeling task, we prepared our questionnaire to enable us to quickly identify and remove users that do not fit the profile needed for our survey.

Most of the questions are closed-ended (i.e., they present a predefined set of alternatives to be selected by the respondent), to ease the participant's completion of the questionnaire. Also, we avoided the use of double-barrelled and leading questions (i.e., questions that ask two questions in one, and that lead the respondent to a certain answer, respectively). Only two questions are open-ended (i.e., the respondents need to answer the question with their own words), more precisely the ones we defined to gather information about modelers' demands. All questions were grouped according to their semantics. Table B.1, in Appendix B, presents all the questions comprising the questionnaire with their respective answering alternatives, whenever it is the case.

Each survey research question is answered by a set of questionnaire questions. Questions 1 up to 16 are used to build the user profile of the participants. Questions 17 and 18, are used to answer $RQ_{4.1}$. Questions 19 up to 28 are used to answer $RQ_{4.2}$ and $RQ_{4.3}$.

4.2.2 Interview elaboration

To design and conduct the interview, we based our work on (BOYCE; NEALE, 2006; ROBSON; MCCARTAN, 2016; BABBIE, 2015). According to Kvale (KVALE, 1996), the minimum number of interviewees depends on what the researcher needs. So, we focused on individual interviews with few interviewees, aiming at exploring their perspectives, experiences and/or opinions on a specific aspect (BOYCE; NEALE, 2006), i.e., the feedback the tools they have been using provide about problems in process models. The interviews were planned to complement the answers to the two open-ended questions given by the respondents in the questionnaire.

We decided to conduct a semi-structured interview, where some questions to be answered by the interviewee are predefined, and others may arise along the interview (ROBSON; MCCARTAN, 2016). All questions comprising the interview are open-ended, and the predefined ones are the same open-ended questions of the questionnaire (questions

17 and 18, in Table B.1, Appendix B). Table C.1 contains the interview script, with the predefined questions. In the interview, we also avoided double-barrelled and leading questions, as we did with the questionnaire.

We arranged the interview script to be consistent with the key interview components as proposed by Boyce et al. (BOYCE; NEALE, 2006) (i.e., we started the interview with an initial thanking for participating, we present the purpose of the interview, confidentiality terms, and so on). Differently from the questionnaire, the interview questions are used to help answering the $RQ_{4.1}$, specifically.

4.2.3 Pilot testing

The questionnaire and interview questions were pilot tested (FINK, 2003; FOWLER, 2007; BRYMAN, 2013) with three colleagues: a MSc student from our research group, and a PhD student and a professor, both outside the group. None of them participated in the data collection process itself. Their feedback were valuable so we could reach the final questionnaire content and interview design. We could also verify if the answers being collected were enough to respond to the survey research questions. For example, after the pilot test, we removed two questions (for example, the one where we asked the participant which company he was linked to), and updated the answering alternatives of others by adding or removing options. The questions that were removed were those that brought doubt to the respondents and what we wanted to extract from the answers to them was already being gathered by the other questions.

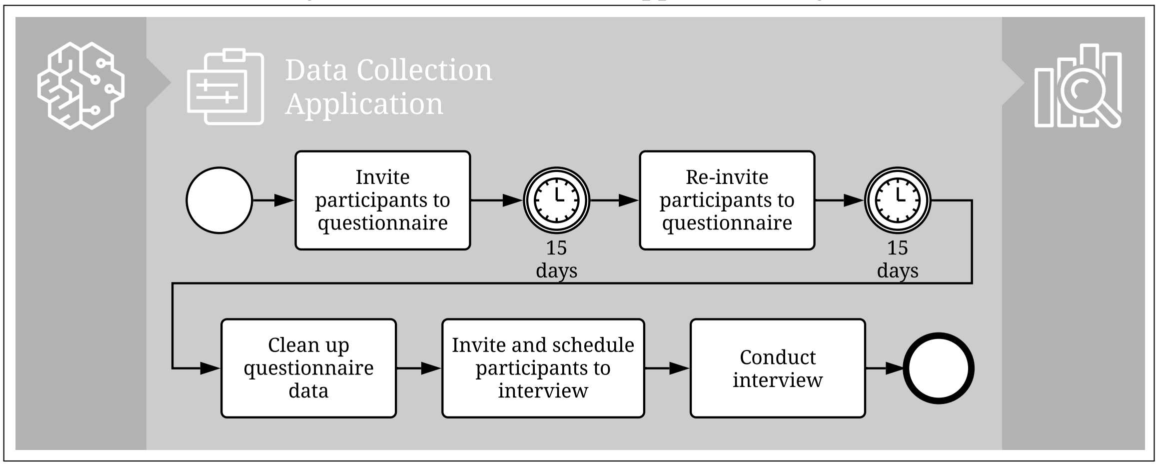
4.3 Data Collection Application

The main stages of our data collection application process (Figure 4.3) are: apply the questionnaire and perform the interview, which are further detailed in this section.

4.3.1 Questionnaire application

The questionnaire was shared in groups and forums previously known as being composed by academics and professionals in Business Process Management. The means

Figure 4.3: Data collection application stages.



used to gather participants to answer the questionnaire were: LinkedIn [1], Facebook [2], Twitter [3], and other peer-to-peer contact platforms such as e-mail, WhatsApp and Hangout. Moreover, five experts (researchers and professionals) were directly invited to answer the questionnaire. After fifteen days of the first sharing of the questionnaire (which happened on October 11, 2018), we re-invited the community to participate and, after another fifteen days, we stopped accepting answers.

After the data were collected, we performed a data cleaning process by removing incomplete questionnaires or those that have some characteristic that was not valid for the survey (for example, from respondents with no experience with business process modeling task). From a total of 61 respondents, 57 were considered for the data analysis phase.

4.3.2 Interview

The five experts directly invited to answer the questionnaire were also invited to be interviewed. All agreed to participate in this stage of our research too. Each one represents a distinct persona and has a different experience in business process modeling task. One of them is a professor in a renowned university in Brazil. Two of them are professors too, but they also have experience in companies, working directly in both environments on a daily basis. The fourth is a PhD student, and the fifth is a process analyst in a known enterprise with a branch in the state of Rio Grande do Sul, Brazil.

We scheduled the session with each invited participant managing to interview them

[1] https://www.linkedin.com/
[2] https://www.facebook.com/
[3] https://www.twitter.com/

at their preferred date and time. Nevertheless, we did not schedule more than one interview for the same day. The interviews were conducted following the interview script presented in Table C.1. Moreover, based on Boyce et al. (BOYCE; NEALE, 2006), during the semi-structured interviews and according to the interviewee answers, other questions were added; and, whenever we felt necessary, we used probe questions (e.g., *Could you explain that further? Can you give an example?*).

4.4 Data Analysis

In the data analysis stage, for the descriptive analysis we followed Babbie (BABBIE, 2015); for the satisfaction and learnability analysis, the studies of Brooke (BROOKE, 1996) and Lewis and Sauro (LEWIS; SAURO, 2009); and, for the content analysis of the answers to the two open-ended questions (questions 17 and 18, Table B, Appendix B), we followed mainly (ELO; KYNGÄS, 2008; BRYMAN, 2013). Figure 4.4 presents the stages of our data analysis, and Table 4.1 shows which data analysis method was used to answer each research question.

Figure 4.4: Data analysis stages.

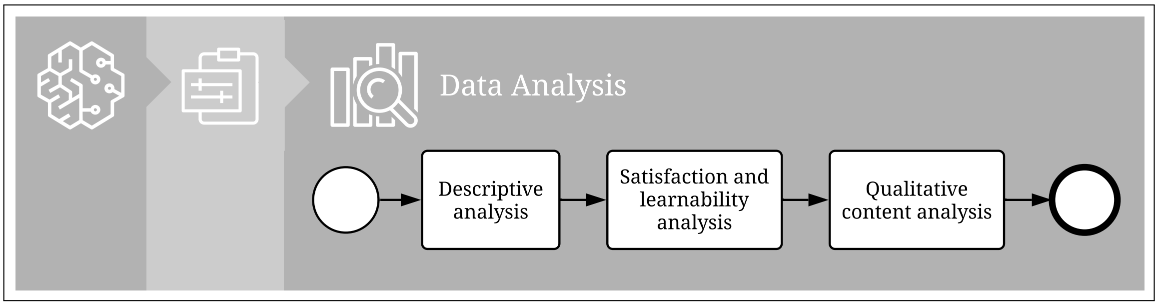


Table 4.1: Data analysis methods used to answer each research question.

Data analysis method	$RQ_{4.1}$	$RQ_{4.2}$	$RQ_{4.3}$	User profile
Descriptive analysis				x
System usability scale		x	x	
Content analysis	x			

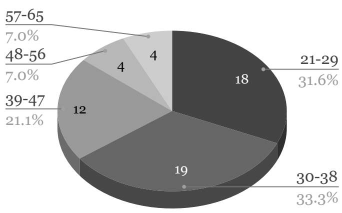
4.4.1 Descriptive analysis

In this section, we present observed relations between the level of process modeling experience of the questionnaire respondents and their satisfaction regarding four different ways of feedback regarding problems in process model.

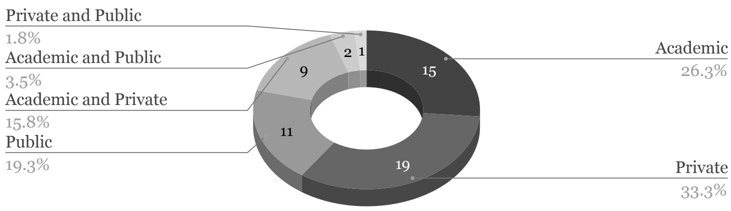
Figure 4.5 presents basic data about participants' profile. The majority of the participants, 37 out of 57 (i.e., 64.90%), are aged between 21 and 38 (Figure 4.5a). The Private sector, followed by Academia, are the ones with most respondents (33.30% and 26.30%, respectively, Figure 4.5b). The participants are mainly Process Analysts, Students, or Managers (28.07%, 21.05%, and 15.79%, respectively, Figure 4.5c).

Figure 4.5: Participants general information. (a) Age, (b) Main occupation, and (c) Sector to which the participant is linked to (e.g., academic, private, public).

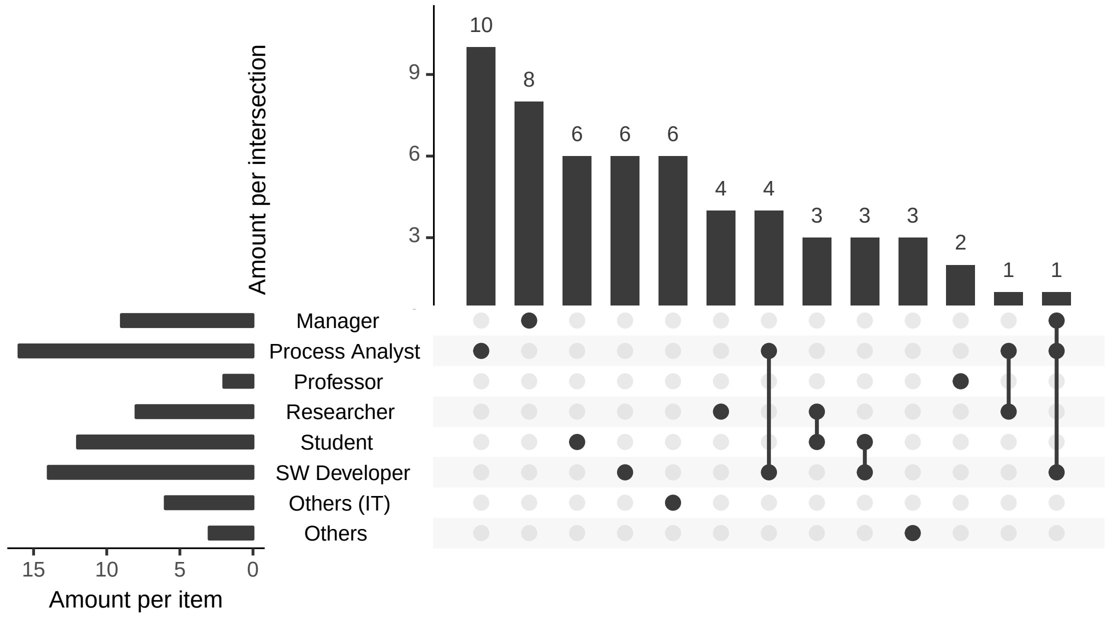
(a) Age



(b) Sector



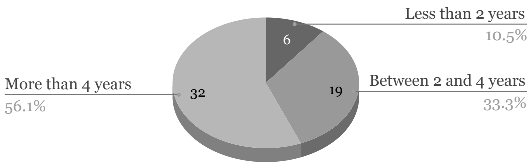
(c) Main occupation. To answer the question regarding this subject, respondents could select more than one possible answer. The vertical lines connect the multiple answers. In the plot, for example, four respondents answered both "Process Analyst" and "Software (SW) Developer".



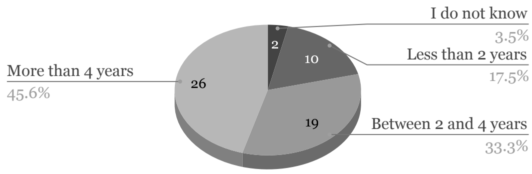
Regarding participants' level of experience with the process modeling task (Figure 4.6), 56.10% of the participants have for four years or more of knowledge in business

Figure 4.6: Participants experience in process modeling.

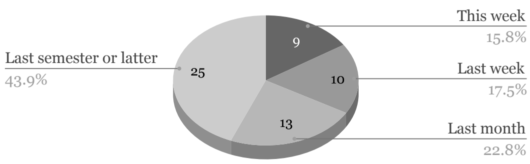
(a) For how long have you known business process modeling?



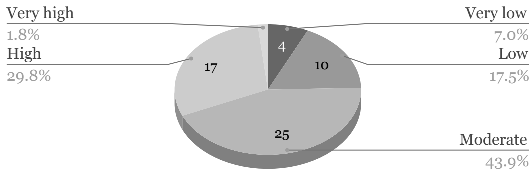
(b) For how long have you known BPMN?



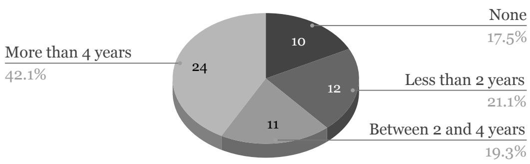
(c) When was the last time you modeled a business process?



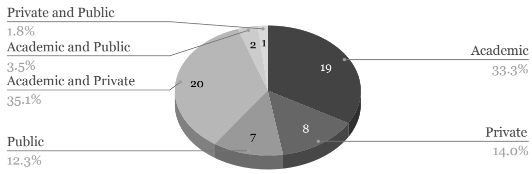
(d) How do you consider your level of experience in business process modeling



(e) How many years of professional experience (employee/trainee) in business process modeling do you have?



(f) Where did your process modeling learning take place in?

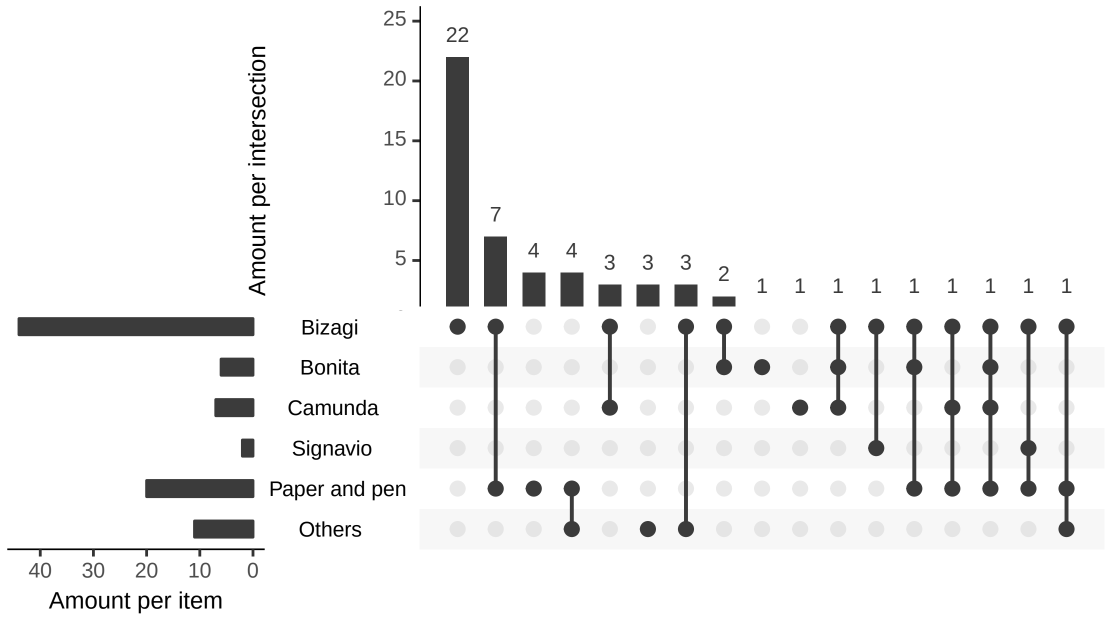


process modeling (Figure 4.9a), 45.60% know BPMN for four years or more (Figure 4.9b), 43.90% modeled a business process in the last semester or later (Figure 4.9c). Moreover, 43.90% consider their experience with business process modeling moderate, while 29.80% consider it high, and 17.50% consider it low (Figure 4.9d).

When asked about how many years of professional experience with process modeling they have, participants said they hold four or more years in 42.10% of the cases (Figure 4.6e). On the other hand, 17.50% of the respondents hold no professional experience in process modeling. After analyzing these participants, we identified that only one of them considered having high knowledge in the process modeling task, while the majority of the others considered to have low or very low knowledge in this field. Also, exactly half of them are students, and their learning in process modeling took place at the university, which matches with the results when considering the whole set of respondents (Figure 4.6f).

Participants were asked three questions, in a sequence, so we could learn when was the last time they modeled a process model (already discussed above), and which modeling tool and modeling notation they used. Figure 4.7 shows the respondents answers'

Figure 4.7: Modeling tool participants used when they last modeled a process.



regarding which modeling tools they used. The vertical lines show sets of modeling tools used by the same respondents. It can be observed that they used Bizagi and another tool in 71.93% of the cases, while 38.60% of the participants used exclusively Bizagi.

Figure 4.8 presents the intersection among respondents answers regarding which modeling notation was used. The respondents used BPMN and another notation in 89.47% of the cases, while 73.68% of the participants used exclusively BPMN.

We analyzed the level of experience in process modeling (answers to question 9 of the questionnaire), and how satisfied the participants are with four different manners to visualize the same problem in four different process modeling tools (answers to questions 13 to 16 of the questionnaire). Figure 4.9 presents an overview of the findings.

We also divided the participants into three groups, according to their level of experience in process modeling, to identify if modelers with different levels of experience have different levels of satisfaction considering the same feedback about problems in process model (Table 4.2): “Group A”, composed by participants with “High” and “Very high” levels of experience; “Group B”, composed by participants with “moderate” level of experience; and, “Group C” composed by participants with “Low” and “Very low” level of experience. For each group, we show the percentages concerning the level of satisfaction reported by respondents to each different feedback about problems in process model (represented by columns Q13 up to Q16 of Table 4.2).

To conclude, we found that our participants are composed by process analysts or students, with four or more years of experience in BPMN and process modeling task,

Figure 4.8: Notation participants used when they last modeled a process.

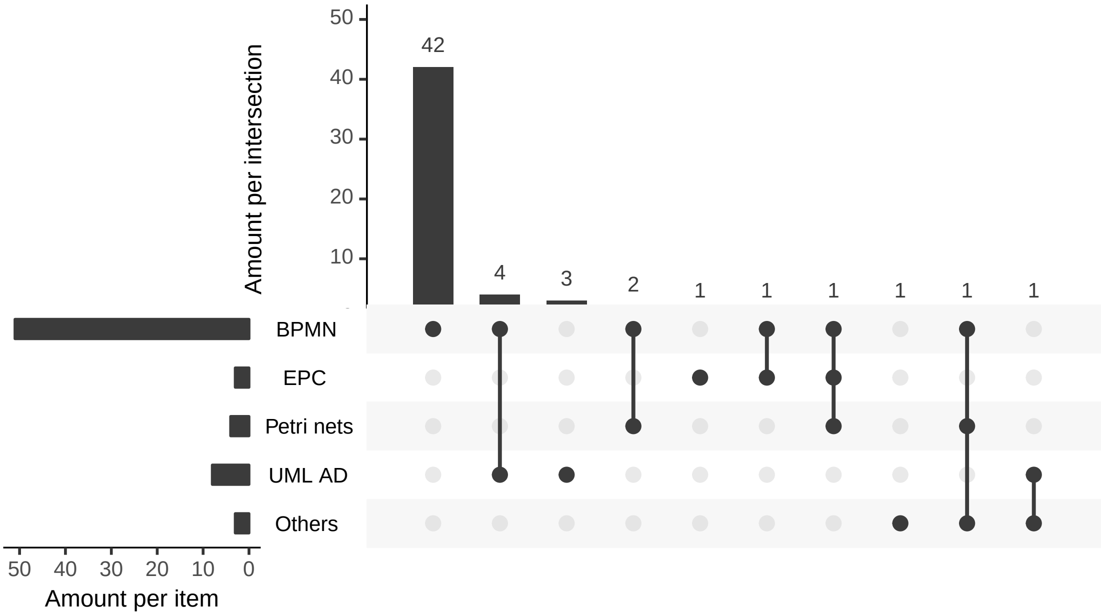
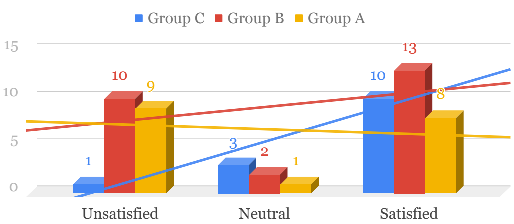
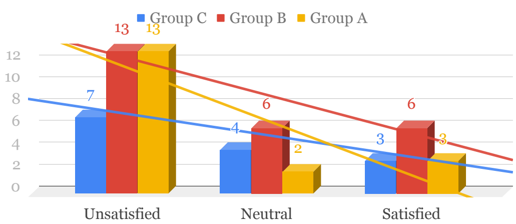


Figure 4.9: Participants satisfaction considering their level of experience and the four different manners to present problems in process models, based on four different BPMN-based process modeling tools. The respondents were not told about which tools were used in this part of the survey: only representative images were shown to them. The columns in the charts (a) to (d) represent the amount of participants per level of experience.

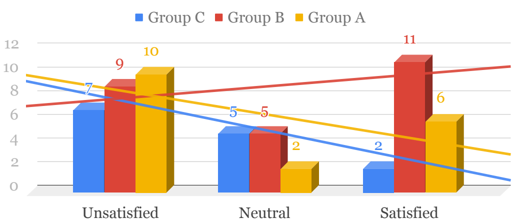
(a) Bizagi (question 13, Figure B.1)



(b) Bonita (question 14, Figure B.1)



(c) Camunda (question 15, Figure B.1)



(d) Signavio (question 16, Figure B.1)

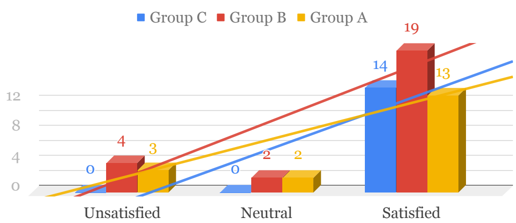


Table 4.2: How satisfied are process modelers with different levels of experience with the process modeling task regarding the manner four different process modeling tools provide feedback about problems in process models (see questions 13 up to 16,in Table B.1). Note that: “Uns.” represents unsafisfied and very unsatisfied; “Neu.” represents neutral; and, “Sat.”, represents satisfied and very satisfied.

Question		**Group** (number of participants in group) *Level of experience*		
		Group A (18) *High and very high*	**Group B** (25) *Moderate*	**Group C** (14) *Low and very low*
Q13	**Uns.**	9 (50.00%)	10 (40.00%)	1 (7.14%)
	Neu.	1 (5.56%)	2 (8.00%)	3 (21.43%)
	Sat.	8 (44.44%)	13 (52.00%)	10 (71.43%)
Q14	**Uns.**	13 (72.22%)	13 (52.00%)	7 (50.00%)
	Neu.	2 (11.11%)	6 (24.00%)	4 (28.57%)
	Sat.	3 (16.67%)	6 (24.00%)	3 (21.48%)
Q15	**Uns.**	10 (55.56%)	9 (36.00%)	7 (50.00%)
	Neu.	2 (11.11%)	5 (20.00%)	5 (35.71%)
	Sat.	6 (33.33%)	11 (44.00%)	2 (14.28%)
Q16	**Uns.**	3 (16.67%)	4 (16.00%)	0 (0.00%)
	Neu.	2 (11.11%)	2 (8.00%)	0 (0.00%)
	Sat.	13 (72.22%)	19 (76.00%)	14 (100.00%)

having modeled a process in the last semester or more recently, using Bizagi and BPMN. They are mainly from the Academic and/or Private sectors. Moreover, 43.90% of the participants consider their level of experience in the business process modeling task as moderate, 31.60% consider it high or very high, and 24.50% consider it low or very low. They are unsatisfied with the manner the tools Bizagi, Bonita, and Camunda present the problems in process models, and satisfied with Signavio.

We observed that participants with different levels of experience do not show the same levels of satisfaction regarding the same feedback about problems in process model presented by different modeling tools. For example, participants with higher levels of experience (Group A) tend to be more unsatisfied than participants with lower levels of experience (Group C) concerning the way Bizagi (Q13) gives feedback about problems, while participants with lower levels of experience are more satisfied with Bizagi.

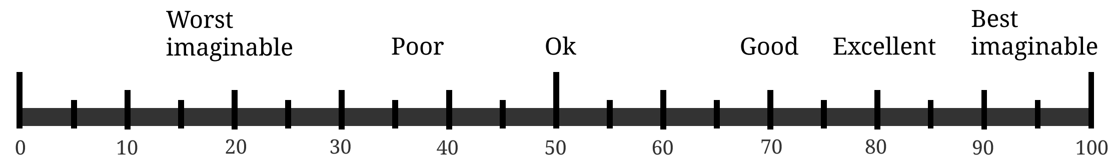
4.4.2 User satisfaction and learnability

We used the System Usability Scale (SUS) questionnaire (BROOKE, 1996) (questions 19 to 28, Table B.1) to identify user satisfaction and learnability regarding the feedback about problems provided by the process modeling tool the modelers use. SUS questionnaire is extremely simple and reliable, intended to get a snapshot of people's satisfaction using a system (BROOKE, 2013), and is composed by ten questions to be answered in a five-point Likert scale, varying from "Strongly disagree" to "Strongly agree". To the range of possible answers, we added the alternative "I never modeled a process", for allowing us to remove respondents with no process modeling experience.

To calculate the SUS score, we sum up the answers per respondent, removing 1 from the answer value to odd questions, and removing the answer value of even questions from 5. For example, if a given respondent answers "agree" to question 19 and "disagree" to question 20, we will have the values 4 and 2, respectively, as answers to these questions; and, we will consider "4 - 1" and "5 - 2" to sum up these answers. After performing this procedure, we multiply the sum by 2.5. This is done so the scale goes from 0 to 100 (BROOKE, 1996). With the scores for each respondent we calculate the mean SUS score.

According to Bangor et al. (BANGOR; KORTUM; MILLER, 2018), a SUS-score between 71.4 and 85.5 is considered to be amid "good" and "excellent" (see Figure 4.10). However, the final SUS-score considering all 57 participants answers was of 50.74, which is a satisfaction among "poor" and "ok".

Figure 4.10: SUS-scores represented by "adjective scale", as proposed by Bangor et al. (BANGOR; KORTUM; MILLER, 2018).



To assess the learnability, we based our analysis on Lewis and Sauro (LEWIS; SAURO, 2009), which says that such information can be extracted from answers to the SUS questionnaire by analyzing two specific questions: 22 and 28 (Table B.1). The final SUS-score to learnability was 44.30, which is also between "poor" and "ok". It is calculated using the same procedure utilized for obtaining the SUS-score, but based only on the questions mentioned above.

After obtaining these results, we wanted to know which is the probability of new respondents answer the questionnaire with a score between "good" and "excellent", considering both satisfaction and learnability. To do that, we built a density plot of the SUS-scores of each respondent, considering satisfaction and learnability, so we could calculate the graphic's area within the desired SUS-score (i.e., 71.4 and 85.5). After that, based on our current data, we estimate that 16.22% of future respondents to our questionnaire may experience satisfaction between "good" and "excellent"; and, 10.13%

of future respondents may experience learnability between “good” and “excellent”.

Based on these findings and the user profiles presented in subsection 4.4.1, users are not satisfied nor think they can learn with the feedback provided by the process modeling tool they use.

4.4.3 Content analysis

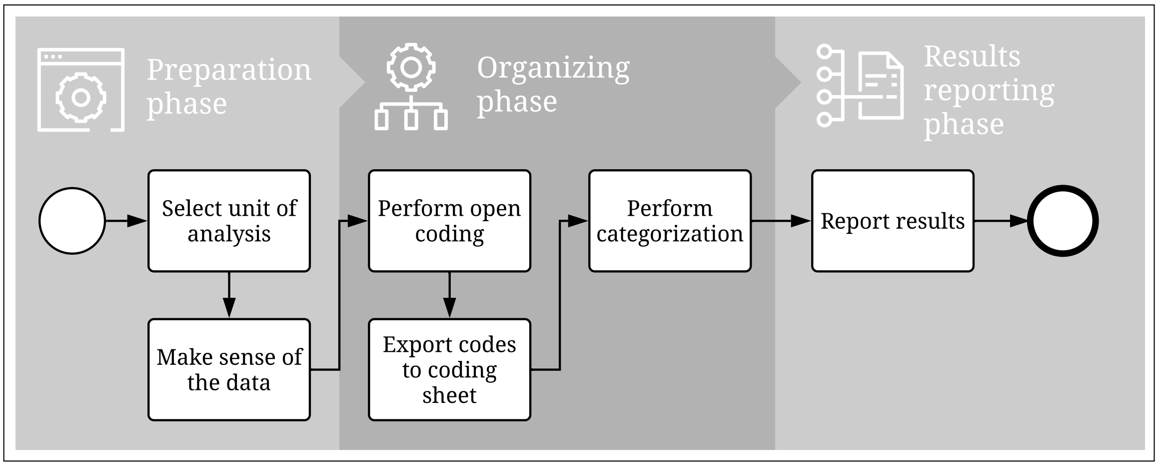
Content analysis (BRYMAN, 2013) is based on coding data, as a means for identifying and grouping similarities in the phenomena being analyzed. Elo and Kyngäs (ELO; KYNGÄS, 2008) differentiate the process of content analysis in deductive and inductive. The distinction between both is that, in the latter, the analyst does not have a previous knowledge about the phenomena being investigated. Therefore, the coding is generated out of the data being analyzed, which is our case: we conducted an inductive content analysis.

Saldaña (SALDAÑA, 2013) presents several approaches for coding data. Two examples of coding are: *holistic coding*, where researcher codes larger units of data to identify possible categories that could be developed; *hypothesis coding*, where the researcher tries to guess, previously to the data analysis, the codes s/he thinks to be able to find in the data. In this study, we used holistic coding. Moreover, since the coder relies on her/his own knowledge when coding, it is important that both the coder and the participants (from which the data was gathered) are from a common field (BRYMAN, 2013), which is the case in our study, since the author has a process modeling background, as the participants of the survey.

There are three main phases when performing a content analysis (Figure 4.11):

- *Preparation*: where the researcher *selects the unit of analysis* (e.g., word, sentence, paragraph, and so on) and *makes sense of the data* (i.e., by immersing in and becoming familiar with the data). We chose to work with sentences, since we knew the responses were short, in general;
- *Organizing*: where the researcher conducts the *open coding*, i.e., passes through the data being analyzed creating codes to represent different sets of units of analysis, *coding sheets*, i.e., organizes the codes into a spreadsheet, and *categorizing*, i.e., where similar codes are grouped together into categories with broader meanings;
- *Results reporting*: where the *results are reported* through categories identified in the process.

Figure 4.11: Inductive content analysis phases. This process was applied twice, one for the analysis of responses to question 17; another for question 18 of the questionnaire (Table B.1).



Since we had answers for two different open-ended questions to analyze, we conducted the content analysis process twice: for question 17; and, for question 18. After the “making sense of the data” stage we filtered respondents, ending up with 47 answers to be coded and categorized for question 17; and 46 answers to be coded and categorized for question 18. The answers removed were empty, too generic (e.g., one of the respondents answered simply “graphically” to question 17), or the respondent said s/he did not know how the subject, or deviated from the topic.

According to Bryman (BRYMAN, 2013), reviewing the codes is important so the researcher can see how well the data gathered fits in the codes and categories. Thus, after one week, we revised the open coding and categorization steps. Tables 4.3 and 4.4 summarize modelers’ demands gathered from questions 17 and 18, respectively, while Figures 4.12 and 4.13 represent the distribution of coded units of analysis (sentences from modelers’ responses) that comprise each category. The figures show that there are intersections between the categories in the sense that the same answer can contain different sentences that generated different codes which, in some cases, were grouped in one or more categories.

It is important to mention that, after the execution of the open coding step, we applied a questionnaire with three assessors: two researchers (one of them, an ex-member of our research group; the other, a researcher from the data mining area) and one professional (from the quality assurance area) specifically to confirm that our codes were coherent with the texts coded. In the questionnaire, for each code we presented the set of answers given by the respondents that were selected as representatives of the code (i.e., units of analysis). The assessor should answer to each extent s/he agreed with the coding, in a five-point

Table 4.3: List containing the set of process modelers' demands, represented by categories defined by the content analysis over the data gathered from the answers of the 47 respondents to the question: "*How do you think modeling tools should feedback on problems within process model?*".

Id	Category	Total supporters (%)
A	Presenting information about the problem and/or problem correction suggestions	32 (68.08%)
B	Highlighting problematic element or flow in diagram	25 (53.19%)
C	Exploring different ways to present the problems	12 (25.53%)
D	Validating	8 (17.02%)
E	Considering the level of experience of the modeler or the level of severity of the problem encountered	3 (6.38%)
F	Enabling interaction with the identified problem	3 (6.38%)
G	Preventing error	2 (4.26%)

Table 4.4: List containing the set of process modelers' demands, separated in categories, after the application of the content analysis process over the data gathered from the answers of 46 respondents to the question: "*What kind of information do you think modeling tools should present about problems within process model?*".

Id	Category	Total supporters (%)
P	Alternatives to correct the problem	22 (47.82%)
Q	Problem description	20 (43.48%)
R	Type of problem	18 (39.13%)

Likert scale from 1, "Totally disagree", to 5, "Totally agree".

After gathering all answers, we calculated the inter-rater reliability (JONATHAN; FENG; HOCHHEISER, 2017), which resulted in a mean of 79.71% of agreement, considered as satisfactory reliability, which indicates that our codes may be considered as representative of the data (at least to this sample of assessors). This result made us confident to categorize the codes and go throughout the whole content analysis process. Table 4.5 presents examples of units of analysis that generated codes, and their parent categories.

To complement the set of modelers' demands, we analyzed the data obtained from the interviews with specialists taking into account the findings from the application of the content analysis process as described. The interviewees are all highly experienced with the process modeling task, use BPMN as the notation and Bizagi as the process modeling tool. Only one of them reported also using Signavio, however, when further asked about this tool in the interview, admitted to only trying to use it once, as a "test". Moreover, regarding feedback about problems in process models, questions 13 to 16 (with no reference to the modeling tools they were based on), in general, all the interviewees are unsatisfied or very unsatisfied with Bizagi, Bonita and Camunda; and, with respect to Signavio, one of them

Figure 4.12: Distribution of coded units of analysis (sentences from modelers' responses) that comprise each category (see Table 4.3 for the categories discovered within the answers to question 17). Each intersection shows the total amount of different units of analysis that support each of the intersected categories. For example, the number ten in the intersection between categories A and B means ten different respondents' answers generated twenty different units of analysis coded and grouped into these two categories (ten of them support category A; the other ten support category B).

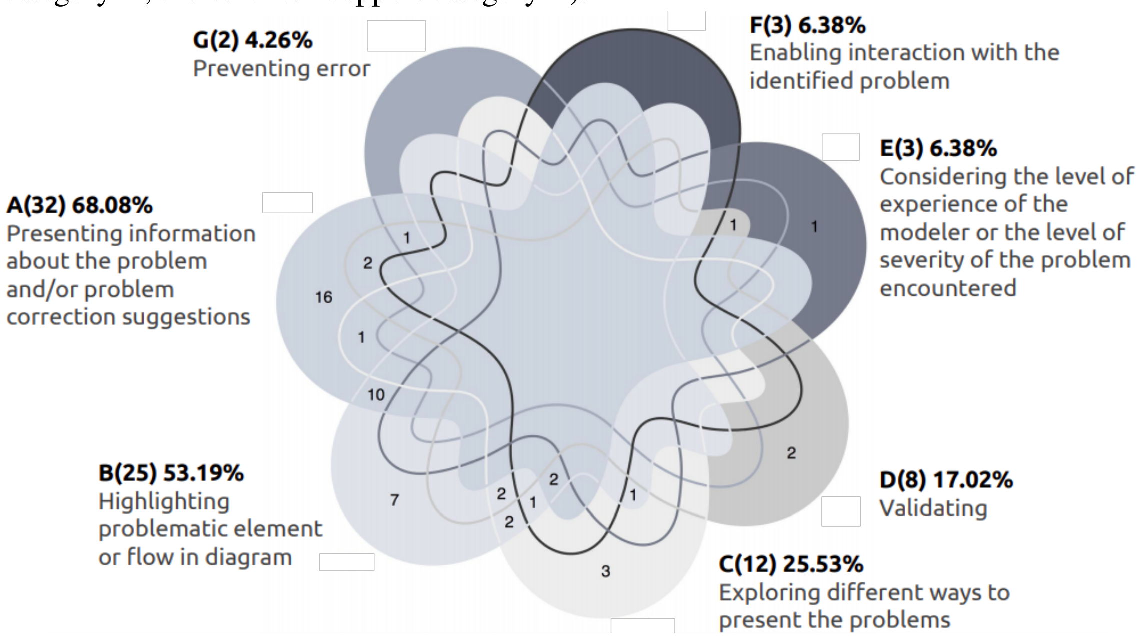


Figure 4.13: Distribution of coded units of analysis (sentences from modelers responses) that comprises each category (see Table 4.4 for the categories identified within the answers to question 18). Each intersection shows the total amount of different units of analysis that support each of the intersected categories. For example, the number two, in the intersection between categories B and C, means two different respondents' answers generated four different units of analysis coded and grouped into these two categories (two of them support category A; the other two support category B)

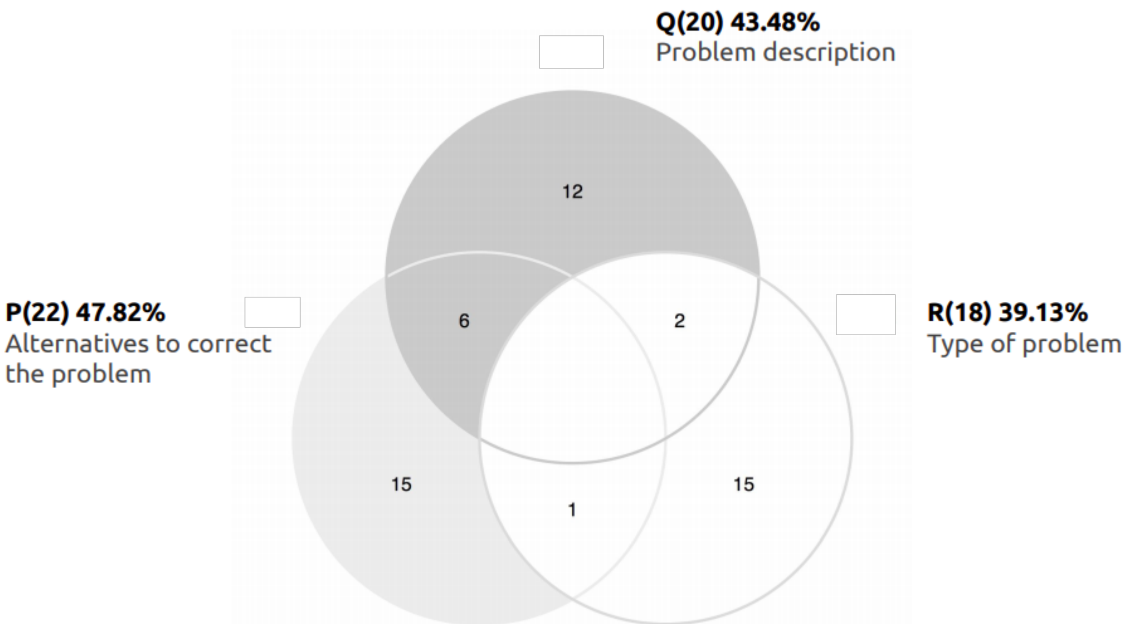


Table 4.5: List containing a subset of examples of process modelers' demands, separated in categories, codes, and units of analysis highlighted from the respondent answer that was coded. These examples are from the categorization applied to the answers to the question "*How do you think modeling tools should feedback on problems within process model?*".

Id	**Category** *Codes*, and *highlighted units of analysis from respondents' answers*
A	**Presenting information about the problem and/or problem correction suggestions** • *Code:* Presentation of details regarding the problem – "Describe the problem in detail, not in a generic way. Indicate the location of the model where the error is." – "Signaling the elements involved in the error (or close to it) and displaying a message describing the error" • *Code:* Presentation of examples to problem correction – "It should run a type of compiler, and where there was inconsistency there should have an exclamation point and, by placing the mouse over it, there should have a dialog box with a clear message about inconsistencies, and if possible with examples of correction." – "Indicating errors clearly and possible solution alternatives"
B	**Highlighting problematic element or flow in diagram** • *Code:* Highlighting problematic element or flow in diagram – "Highlighting in different colors in the drawing" – "Marking the element that has the problem and a message."

is unsatisfied, and the others are satisfied. These results are similar to the findings with the questionnaire analysis, as presented in subsection 4.4.1.

To explain the categories related to the question "*how do you think modeling tools should feedback on problems within process model?*" (Table 4.3), supported by the interviewees' answers, we divide the text into the following subsections, as suggested by Saldaña (SALDAÑA, 2013). Whenever necessary, we will refer to the interviewees as "I_x"; where "x" stands for a natural number from 1 to 5. When referring to, for example, the interviewees 1,3, and 4, we will simply refer to "$I_{1,3,4}$".

The categories related to the question "*what kind of information do you think modeling tools should present about problems within process model?*" are explained further: (P) *alternatives to correct the problem* correspond to modelers demanding modeling tools to provide feedback with textual information or examples of correct process models for the problem correction; (Q) *problem description*, modelers demanding modeling tools to properly describe why the problem occurred, and not simply which element has a problem;

and (R) *type of problem*, modelers demanding modeling tools to be able to typify the problems, according to different known modeling mistakes categorization (e.g., syntactical, pragmatic, semantic (SNOECK et al., 2015)).

4.4.3.1 Category A: Presenting information about problems and/or problem correction suggestions

This category refers to demands concerning to the presentation of textual information explaining why the problem occurred, and suggestions towards problem correction (either textual or through examples of process models that represent the correction).

$I_{1,2,4}$ emphasize that the modeling tools generally provide feedback about problems using too generic textual messages. I_4 says that textual feedback should explicitly explain why the problem occurred. I_1 also suggests that it would be interesting if the tool could provide information such as the impact the problem could cause to the organization if kept in a production environment.

The need for problem correction suggestions is pointed out by $I_{1,2,4,5}$. Only I_3 thinks that the problem correction should be strict to the modeler. I_1 also suggests a correction proposal that could be “accepted and automatically executed” by the tool.

4.4.3.2 Category B: Highlighting problematic elements or flows in diagram

This category refers to highlighting (for example, using different colors) the problematic elements or flows in a fragment of the process model.

Both $I_{2,3}$ say that modeling tools should highlight problematic element/flow with yellow or red (for warning and error hints, respectively). $I_{3,5}$ say the highlight should be not only for an element or a diagram flow, but also for process model fragments. Moreover, I_5 says that, considering s/he usually finds process models comprised by a large number of elements (i.e., tasks, events, and so on), the highlight should be through the usage of a circle around the problem or problematic area so the modeler, when visualizing an overview of the process model, could easily detect its problematic portions.

4.4.3.3 Category C: Exploring different ways to present the problem

This category refers to demands suggesting that the problems in a process model should be presented according to the region of the model that is predominant in the view (i.e., that the problems that are hidden by the limited viewport of the display do not get more attention than the ones that are closer to the modelers' eyes).

I_3 proposes that the problems should be categorized and represented by different icons. S/he says that initially the modeler would need to learn the iconography, which should be accompanied by a textual explanation in some part of the screen easily accessible to the modeler. However, after becoming more involved with the process, it would be faster to for the modeler to identify the types of problems in the model. I_2 also considers that icons should be used to be attached to problems detected in the process model. However, s/he expects a more simplistic manner: yellow icons to represent warnings, and red ones to represent errors.

I_4 thinks that the ideal would be that the problems were not only displayed in the model and problems log, but also through a list of categorized problems that would allow the user to approach and correct these problems according to the categories s/he feels most urged to correct first. These categories could correspond to problematic labeling or usage of message flow. A text informing the problem and where it occurred should be provided with a clickable "help" button, so the user, in case s/he did not understand what that problem refers to, could obtain more information about it. After clicking the "help" button, a pop-up or another tab should be opened, so the user could see more detailed information to assist in fixing the problem. This information could be in the form of text, image or examples of corrected diagrams.

4.4.3.4 Category D: Validating

This category is related to demands about the validation of the process model (i.e., the functions of detection and reporting of problems provided by the modeling tool) being automatic, manual, preemptive, or non-preemptive, for example.

I_1 believes both types of validation are important: the automatic one, that validates the model while the modeler is performing the process modeling task; and, the validation that is executed by the modeler whenever that s/he believes it is appropriate. While I_3 prefers the model to be validated only after the modeling task is completed, I_5 expects that the validation could occur along equal intervals of time, populating some problem log, without preempting the modeler.

4.4.3.5 Category E: Considering the level of experience of the modeler or the level of severity of the detected problem

This category refers to demands for modeling tools capable of giving feedback of problems tailored to the modeler level of experience or the severity of the problem detected, or both. For example, $I_{1,3}$ consider it is interesting to present problems found at two distinct levels: warning, for less severe problems such as not following a process modeling guideline (e.g., using more than 50 elements in a process model (MENDLING; REIJERS; Van Der Aalst, 2009), which is a pragmatic problem); and, error, for severe problems such as a syntactic ones (e.g., disconnected activities in one pool with a sequence flow (ROZMAN; POLANCIC; HORVAT, 2008)).

I_4 considers that the problems should be categorized in syntactic and pragmatic problems too. Then, modelers could decide which type of problem to handle first, out of a categorized list of problems found in the process model. Moreover, I_3 considers that novice users should initially be preempted always, when erroneously modeling a process. The interviewee believes that after detecting the same problem a few times, the tool should stop preempting the novice modeler for a while.

4.4.3.6 Category F: Enabling interaction with the identified problem

This category is related to the possibility of interacting with the problematic element and/or flow with the intention of obtaining further information about that problem. The modelers wish they could, for example, hover the problematic element, and a tooltip would show up directly on the model with more information about the problem. Or clicking the problematic element and/or flow and viewing this information in another window, containing examples of process models that represent the correction.

The same idea, a tooltip with textual information regarding the problem, is shared by both I_3 and I_2, but while I_3 would prefer hover as interaction, I_2 suggests a mouse click to trigger the tooltip.

I_4 believes that there should be a “help” button at the side of the problem detected, which would be presented in a list of categorized problems, so the modeler could click on this button and be sent to a tab or a page where s/he could get more details about the problem, and suggestions for correction, and so on. I_5, for example, considers that modeling tools should enable modelers to click on the problem detected and presented in a problems log, and be conducted to the area where the problem is in the model.

4.4.3.7 Category G: Preventing error

This category refers to demands concerning problems prevention. The modelers that support this category do not want the modeling tool to permit them to reproduce modeling problems in their process models. $I_{1,3}$ consider that it is important to prevent modelers from reproducing modeling problems in their modeling task. Although, I_3 considers this prevention should target novice modelers exclusively, as mentioned in subsection 4.4.3.5.

4.5 Research Reliability and Validity

According to Fink (FINK, 2003), the quality of a survey is related to diminishing threats to *reliability* and *validity*. Reliability is related to the reproduction of consistent results across different moments or by different researchers considering the application of the same instruments (e.g., survey). Validity is related to the extent to which the instruments used to gather data interfere in the data collected. Based on several authors (FINK, 2003; BOYCE; NEALE, 2006; ELO; KYNGÄS, 2008; BRYMAN, 2013), we took some cautions while conducting our survey, although we cannot guarantee removal of all threats to reliability and validity. For example, we conducted a pilot test on our questionnaire and interview processes and data collection to mitigate these threats. Moreover, we tried to be cautious when elaborating the questions composing the questionnaire and interview.

The questionnaire was widely shared with the community of BPM academics and professionals, and the people that participated in this stage of our research did so by their own choice. Also, they decided when to respond to the questionnaire (i.e., which time of the day it was most appropriate to them) and, even though the questionnaire were not short in terms of number of questions, it was thought to enable users to rapidly finish responding to all the questions (i.e., taking no more than 10 minutes to answer). The majority of the questions were closed-ended, with few alternatives to be chosen from, and semantically grouped into different categories. Closed-ended questions enhance the reliability since all respondents answer in terms of the same alternatives (FINK, 2002). All closed-ended questions, except for the Likert scale-based ones, had an option entitled "Other", which allowed respondents to add up an answer, in case s/he did not agree with any of the pre-defined alternatives.

The interview also had participants that, besides being invited, opted to contribute to this other stage of our research. Moreover, the interviewee was the ones that had the choices of at which day, hour and by which means (i.e., Skype, WhatsApp call, phone call, face-to-face) the interview would take place. The interviews took no more than 15 minutes each, considering the introduction and conclusion parts.

We were particularly concerned with reducing the threats to reliability with the experts interviewees. To achieve this, for example, we managed to perform only one interview per day, aiming at avoiding the interviewer getting tired and hampering the appropriate answering of our questions. Besides that, in the data analysis phase, especially in the content analysis one, we coded no more than ten answers a day, and reviewed the codes and categorizations as a whole after a week, to identify any possible misinterpretation of the open-ended questions in both interview and questionnaire stages.

4.6 Final comments

In this chapter we presented the process we adopted for conducting our survey, and we reported and discussed our findings. The results gave us a panorama of the process modelers we sampled. They are between 21 to 38 years old, process analysts, students or managers, know business process modeling and the BPMN notation for four years or more. Their knowledge about the field came mainly from the academic sector, and consider their process modeling experience to be moderate.

Our participants reported to be unsatisfied with the manner the tools Bizagi, Bonita, and Camunda provide feedback about problems in process models, and satisfied with Signavio. These results were obtained by presenting to the modelers 4 different ways of providing feedback about problems based on these tools, but without explicitly displaying which were the tools we used for building the examples. Regarding satisfaction and learnability, extracted from the answers to a SUS questionnaire, we obtained SUS-scores 50.74 and 44.30, respectively. Which is in a gap between "poor" and "ok".

Finally, the content analysis of the answers to the open-ended questions (questions 17 and 18) of the questionnaire), complemented with the analysis of the information gathered with the interview, allowed us to identify a set of 10 different categories, which represent the process modelers' demands regarding how the feedback should be provided by the tools, and what kind of information they should present. For example, they wish they could have examples and/or instructions helping them to correct the detected problems, and not only textual information. In the next chapter, we present a mapping of the modelers' demands to the set of modeling tools used in this work and the literature.

5 CASE STUDIES

In this chapter, we present two case studies and a mapping of the previously identified demands to the feedback features that the process modeling tools provide, and also to the solutions the literature describes regarding feedback about problems in process models.

The two case studies showed how the BPMN-based process modeling tools deal with different modeling problems. The first case study (Sect. 5.1) analyzed the behavior of the four modeling tools (Bizagi, Bonita, Camunda, and Signavio) regarding the features they provide for displaying modeling problems. For the second one (Sect.5.2, we conducted a multiple case study following an embedded approach (ROBSON; MCCARTAN, 2016), where we analyze one characteristic (*how modeling tools provide feedback about problems in process model?*) across multiple cases (the modeling tools). Then, we took the results from our SLR (Chapter 3, Sect.5.3) regarding solutions presented in the literature concerning visual feedback about problems in process models, and we built a matrix (Sect. 5.4 to compare the findings from the case studies with those identified in our SLR, and the modelers' demands obtained through our survey, reported in chapter 4.

5.1 Analysis of the Behaviour of BPMN-based Process Modeling Tools Concerning Feedback about Problems in Process Models

In the study conducted by Dias (DIAS, 2018) [1] , the author selected four BPMN-based process modeling tools. The selection was the result from a literature search considering a set of predefined inclusion criteria, such as: the tools should be open-source or provide a trial version, should be recognized as a reference both in academia and industry regarding BPM, have recent updates, among others. A study that helped the selection process was from Snoeck et al. (SNOECK et al., 2015), which analyzed 117 process modeling tools exhaustively. As a result, the four BPMN-based process modeling tools selected to be analyzed as well as used in our study were: Bizagi, Bonita, Camunda, and Signavio.

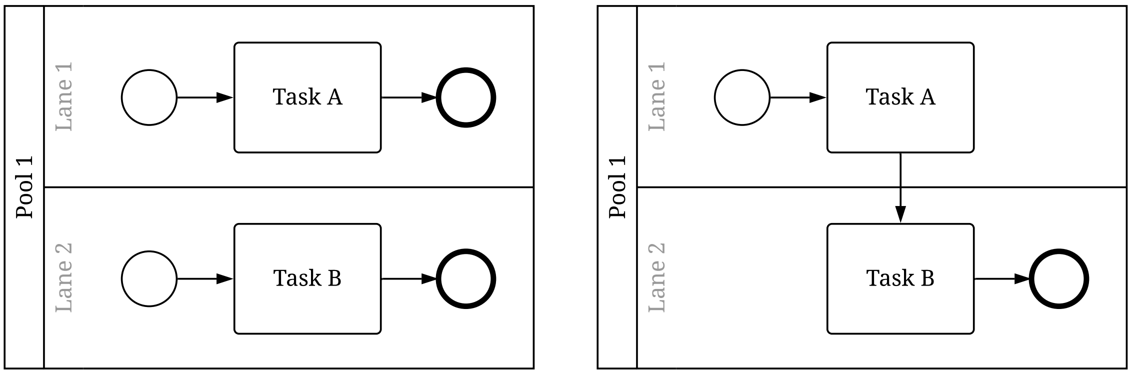
Furthermore, a set of common process modeling problems (also known as *anti-patterns*) was selected and used to identify differences of behavior between the four different selected process modeling tools. Figure 5.1 presents an example of anti-pattern

[1] Study advised by Lucinéia Heloisa Thom, and co-advised by Vinicius Stein Dani.

used by Dias (DIAS, 2018). It is worthy of mentioning that a characteristic of an anti-pattern is that, usually, it comes with a proposed solution.

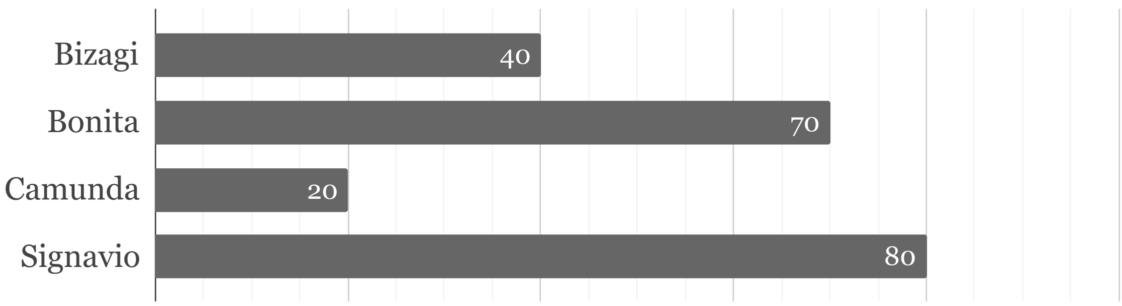
Figure 5.1: Example based on Rozman et al. (ROZMAN; POLANCIC; HORVAT, 2008) of one of the anti-patterns, and its proposed solution, as selected by Dias (DIAS, 2018). This anti-pattern is regarding to "*each swimlane in the pool containing one start event*".

(a) Anti-pattern. (b) Proposed solution



After defining and following a tool analysis method, a summary of the results obtained by Dias (DIAS, 2018) is shown in Figure 5.2. Signavio provided an warning or error feedback in 80% of the cases, followed by Bonita (70%), Bizagi (40%), and Camunda (20%). Other findings presented by the author is that the modeling tools analyzed do not present the same behaviour regarding the same problems in a process modeled in BPMN 2.0. Since this version of BPMN is an ISO standard, the tools should show the same problems in a similar way, but usually they do not do so.

Figure 5.2: Extent to which modeling tools provide warning or error feedback about problems in process models (DIAS, 2018).



5.2 Analysis of BPMN-based Process Modeling Tools Concerning the Occurrence of Problems in Process Models

We wanted to know if the anti-patterns analyzed by Dias (DIAS, 2018) still occur among modelers. To obtain this information, we performed a case study using a subset of 115 process models from the dataset [2] provided by Camunda through its “BPMN for Research” initiative. These models were built by professionals and students in response to proposed exercises in training sessions provided by this company. In these exercises, modelers were provided with a textual description of a process, which they should use to create the model.

We chose two exercises from this dataset. The dataset provides: the textual description of the exercise, the models built by the participants in their training sessions, and a proposed solution to the exercise (which the participants were not aware of). To analyze the models we opened each one in the four modeling tools: Bizagi, Bonita, Camunda, and Signavio.

The results show that:

- out of 66 models analyzed for exercise 1, 19 (i.e., 28.79%) reproduced at least one anti-pattern;
- out of 49 models for exercise 2, 20 (i.e., 40.82%) reproduced at least one anti-pattern.

Although it may seem not significant, 28.79% indicate that, at least one, out of every four modelers, fall into reproducing anti-patterns. That is, there still is a non-neglectable number of modelers making mistakes considered common in the literature. In other words, the modelers could be better informed and directed by the modeling tools in terms of avoiding the reproduction of modeling problems (i.e., use of anti-patterns).

5.3 Analysis of the Literature Regarding Visual Feedback about Problems in Process Models

As one of the results of our SLR, we found four different studies exploring different manners to give feedback about problems in process models (AWAD; WESKE, 2010; LAUE; AWAD, 2011; WITT et al., 2015; CORRADINI et al., 2017). In the study by Laue and Awad (LAUE; AWAD, 2011), as previously mentioned, they propose the attachment

[2] https://github.com/camunda/bpmn-for-research

of a graphic symbol, more precisely a white “x” surrounded by a red circle (indicated by the dotted circle in the figure), to the process model‘s element that generated the process modeling problem. Besides that, a textual description about the problem is triggered by hover action. This description of the problem, however, is rather simplistic.

The process model’s elements that are part of the detected process modeling problems are highlighted in red in other proposals (AWAD; WESKE, 2010; CORRADINI et al., 2017). Moreover, Corradini et al. (CORRADINI et al., 2017) present a list of guidelines violated in a process model and allows the user to select each of them separately to see the process model’s fragment that is related to the violated guideline highlighted in red. Finally, Corradini et al. (CORRADINI et al., 2017) also suggest the display of an example of process model correction, by displaying the problematic fragment of the model side-by-side with a counter-example in the expected corrected way.

5.4 Mapping Process Modelers Demands to What the Literature and the Process Modeling Tools Provide Regarding Feedback about Problems in Process Models

In this section, we present a comparison matrix, where we map which of the modelers’ demands, identified through the survey, are supplied by the literature reviewed and by each modeling tool addressed in our work.

Table 5.1: Comparison matrix of the process modelers’ demands and the manner the literature and different BPMN-based process modeling tools provide feedback about problems in process models. The categories represent the findings presented in section 4.4.3. Whereas the tool and/or the literature totally support the demand, we mark the respective table cell with “+”; if the tool and/or the literature totally do not support the demand, we mark with “–”; otherwise, we mark it with “+/–”.

		Tools				
Id	**Category**	**Bizagi**	**Bonita**	**Camunda**	**Signavio**	**Literature**
	Textual explanatory information about the problem	+	+	–	+	+
	Correction suggestions through text	–	–	–	+	–
	Correction suggestions through process model fragment representing the correction	–	–	–	+	+
A	**Presenting information about the problem and/or problem correction suggestions**	**+/–**	**+/–**	**–**	**+**	**+/–**

Table 5.1 continued from previous page

Id	Category	Tools: Bizagi	Bonita	Camunda	Signavio	Literature
	Highlight through the use of different colors	–	–	+	+	+
	Highlight through the use of different icons	–	–	–	+	–
	Highlight through the use of shapes around the problematic element, flow, fragment	–	–	–	–	–
	Highlight of problematic element	–	–	+	–	+
	Highlight of problematic flow	–	–	–	–	+
	Highlight of problematic fragment	–	–	–	–	–
B	**Highlighting problematic element or flow in diagram**	**–**	**–**	**+/–**	**+/–**	**+/–**
	Prioritize feedback about problems found in the problematic area in focus (i.e., being visualized by the modeler)	–	–	–	–	–
	Use of different icons to display warnings and errors	–	–	–	+	–
	Use of list divided by categories of problems	–	–	–	–	–
	Use of external page to present more information about the problem	–	–	–	+	–
C	**Exploring different ways to present the problems**	**–**	**–**	**–**	**+/–**	**–**
	Modeler may activate automatic validation	–	–	–	–	–
	Preemptive feedback	–	–	–	–	–
	Non-preemptive feedback	+	+	+	+	–
D	**Validating**	**+/–**	**+/–**	**+/–**	**+/–**	**–**
	Feedback about problems according to modeler level of experience	–	–	–	–	–
	Feedback about problems according to the severity of the problem	–	–	–	+	–
E	**Considering the level of experience of the modeler or the level of severity of the detected problem**	**–**	**–**	**–**	**+/–**	**–**

Table 5.1 continued from previous page

Id	Category	Tools: Bizagi	Bonita	Camunda	Signavio	Literature
	Explore by hovering the problematic element, flow, or fragment to present further information	–	–	–	+	+
	Explore by clicking the problematic element/flow/fragment to present further information	–	–	–	+	–
	Explore by displaying a "Help" or "More information" button to present further information about the detected problem	–	–	–	+	+
F	**Enabling interaction with the identified problem**	**–**	**–**	**–**	**+**	**+/–**
	Do not permit problems to be reproduced *It is noteworthy that all columns received "+/–" because the tools and the literature try to minimize the reproducibility of, at least, syntax errors to some extent*	+/–	+/–	+/–	+/–	+/–
G	**Preventing error**	**+/–**	**+/–**	**+/–**	**+/–**	**+/–**
	Through presentation of textual information	–	–	–	+	–
	Through presentation of examples of correct process models	–	–	–	+	+
P	**Alternatives to correct the problem**	**–**	**–**	**–**	**+**	**+/–**
	Properly describe why the problem was occurred	+	+	–	+	+
Q	**Problem description**	**+**	**+**	**–**	**+**	**+**
	Typify the problem (e.g., syntactic, pragmatic, semantic, or any other typification that may categorize groups of similar problems)	–	–	–	+	–
R	**Type of problem**	**–**	**–**	**–**	**+**	**–**

5.5 Final comments

In this chapter, we presented two case studies we conducted and a matrix where we map the process modelers demands to the literature and a set of process modeling tools, regarding feedback about problems in process models.

We observe that, as the majority of the respondents are mainly users of Bizagi, the participants, in general, seem to have raised demands regarding their experience with this tool. This does not mean that other tools nor the literature have already solved a certain demand, without having come to the knowledge of the modeler. In this sense, one of the contributions of our work is to allow the reader to identify which category represents her/his needs, and easily identify which modeling tool fits best to the case in hand.

Table 5.1 shows that literature is behind the modeling tools in terms of supporting modelers' demands. Signavio is the tool that fits the most with the modelers' demands, fully supporting 2 out of 7 of the demands concerning how the feedback about modeling problems should be, and fully supporting all the demands regarding which information should be displayed within the feedback.

Moreover, we show that the recurrent problems in the process modeling task identified and characterized as anti-patterns back in 2008 still occur nowadays, and that the modeling tools differ when it comes to how they give feedback about problems in process models, even when we analyzed the same problem across different tools.

6 RECOMMENDATIONS ON HOW TO PRESENT FEEDBACK ABOUT PROBLEMS IN PROCESS MODELS

In this chapter, based on all our main findings, we propose a set of recommendations to present feedback about modeling problems during the process modeling task. The main source for these recommendations is the comparison of how the literature and the four selected BPMN-based process modeling tools fulfill modelers' demands, as presented in Table 5.1. More than aiming at fulfilling these modelers' demands, we focus on providing recommendations not linked to any particular tool. The intention with this is twofold: (i) the providers of modeling tools could use our recommendations to provide a more standardized way to feedback modelers about problems in process models; (ii) our recommendations could serve as a starting point for the implementation of a standalone complement for the process modeler, allowing a separate verification of a process model being built.

We present our recommendations using different usage scenarios regarding modelers' demands. To exemplify our recommendations, and based on (LEOPOLD; MENDLING; GÜNTHER, 2016; Object Management Group, 2015), we consider a task with no label as a warning; and, a task with no incoming and/or outgoing sequence flow as an error.

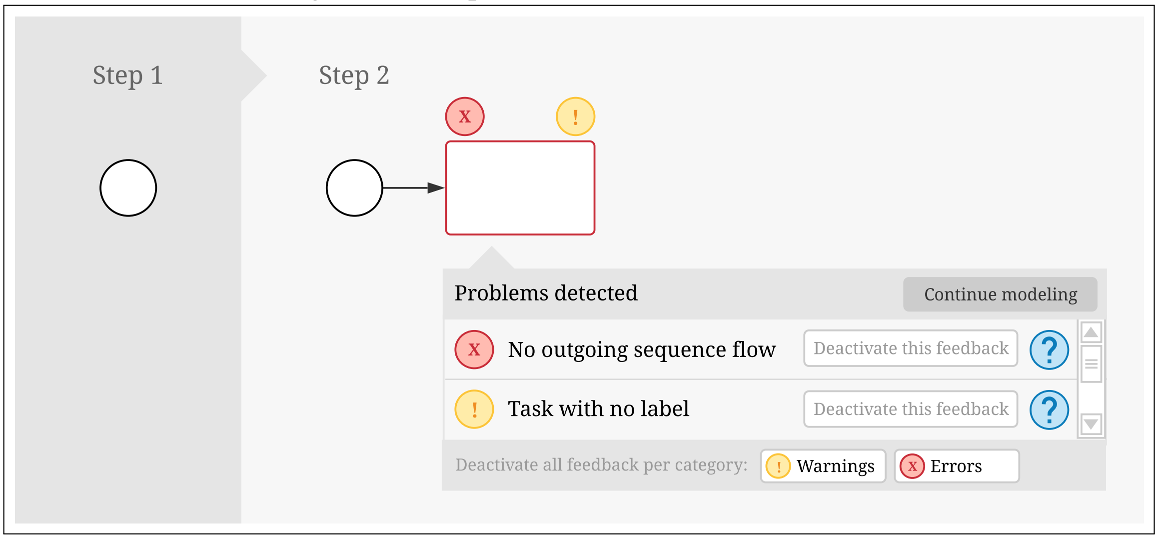
We based only the general lines of our proposed interfaces in the set of modeling tools analyzed. We reused what already is common sense along with these tools (i.e., the menu at the top of the screen, toolbar at the left, modeling area at the right, further information at the bottom) and added elements intending to fulfill the not yet supported (or fully supported) modelers' demands. Also, we adopted the colors yellow and red to represent warning and error messages, respectively, in our recommendations, not only based on modelers' demands but also because it is a widely used color convention (GROSSMAN, 1992). Moreover, both the icons representing warning and error are circled. Furthermore, we divide the problems into two "categories": warnings and errors.

6.1 Scenario 1: Initialization of the Modeler Tool

The modeler is asked which is her/his level of experience in process modeling. Depending on the answer, the feedback will be accordingly. In the case of s/he chooses "very low" or "low", the validation will be set to *automatic* and *preemptive*. And, at any moment, the modeler may choose not to be preempted any more when incurring in a category of problems or any specific problem. Moreover, the modeler may be able to switch back to preemptive feedback at any moment. This scenario was built based on the demands represented by Categories D (see Subsection 4.4.3.4), E (see Subsection 4.4.3.5), and G (see Subsection 4.4.3.7).

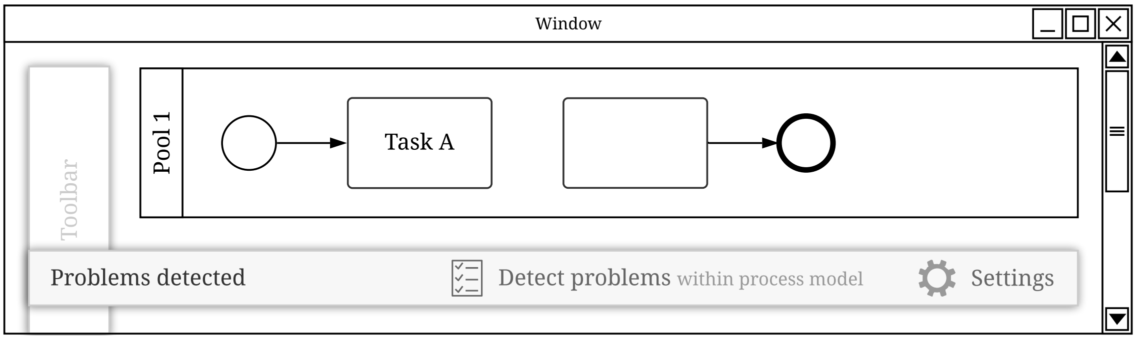
Figure 6.1 presents an example of how the *automatic* and *preemptive* feedback would behave, considering a modeling task composed of two steps executed sequentially. In the first step, the modeler places a start event in the modeling area; in the second step, s/he places a task, as a sequence of the start event placed in the first step, and the preemption occurs (even before the modeler defines a label to the task). A warning and error feedback is presented to the modeler, in the form of a tooltip attached to the element that generated the problem, demanding action from the modeler. Thus, s/he may decide to: (i) continue modeling; (ii) visualize further information about the problem (such as why the feedback was given and correction suggestions); (iii) deactivate the feedback concerning such kind of problem; (iv) or, deactivate the feedback concerning the category of problems in which that problem relies on (in this case, s/he may choose to silence feedback on warnings and/or errors, for example).

Figure 6.1: Two step example of the *preemptive automatic* feedback. In "Step 1", the modeler places a start event; in "Step 2", the modeler is required to take an action through the tooltip displayed right after he placed the task. This tooltip presents the warning and the error detected in the model at this stage. And, one of the actions the modeler may take is to "Continue modeling", for example.



On the other hand, in the case of s/he chooses any other level of experience (i.e., "moderate", "high", or "very high"), the feedback will be set to *automatic* and *non-preemptive*. In this case, the highlighting of problematic elements inserted in the model and the populating of the list of problems detected continue to occur, but the tooltip demanding action of the modeler ceases to appear. Independently of the initial setting, the modeler should be able to reset her/his level of experience and/or change the tool behavior regarding feedback. Note that the modeler should be able to choose the feedback to be *manual* and *non-preemptive*, in complement to the other two possible settings presented before. In this case, the categorized list of problems should be populated, and the highlighting of the problematic element, flow and/or fragment should proceed only after the modeler explicitly demands (by clicking "Detect problems within process model"). Figure 6.2 shows an example of a process modeled with problems in a manual feedback setting. In this case, no highlighting occurs, nor the problems detected list is populated. Moreover, the button "Detect problems within process model" is kept visible in the feedback log window.

Figure 6.2: Example of the *non-preemptive manual* feedback setting. The modeler, at the modeling stage represented by this example, have not clicked yet on the "Detect problems within process model" button.

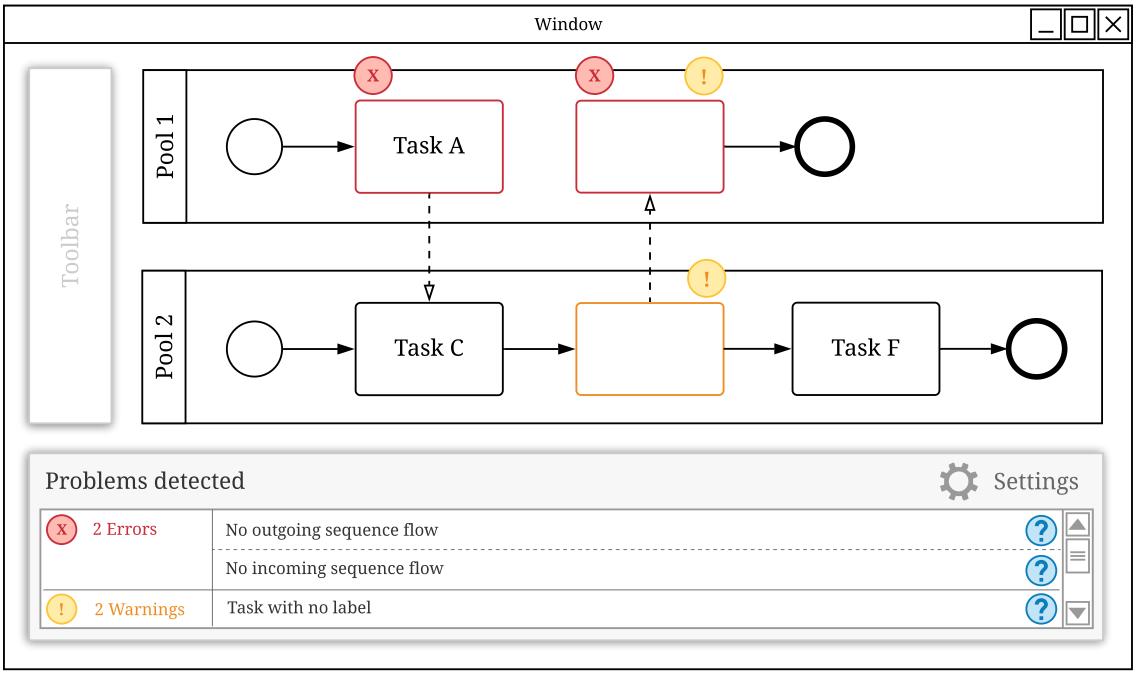


The "Settings" button is a simple form that enables the modeler to decide the feedback setting by changing her/his level of experience or by directly switching between automatic and manual feedback. Finally, in the case of automatic feedback being selected, the modeler can also decide between preemptive and non-preemptive feedback.

6.2 Scenario 2: Modeler Encounters a Set of Problems in a small Process Model

In this scenario, the problems are displayed in a list of problems ordered by categories, which is presented in a feedback log, below the modeling area. As can be seen in Figure 6.3, the process model is fully visible in the process modeling area. This and the next scenarios consider a *non-preemptive automatic* feedback. This scenario was built based on the demands represented by Categories C (see Subsection 4.4.3.3) and F (see Subsection 4.4.3.6).

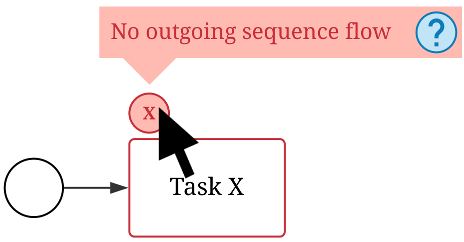
Figure 6.3: Example of the process model fully visible in the modeling area, in a *non-preemptive automatic* feedback scenario. The feedback log (presented at the bottom) accumulates the problems detected in the process model as a list, without demanding action from the modeler while s/he models the process.



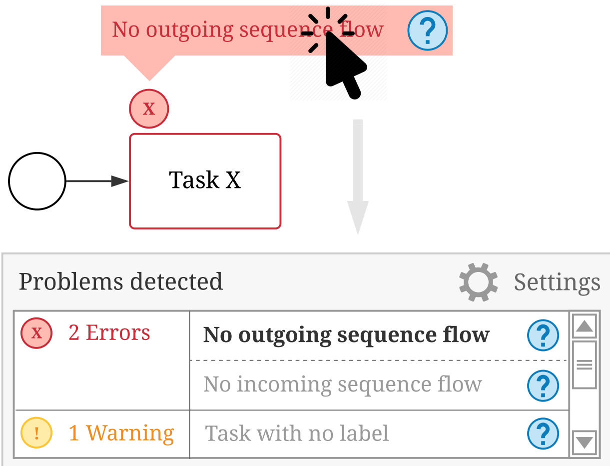
The modeler may hover any of the icons to read which is the problem detected (Figure 6.4a). S/he can also click on any icon to view where is the problem detected in the feedback list (Figure 6.4b). And, finally, the modeler may also double click on the icon or click on a detected problem in the feedback list to view further information and correction suggestions about it. These last interactions will be detailed in the next scenario.

Figure 6.4: Modelers actions over the icons representing detected problems. (a) Mouse over, and (b) Click.

(a) The modeler visualizes information about the problem through a tooltip after hovering the icon.



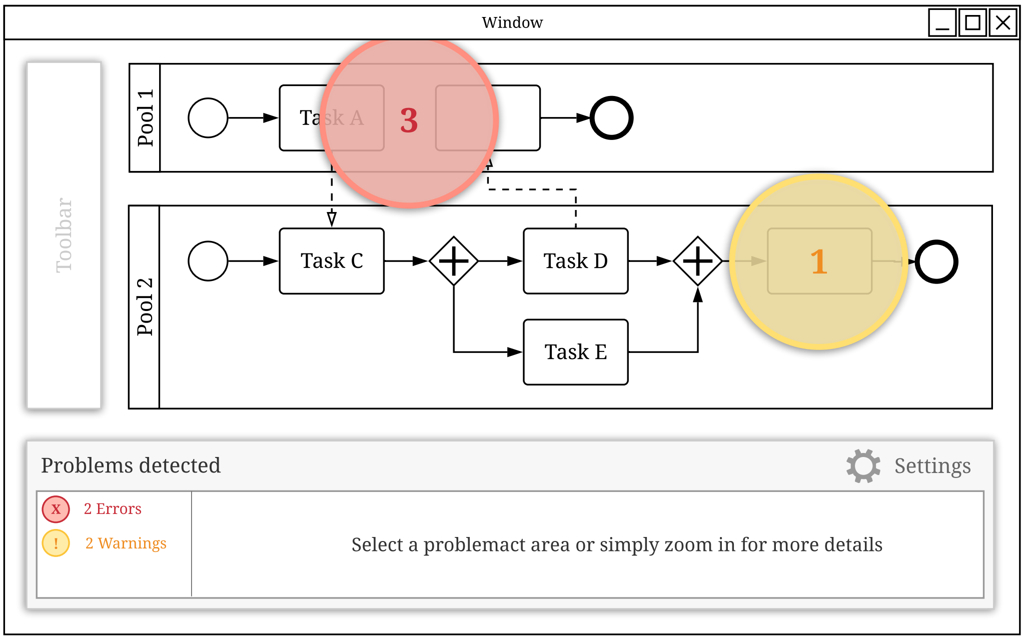
(b) The modeler visualizes where the information about the problem that appeared after the hovering is, in the feedback list, after clicking on it.



6.3 Scenario 3: Modeler Encounters a Set of Problems in a large Process Model

This usage scenario exemplify a situation of modeling a large process model, which cannot be adequately visualized in the process modeling area. In this scenario, we explore two possibilities. The first one concerns the zoom out of the model (Figure 6.5); in this case, the modeler sees yellow or red circles, depending on the type of the problems, within the problematic areas highlighted. The modeler can click on any circle to zoom in that portion of the process model, gaining information about its context. This scenario was built based on the demands represented by Category B (see Subsection 4.4.3.2), C (see Subsection 4.4.3.3), and F (see Subsection 4.4.3.6).

Figure 6.5: Overview of a large process model. When zoom is out, and the model is fully visible in the modeling area, the modeler visualizes problematic areas of the model highlighted. In this example, there are two problematic areas: one with at least one error; the other, with only warnings (in this case, precisely one).



After this action, the modeler is presented with the second possibility, which concerns the zoomed portion of the model (Figure 6.6). In this case, the modeler sees basically what s/he would see in Scenario 2, presented before, with the following addition: arrows pointing to directions where there are other problems in the model. These arrows contain numbers so the modeler can decide to attack the least or the most problematic area. Moreover, these arrows are colored in yellow or red, too. In this case, there is one warning to the left. The feedback list, displayed in Figure 6.6, presents information about the problems that are visible in the modeling area, in a similar way that Google Docs and Microsoft Word present the most used Text Fonts to the user: on top of the list containing the full set of fonts. In our case, the list presents the full set of detected problems, organized per categories, with the problems that are visible in the modeling area located at the top of this list.

Figure 6.6: Zoom in the large process model already presented in Figure 6.5. In this case, the modeler zoomed in the red portion of the model, which contains two errors and one warning. The problems in focus are listed on top of the feedback list, at the bottom of the interface.

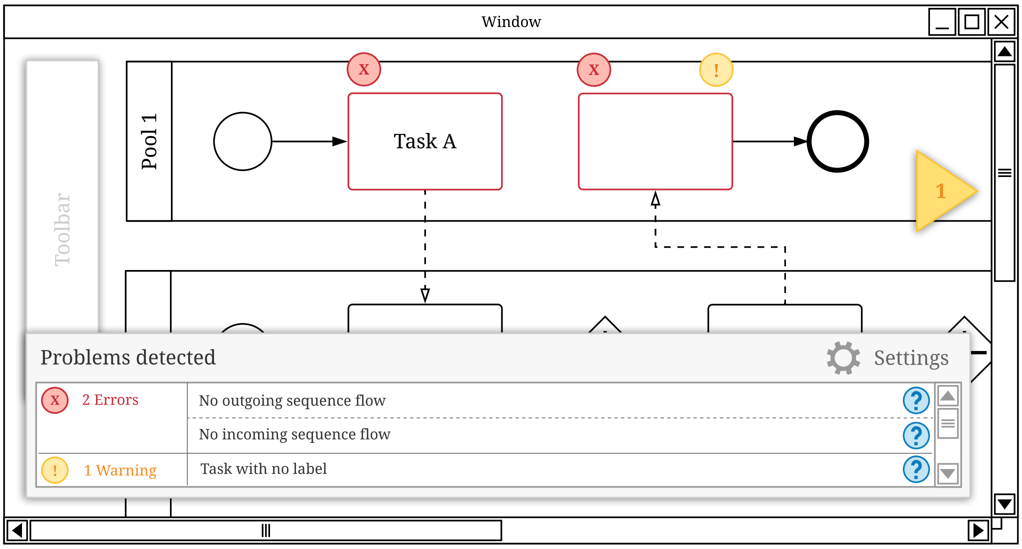
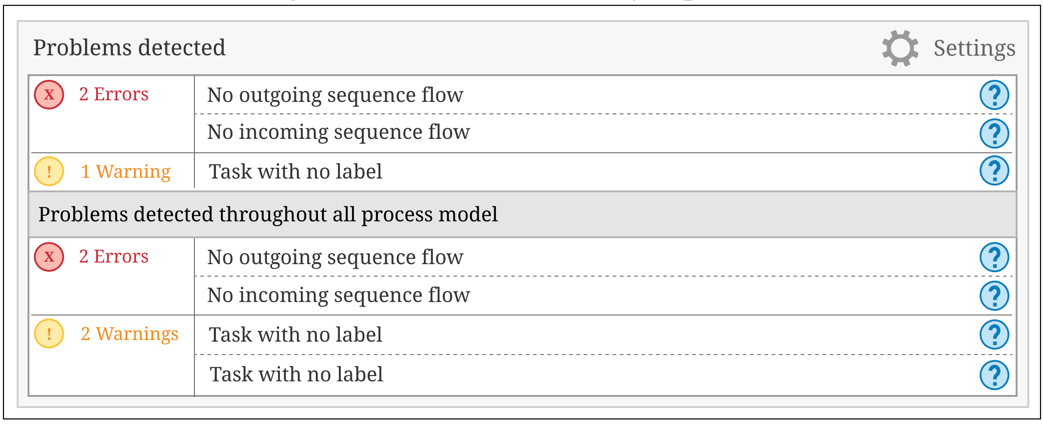


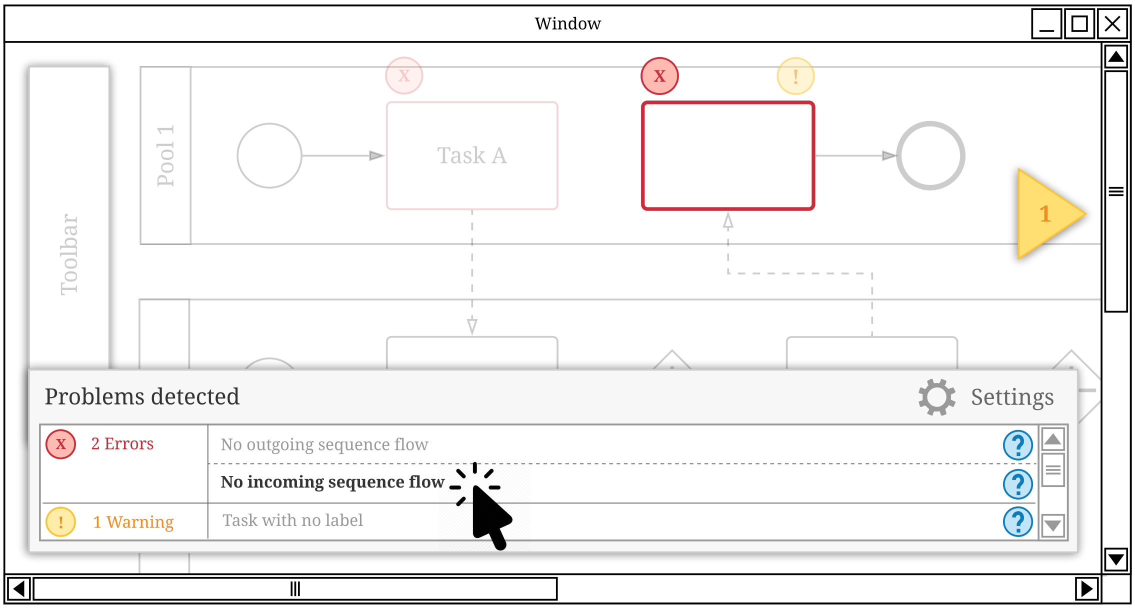
Figure 6.7 presents an example of how the full list of problems should be presented if it could be fully expanded and viewed at once.

Figure 6.7: Feedback list fully expanded.



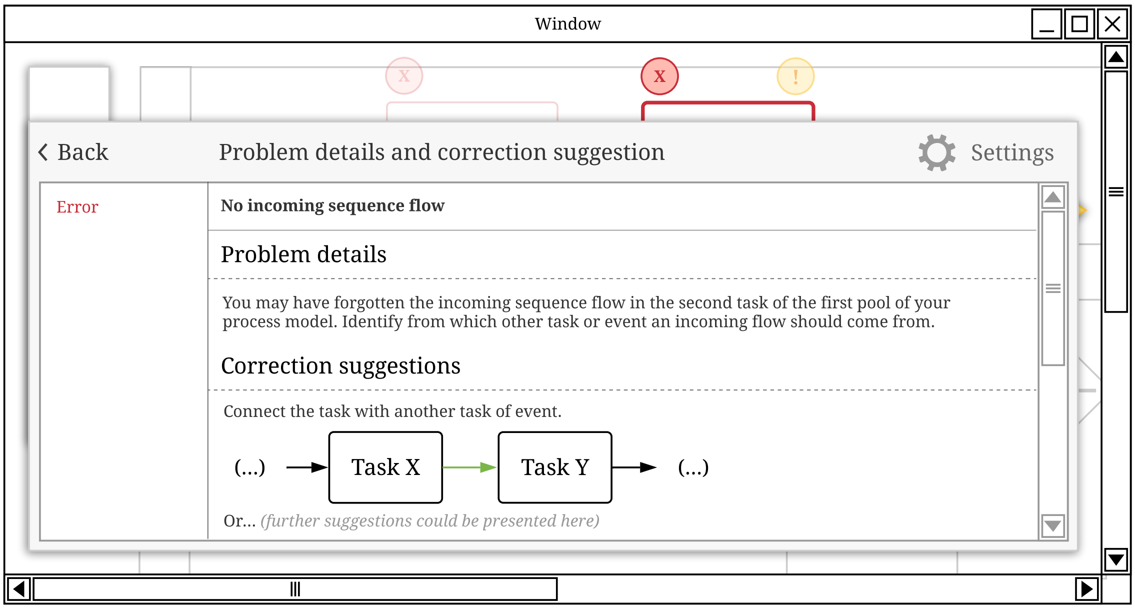
When the modeler hover a detected problem in the feedback list, that line is highlighted and the whole model opacity is decreased, except for the one relative to the hovered item in the list (Figure 6.8). This recommendation may be particularly complicated to implement, in the case of the problematic area being composed by different (and distant) fragments of the process model (e.g., in the case of the existence of a xor-split without a xor-join).

Figure 6.8: Mouse over action on any problem in the feedback list will cause the model to be transparent, except for the problem being hovered.



After clicking the hovered problem, a view containing details about the problem and correction suggestions are shown (Figure 6.9). We suggest this information, more precisely the correction recommendation, to be presented within the modeling tool expecting that, in the future, the modeling tools' providers may deliver features to suggest the auto-correction of process models, for example.

Figure 6.9: Example of detailed information and correction suggestion for one detected problem.



6.4 Final comments

In this chapter we presented recommendations to provide visual feedback to modelers about problems in process models. To compose the set of recommendations, we considered the modelers' demands obtained through our survey (see Chapter 4) and the results of the mapping presented in Section 5.4. Considering the demands that were not fully supported by at least one of the tools or the literature, we described three scenarios to present our recommendations.

7 CONCLUSIONS

A well-modeled process can be implemented to generate significant results for organizations, which can be savings or impact in society, for example. On the other hand, modeling problems can induce wrong implementations that generate execution errors, creating extra costs and unsuccess.

As was observed in our second case study, modelers keep making the same mistakes during the process modeling task (e.g., missing the proper use of sequence and/or message flows, misusing timer events, and so on), even though BPMN 2.0 is a standard which many well-known process modeling tools comply to. Moreover, there is no standard on the way each tool provide feedback to the modelers about the frequently reproduced mistakes.

To the best of our knowledge, this is the first study investigating how satisfied are the modelers specifically with the manner the modeling tool they use provide feedback to them and the first to investigate what are the modelers demands on feedback about problems in process models. We also present the first set of recommendations towards supporting modelers' demands.

After conducting a systematic literature review on visualization of business process models, we identified there were few studies on feedback about problems in process models. This corroborated our investigation on modelers' satisfaction and learnability regarding the feedback provided by the modeling tools they use, and the investigation on what are their demands regarding such feedback. We applied a questionnaire to 57 participants, and interviewed five experts (which also responded to the questionnaire). The analysis of their responses to two open-ended questions enabled us to identify 7 categories of demands on how the feedback should be provided; and 3 categories of demands on which kind of information should be provided. We further complemented these categories with the results obtained from the interviews. As a result, we were able to map the modelers' demands to what the literature and the BPMN-based process modeling tools provide regarding feedback about modeling problems. With such mapping, we found that 5 out of 7 demands were not yet fully addressed in the literature nor by the four selected modeling tools, which led us towards proposing a set of recommendations to fulfill these gaps.

Moreover, we confirmed that the modeling problems identified by Rozman et al (ROZMAN; POLANCIC; HORVAT, 2008) still occur; we point out that the standard BPMN 2.0, to which many modeling tools comply with, do not include elements for providing feedback about problems in process models. Then, we advanced towards the definition of a set of recommendations on how to visually represent the problems in process models in a way coherent to the process modelers' demands.

Finally, we believe that this study, which is in a boundary zone between BPM and information visualization, is valuable to the academic and industry communities, specially for professionals and students initiating their studies on business process modeling. Providing adequate feedback regarding problems in process models can facilitate learnability and reduce modeling problems occurrence.

7.1 Limitations

As a possible limitation of this work, we consider that the size of the questionnaire may have introduced the variable "fatigue" in the respondents, which may have caused the last questions to be answered with less attention or dedication. As a possible suggestion to improve this, one could carry out the application of the questionnaire in two stages, for example.Nevertheless, we cannot guarantee the respondents answers were reliable and valid, although we have taken precautions (e.g., pilot testing, questionnaire with majority of questions in a closed-ended form and with few pre-defined alternatives to be chosen from) to mitigate these possible threats.

Another limitation is that only one researcher performed the open-coding and the categorization of the modelers' demands. Although we have conducted an analysis of the agreement of other researchers to the coding and the units of analysis defined in our work, it would be better if two or more researchers had conducted the open-coding process separately for later comparison and discussion. Finally, although the set of modeling tools we analyzed are the most used ones, it can also be considered as a limitation, since it could be larger.

7.2 Future works

Considering our SLR, one of the results we obtained is that 65.12% of the papers report features regarding user-interaction on their approaches. We believe there could be opportunities for future research on this topic. Besides, considering that only 16.28% of the selected studies explore ways of displaying information about process models linked to the process model itself, we believe there is space for further exploration here too. Moreover, regarding our SLR as a whole, it would be interesting to broaden its scope by investigating visualization of conceptual models as well. Finally, we desire to deepen our analysis of the selected papers to point out similarities among studies across categories, and identify explicitly what other notations may still contribute to BPMN.

Regarding our survey, we intend to use its results to elaborate and carry out a similar one, in English, to widen the sample of participants. We also intend to replicate the one presented herein in a near future to assess its reliability, and identify if our predictions concerning future respondents' satisfaction and learnability will be confirmed. Moreover, we intend to have at least two researchers on the open-coding phase. As for the case studies, we wish to complement the set of modeling tools and anti-patterns analyzed, and also have two or more researchers conducting the case studies. Broadening the set of tools being analyzed across all stages of the study is also an interesting future work.

Another interesting future work is to implement a fully functional prototype of our recommendations, following a user-centered design approach, which could be evaluated by modelers with different levels of experience. Also, we could reproduce our survey, preferably with the same set of respondents, with the intention of identifying if our recommendations improve satisfaction and/or learnability regarding the feedback about modeling problems. Finally, we intend to provide our recommendations as a list that may serve as a classifier for evaluating process modeling tools in terms of visualizing problems within process models.

REFERENCES

AWAD, A.; WESKE, M. Visualization of compliance violation in business process models. **Lecture Notes in Business Information Processing**, v. 43 LNBIP, p. 182–193, 2010. ISSN 18651348.

AYSOLMAZ, B.; REIJERS, H. Towards an integrated framework for invigorating process models: A research agenda. **Lecture Notes in Business Information Processing**, v. 256, p. 552–558, 2016. ISSN 18651348.

BABBIE, E. R. **The practice of social research**. [S.l.]: Nelson Education, 2015.

BANGOR, A.; KORTUM, P.; MILLER, J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. **Cochrane Database of Systematic Reviews**, v. 2018, n. 6, p. 114–123, 2018. ISSN 1469493X.

BARBOSA, S.; SILVA, B. **Interação Humano-Computador**. [S.l.]: Elsevier Brasil, 2010. ISBN 9788535211207.

BECKER, J.; ROSEMANN, M.; Von Uthmann, C. Guidelines of Business Process Modeling. **LNCS**, Springer Berlin Heidelberg, v. 1806, p. 30–49, 2000.

BOYCE, C.; NEALE, P. Conducting in-depth interviews: A Guide for designing and conducting in-depth interviews. **Evaluation**, v. 2, n. May, p. 1–16, 2006. ISSN 1461-6734.

BROOKE, J. A quick and dirty usability scale. **Usability evaluation in industry**, 1996.

BROOKE, J. SUS: A Retrospective. v. 8, n. 2, p. 29–40, 2013.

BRYMAN, A. Social research methods. **Journal of Chemical Information and Modeling**, v. 53, n. 9, p. 1689–1699, 2013. ISSN 1098-6596.

CABALLERO, H. S. G. et al. Visual analytics for soundness verification of process models. In: **Lecture Notes in Business Information Processing**. [S.l.: s.n.], 2018. v. 308, p. 744–756. ISBN 9783319740294. ISSN 18651348.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. (Ed.). **Readings in Information Visualization: Using Vision to Think**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999. ISBN 1-55860-533-9.

CORDES, C.; VOGELGESANG, T.; APPELRATH, H.-J. A generic approach for calculating and visualizing differences between process models in multidimensional process mining. **Lecture Notes in Business Information Processing**, v. 202, p. 383–394, 2015. ISSN 18651348.

CORRADINI, F. et al. A Guidelines Framework for Understandable BPMN Models. **Data & Knowledge Engineering**, 2017. ISSN 0169-023X.

CORRADINI, F. et al. Animating multiple instances in BPMN collaborations: From formal semantics to tool support. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2018. v. 11080 LNCS, p. 83–101. ISBN 9783319986470. ISSN 16113349.

DANI, V. S.; FREITAS, C. M. D. S.; THOM, L. H. Ten years of visualization of business process models: A systematic literature review. **Computer Standards & Interfaces**, 2019. ISSN 0920-5489. Available from Internet: <http://www.sciencedirect.com/science/article/pii/S0920548918303295>.

DIAS, C. L. d. B. **Behavior Analysis of Process Modeling Tools Based on Anti-Patterns**. Bachelor‘s Thesis — Federal University of Rio Grande do Sul, 2018.

DIKICI, A.; TURETKEN, O.; DEMIRORS, O. Factors Influencing the Understandability of Process Models: A systematic Literature Review. **Information and Software Technology**, v. 2319, n. 0, p. 0–1, 2017.

DUMAS, M. et al. **Fundamentals of Business Process Management**. [S.l.]: Springer Berlin Heidelberg, 2013.

ECKLEDER, A. et al. Realtime detection and coloring of matching operator nodes in workflow nets. In: **CEUR Workshop Proceedings**. Karlsruhe: [s.n.], 2009. v. 501, p. 56–61.

EFFINGER, P.; SPIELMANN, J. Lifting business process diagrams to 2.5 dimensions. In: **Proceedings of SPIE - The International Society for Optical Engineering**. San Jose, CA: [s.n.], 2010. v. 7530. ISBN 978-0-8194-7923-5.

ELO, S.; KYNGÄS, H. The qualitative content analysis process. **Journal of Advanced Nursing**, v. 62, n. 1, p. 107–115, 2008. ISSN 03092402.

EMENS, R.; VANDERFEESTEN, I.; REIJERS, H. The dynamic visualization of business process models: A prototype and evaluation. **Lecture Notes in Business Information Processing**, v. 256, p. 559–570, 2016. ISSN 18651348.

FIGL, K.; KOSCHMIDER, A.; KRIGLSTEIN, S. Visualising process model hierarchies. In: **ECIS 2013 - Proceedings of the 21st European Conference on Information Systems**. Utrecht: Association for Information Systems, 2013.

FINK, A. **How to Conduct Surveys - Arlene Fink**. 6. ed. [S.l.: s.n.], 2002. ISBN 9781483378480.

FINK, A. **The Survey Handbook**. Thousand Oaks, California: [s.n.], 2003.

FOWLER, F. J. J. **Survey Research Methods**. [S.l.: s.n.], 2007. 635–646 p. ISBN 9781452259000.

FOWLER, M.; SCOTT, K. **UML Distilled: A Brief Guide to the Standard Object Modeling Language**. 2. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999. ISBN 0-201-65783-X.

GALL, M. et al. A study of different visualizations for visualizing differences in process models. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9382, p. 99–108, 2015. ISSN 03029743.

GEIGER, M. et al. BPMN 2.0: The State of Support and Implementation. **Future Generation Computer Systems**, 2017.

GROSSMAN, J. D. Color conventions and application standards. In: ____. **Color in Electronic Displays**. Boston, MA: Springer US, 1992. p. 209–218. ISBN 978-1-4757-9754-1.

GULDEN, J.; ATTFIELD, S. Business process models for visually navigating process execution data. **Lecture Notes in Business Information Processing**, v. 256, p. 583–594, 2016. ISSN 18651348.

GULDEN, J.; REIJERS, H. Toward advanced visualization techniques for conceptual modeling. In: **CAiSE 2015 Forum at the 27th International Conference on Advanced Information Systems Engineering**. CEUR Workshop Proceedings, 2015. v. 1367, p. 33–40. Available from Internet: <http://ceur-ws.org/Vol-1367/paper-05.pdf>.

GUO, H.; BROWN, R.; RASMUSSEN, R. Virtual worlds as a model-view approach to the communication of business processes models. In: **CEUR Workshop Proceedings**. Gdansk: CEUR-WS, 2012. v. 855, p. 66–73.

HAMMER, M. What is business process management? In: ____. **Handbook on Business Process Management 1: Introduction, Methods, and Information Systems**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 3–16. ISBN 978-3-642-00416-2.

HIPP, M. et al. Navigating in process model repositories and enterprise process information. In: **Proceedings - International Conference on Research Challenges in Information Science**. Marrakesh: IEEE Computer Society, 2014. ISBN 978-1-4799-2393-9.

HIPP, M.; MUTSCHLER, B.; REICHERT, M. Navigating in process model collections: A new approach inspired by google earth. **Lecture Notes in Business Information Processing**, v. 100 LNBIP, n. PART 2, p. 87–98, 2012. ISSN 18651348.

HIPP, M. et al. Enabling a user-friendly visualization of business process models. **Lecture Notes in Business Information Processing**, v. 202, p. 395–407, 2015. ISSN 18651348.

HOFSTEDE, A. H. ter; AALST, W. M. van der. Yawl: Yet another workflow language. **Information Systems**, Elsevier, v. 30, n. 4, p. 245–275, June 2005.

HOLZMÜLLER-LAUE, S. et al. Visual simulation for the BPM-based process automation. **Lecture Notes in Business Information Processing**, v. 158 LNBIP, p. 48–62, 2013. ISSN 18651348.

IVANCHIKJ, A.; FERME, V.; PAUTASSO, C. BPMeter: Web service and application for static analysis of BPMN 2.0 collections. In: S., D. F. Z. (Ed.). **CEUR Workshop Proceedings**. [S.l.]: CEUR-WS, 2015. v. 1418, p. 30–34.

JABLONSKI, S.; GOETZ, M. Perspective oriented business process visualization. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2008. v. 4928 LNCS, p. 144–155. ISBN 3540782370. ISSN 03029743.

JONATHAN, L.; FENG, J. H.; HOCHHEISER, H. **Research Methods in Human-Computer Interaction**. 2. ed. [S.l.]: Morgan Kaufmann, 2017. 540 p.

JOSCHKO, P.; WIDOK, A.; PAGE, B. A simulation tool for maintenance processes of offshore wind farms. In: **International Conference on Harbour, Maritime and Multimodal Logistics Modelling and Simulation**. [S.l.]: I3M Conference, 2013. v. 1, p. 83–89.

JOŠT, G. et al. Improving cognitive effectiveness of business process diagrams with opacity-driven graphical highlights. **Decision Support Systems**, v. 103, p. 58–69, 2017. ISSN 01679236.

JÚNIOR, V. H. G. et al. An interface prototype proposal to a semiautomatic process model verification method based on process modeling guidelines. In: HAMMOUDI, S. et al. (Ed.). **Enterprise Information Systems**. Cham: Springer International Publishing, 2018. p. 611–629. ISBN 978-3-319-93375-7. ISSN 1865-1348.

KABICHER-FUCHS, S.; KRIGLSTEIN, S.; FIGL, K. Timeline visualization for documenting process model change. In: M., R.-M. S. W. (Ed.). **Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)**. [S.l.]: Gesellschaft fur Informatik (GI), 2012. P-206, p. 95–108. ISBN 978-3-88579-600-8.

KABICHER, S.; KRIGLSTEIN, S.; RINDERLE-MA, S. Visual change tracking for business process models. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2011. v. 6998 LNCS, p. 504–513.

KATHLEEN, N.; ROSS, B.; KRIGLSTEIN, S. Storyboard augmentation of process model grammars for stakeholder communication. In: **IVAPP 2014 - Proceedings of the 5th International Conference on Information Visualization Theory and Applications**. Lisbon: SciTePress, 2014. p. 114–121. ISBN 978-989-758-005-5.

KELLER, G.; NÜTTGENS, M.; SCHEER, A. W. **Semantische Prozessmodellierung auf der Grundlage" ereignisgesteuerter Prozessketten (EPK)**. [S.l.]: Iwi, 1992.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for Performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007.

KOLB, J.; REICHERT, M. A Flexible Approach for Abstracting and Personalizing Large Business Process Models. **SIGAPP Appl. Comput. Rev.**, v. 13, n. 1, p. 6–18, mar. 2013. ISSN 1559-6915.

KOSCHMIDER, A.; FIGL, K.; SCHOKNECHT, A. A comprehensive overview of visual design of process model element labels. **Lecture Notes in Business Information Processing**, v. 256, p. 571–582, 2016. ISSN 18651348.

KOSCHMIDER, A.; KRIGLSTEIN, S.; ULLRICH, M. Investigations on user preferences of the alignment of process activities, objects and roles. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8185 LNCS, p. 57–74, 2013. ISSN 03029743.

KRENN, F.; KEPLER, J. Dealing with Process Complexity: A Multiperspective Approach. ACM, p. 10, 2018.

KRIGLSTEIN, S.; RINDERLE-MA, S. Change visualizations in business processes: Requirements analysis. In: **GRAPP 2012 IVAPP 2012 - Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications**. Rome: [s.n.], 2012. p. 584–593. ISBN 978-989-8565-02-0.

KRIGLSTEIN, S.; RINDERLE-MA, S. A visualization concept for high-level comparison of process model versions. **Lecture Notes in Business Information Processing**, v. 132 LNBIP, p. 465–476, 2013. ISSN 18651348.

KRIGLSTEIN, S.; WALLNER, G.; RINDERLE-MA, S. A visualization approach for difference analysis of process models and instance traffic. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 8094 LNCS, p. 219–226, 2013. ISSN 03029743.

KUMMER, T.-F.; RECKER, J.; MENDLING, J. Enhancing understandability of process models through cultural-dependent color adjustments. **Decision Support Systems**, v. 87, p. 1–12, 2016. ISSN 01679236.

KVALE, S. **InterViews. An introduction to qualitative research writing**. [S.l.: s.n.], 1996. 267–270 p. ISBN 0803958196.

LAUE, R.; AWAD, A. Visual suggestions for improvements in business process diagrams. **Journal of Visual Languages and Computing**, v. 22, n. 5, p. 385–399, 2011. ISSN 1045926X.

LEITNER, M. et al. An experimental study on the design and modeling of security concepts in business processes. **Lecture Notes in Business Information Processing**, v. 165 LNBIP, p. 236–250, 2013. ISSN 18651348.

LEOPOLD, H.; MENDLING, J.; GÜNTHER, O. Learning from Quality Issues of BPMN Models from Industry. **IEEE Software**, Institute of Electrical and Electronics Engineers (IEEE), v. 33, n. 4, p. 26–33, jul 2016.

LEWIS, J. R.; SAURO, J. The factor structure of the system usability scale. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2009. v. 5619 LNCS, p. 94–103. ISBN 3642028055. ISSN 03029743.

MEIDAN, A. et al. A Survey on Business Processes Management Suites. **Computer Standards & Interfaces**, 2017.

MENDLING, J.; REIJERS, H. A.; Van Der Aalst, W. M. P. Seven Process Modeling Guidelines (7PMG). **Information and Software Technology**, v. 52, p. 127–136, 2009.

MERINO, M. et al. Extending BPMN model for improving expressiveness and machine-understandability. **Advances in Intelligent Systems and Computing**, v. 444, p. 297–306, 2016. ISSN 21945357.

Moreno-Montes De Oca, I. et al. A Systematic Literature Review of Studies on Business Process Modeling Quality. **Information and Software Technology**, v. 58, p. 187–205, 2015.

MUELLER-WICKOP, N. et al. Towards automated financial process auditing: Aggregation and visualization of process models. In: **Proceedings of the 4th International Workshop on Enterprise Modelling and Information Systems Architectures, EMISA 2011**. Hamburg: [s.n.], 2011. p. 121–134. ISBN 978-3-88579-284-0.

MUNZNER, T. **Visualization Analysis and Design**. [S.l.]: CRC press, 2014.

OBERHAUSER, R.; POGOLSKI, C.; MATIC, A. VR-BPMN: Visualizing BPMN Models in Virtual Reality. In: **Lecture Notes in Business Information Processing**. [S.l.: s.n.], 2018. v. 319, p. 83–97. ISBN 9783319942131. ISSN 18651348.

Object Management Group. **BPMN Specification - Business Process Model and Notation**. 2015.

OMG (Object Management Group). **Business process modeling notation (BPMN) 2.0 beta 1**. 2009. Available from Internet: <http://www.omg.org/cgi-bin/doc?dtc/09-08-14.pdf>.

ONGGO, B. S. S.; KARPAT, O. Agent-based Conceptual Model Representation Using BPMN. In: **Proceedings of the Winter Simulation Conference**. Phoenix, Arizona: Winter Simulation Conference, 2011. (WSC '11), p. 671–682.

PERALTA, A. et al. A tool for assessing quality of rescue plans by combining visualizations of different business process perspectives. **Lecture Notes in Business Information Processing**, v. 233, p. 155–166, 2015. ISSN 18651348.

PESIC, M.; SCHONENBERG, H.; AALST, W. M. P. van der. Declare: Full support for loosely-structured processes. In: **Proceedings of the 11th IEEE International Enterprise Distributed Object Computing Conference**. Washington, DC, USA: IEEE Computer Society, 2007. (EDOC '07), p. 287–. ISBN 0-7695-2891-0. Available from Internet: <http://dl.acm.org/citation.cfm?id=1317532.1318056>.

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for Conducting Systematic Mapping Studies in Software Engineering: An Update. **Information and Software Technology**, v. 64, p. 1–18, 2015.

PETRI, C. A. Fundamentals of a theory of asynchronous information flow. In: **IFIP Congress**. [S.l.: s.n.], 1962. p. 386–390.

PHILIPPI, S.; HILL, H. J. Communication support for systems engineering-process modelling and animation with APRIL. 2007.

PINI, A.; BROWN, R.; WYNN, M. Process visualization techniques for multi-perspective process comparisons. **Lecture Notes in Business Information Processing**, v. 219, p. 183–197, 2015. ISSN 18651348.

POLDERDIJK, M. et al. A visualization of human physical risks in manufacturing processes using BPMN. In: **Lecture Notes in Business Information Processing**. [S.l.: s.n.], 2018. v. 308, p. 732–743. ISBN 9783319740294. ISSN 18651348.

REICHERT, M. Visualizing large business process models: Challenges, techniques, applications. **Lecture Notes in Business Information Processing**, v. 132 LNBIP, p. 725–736, 2013. ISSN 18651348.

REIJERS, H. A. et al. Syntax highlighting in business process models. **Decision Support Systems**, v. 51, n. 3, p. 339–349, jun. 2011. ISSN 0167-9236.

REIJERS, H. A.; MENDLING, J. Modularity in Process Models: Review and Effects. In: **Lecture Notes in Computer Science**. [S.l.]: Springer Berlin Heidelberg, 2008. p. 20–35. ISBN 978-3-540-85758-7. ISSN 0302-9743.

RINDERLE, S. B. et al. Business Process Visualization - Use Cases, Challenges, Solutions. In: **Proceedings of the Eighth International Conference on Enterprise Information Systems (ICEIS'06): Information System Analysis and Specification**. [S.l.]: INSTICC Press, 2006. p. 204–211. ISBN 9789728865436.

RITTGEN, P. Quality and Perceived Usefulness of Process Models. In: **Proceedings of the 2010 ACM Symposium on Applied Computing - SAC 1́0**. New York, NY, USA: ACM, 2010. ISBN 978-1-60558-639-7.

ROBSON, C.; MCCARTAN, K. **Real World Research, 4th Edition**. [S.l.: s.n.], 2016. ISBN 978-1-118-74523-6.

ROSA, M. L. et al. Managing Process Model Complexity via Concrete Syntax Modifications. **IEEE Transactions on Industrial Informatics**, v. 7, n. 2, 2011.

ROSA, M. L. et al. Managing Process Model Complexity Via Abstract Syntax Modifications. **IEEE Transactions on Industrial Informatics**, v. 7, n. 4, 2011.

ROY, S. et al. An Empirical Study of Error Patterns in Industrial Business Process Models. **IEEE Transactions on Services Computing**, v. 7, n. 2, p. 140–153, 2014. ISSN 19391374.

ROZMAN, T.; POLANCIC, G.; HORVAT, R. V. Analysis of most common process modeling mistakes in bpmn process models. In: **2008 BPM and Workflow Handbook**. [S.l.]: University of Maribor Slovenia, 2008.

SADIQ, S.; GOVERNATORI, G.; NAIMIRI, K. **Modeling Control Objectives for Business Process Compliance**. [S.l.], 2007.

SALDAÑA, J. **The Coding Manual for Qualitative Researchers**. [S.l.: s.n.], 2013. 329 p. ISSN 01631829. ISBN 9781446247365.

SALNITRI, M.; DALPIAZ, F.; GIORGINI, P. Designing secure business processes with SecBPMN. **Software and Systems Modeling**, v. 16, n. 3, p. 737–757, 2017. ISSN 16191366.

SARAEIAN, S.; SHIRAZI, B.; MOTAMENI, H. Towards an Extended BPMS Prototype: Open Challenges of BPM to Flexible and Robust Orchestrate of Uncertain Processes. **Computer Standards & Interfaces**, v. 57, p. 1–19, 2017.

SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. In: **Proceedings of the 1996 IEEE Symposium on Visual Languages**. Washington, DC, USA: IEEE Computer Society, 1996. (VL '96). ISBN 0-8186-7508-X. Available from Internet: <http://dl.acm.org/citation.cfm?id=832277.834354>.

SMUTS, M.; BURGER, C.; SCHOLTZ, B. Composite, Real-time Validation for Business Process Modelling. **Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology - SAICSIT '14**, p. 93–103, 2014.

SNOECK, M. et al. Testing a selection of bpmn tools for their support of modelling guidelines. In: SPRINGER. **IFIP Working Conference on The Practice of Enterprise Modeling**. [S.l.], 2015. p. 111–125.

STORCH, A.; LAUE, R.; GRUHN, V. Measuring and visualising the quality of models. In: **2013 IEEE 1st International Workshop on Communicating Business Process and Software Models: Quality, Understandability, and Maintainability, CPSM 2013**. Eindhoven: IEEE Computer Society, 2013. ISBN 978-1-4799-2941-2.

STROPPI, L.; CHIOTTI, O.; VILLARREAL, P. A BPMN 2.0 extension to define the resource perspective of business process models. In: **14th Ibero-American Conference on Software Engineering and 14th Workshop on Requirements Engineering, CIbSE 2011**. Rio de Janeiro: [s.n.], 2011. p. 25–38.

SUCHENIA, A.; LIGEZA, A. Event anomalies in modeling with bpmn. **International Journal of Computer Technology And Applications**, v. 6, n. 5, p. 789–797, 2015.

"SUCHENIA, A. et al. Selected approaches towards taxonomy of business process anomalies. In: ____. **Advances in Business ICT: New Ideas from Ongoing Research**. Cham: Springer International Publishing, 2017. p. 65–85. ISBN 978-3-319-47208-9. Available from Internet: <https://doi.org/10.1007/978-3-319-47208-9_5>.

TUFTE, E. **Envisioning Information**. [S.l.]: Graphics Pr, 1990. ISBN 0961392118.

Van Der Aalst, W. M. P. Business Process Management: A Comprehensive Survey. **ISRN Software Engineering**, Hindawi Publishing Corporation, v. 507984, 2013.

VIDACIC, T.; STRAHONJA, V. Taxonomy of anomalies in business process models. In: ESCALONA, M. J. et al. (Ed.). **Information System Development**. Cham: Springer International Publishing, 2014. p. 283–294. ISBN 978-3-319-07215-9.

WESKE, M. **Business Process Management. Concepts, Languages, Architectures**. [S.l.]: Springer Berlin Heidelberg, 2012. ISSN 10744770. ISBN 978-3-642-28615-5.

WITT, S. et al. Visualization of checking results for graphical validation rules. **Communications in Computer and Information Science**, v. 532, p. 120–136, 2015. ISSN 18650929.

YOUSFI, A. et al. uBPMN: A BPMN Extension for Modeling Ubiquitous Business Processes. **Information and Software Technology**, v. 74, p. 55–68, 2016.

APPENDIX A — VISUALIZATION ANALYSIS DATA EXTRACTION

This Appendix displays the data extraction Tables A.1 and A.2 generated from the visualization analysis perfomed in the systematic literature review we conducted (Chapter 3).

Table A.1: Data extracted regarding aspects "Why?" for the visualization-related analysis based on Munzner's framework (MUNZNER, 2014): each mark indicates that the paper, referenced in the line, describes visual representations and/or interactive features that can be mapped to the framework's concept in the corresponding column. Totals and percentages for each column are presented; the dark bar represents the number of papers with the corresponding feature.

	Why														
	Actions						**Targets**								
	Analyze			**Query**			**All Data**			**Attributes**					
Articles	**Discover**	**Present**	**Search**	**Identify**	**Compare**	**Summarize**	**Trends**	**Outliers**	**Features**	**Distribution**	**Dependency**	**Correlation**	**Similarity**	**Network Data**	**Spatial Data**
(JÚNIOR et al., 2018)	x	x	x	x	x	x		x		x			x		
(ECKLEDER et al., 2009)	x	x							x			x		x	
(REIJERS et al., 2011)	x	x							x			x		x	
(KUMMER; RECKER; MENDLING, 2016)		x							x					x	
(EMENS; VANDERFEESTEN; REIJERS, 2016)		x	x	x					x		x			x	
(HIPP; MUTSCHLER; REICHERT, 2012)	x	x	x			x			x					x	
(STORCH; LAUE; GRUHN, 2013)	x	x	x		x	x	x	x		x		x			
(HIPP et al., 2014)	x	x	x	x	x	x		x	x		x			x	
(IVANCHIKJ; FERME; PAUTASSO, 2015)	x	x	x		x	x		x		x		x			
(GULDEN; ATTFIELD, 2016)	x	x	x	x	x	x	x		x					x	

Table A.1 continued from previous page

Articles	Discover	Present	Search	Identify	Compare	Summarize	Trends	Outliers	Features	Distribution	Dependency	Correlation	Similarity	Network Data	Spatial Data
(AWAD; WESKE, 2010)	x	x	x						x					x	
(LAUE; AWAD, 2011)	x	x	x	x					x					x	
(WITT et al., 2015)	x	x	x	x					x		x			x	
(CORRADINI et al., 2017)	x	x	x			x			x			x		x	
(ONGGO; KARPAT, 2011)									x					x	
(MUELLER-WICKOP et al., 2011)	x	x	x						x					x	
(JOSCHKO; WIDOK; PAGE, 2013)		x							x						
(KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013)	x	x	x	x								x		x	
(HIPP et al., 2015)	x	x	x						x					x	x
(MERINO et al., 2016)	x	x	x						x					x	
(SALNITRI; DALPIAZ; GIORGINI, 2017)	x	x	x						x					x	
(EFFINGER; SPIELMANN, 2010)	x	x	x											x	x
(STROPPI; CHIOTTI; VILLARREAL, 2011)		x	x						x					x	
(KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011)	x	x	x		x				x			x	x	x	
(KRIGLSTEIN; RINDERLE-MA, 2012)	x	x	x						x			x	x	x	

Table A.1 continued from previous page

Articles	Discover	Present	Search	Identify	Compare	Summarize	Trends	Outliers	Features	Distribution	Dependency	Correlation	Similarity	Network Data	Spatial Data
(KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012)	x	x	x	x	x			x	x			x	x	x	
(KOLB; REICHERT, 2013)	x	x	x			x								x	
(KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013)		x	x		x		x	x	x			x	x	x	
(KRIGLSTEIN; RINDERLE-MA, 2013)	x	x	x		x	x	x			x		x			
(REICHERT, 2013)	x	x	x	x					x					x	
(CORDES; VOGELGESANG; APPELRATH, 2015)		x	x		x				x			x	x	x	
(GALL et al., 2015)		x			x				x			x	x	x	
(PERALTA et al., 2015)	x	x	x	x				x	x		x			x	x
(PINI; BROWN; WYNN, 2015)	x	x	x	x	x	x	x	x	x			x		x	
(GUO; BROWN; RASMUSSEN, 2012)	x	x	x						x			x		x	x
(HOLZMÜLLER-LAUE et al., 2013)	x	x	x	x					x					x	
(KATHLEEN; ROSS; KRIGLSTEIN, 2014)	x	x	x						x			x		x	
(JOŠT et al., 2017)	x	x	x	x					x		x			x	
(POLDERDIJK et al., 2018)	x	x	x	x					x					x	
(CORRADINI et al., 2018)	x	x	x						x		x			x	

Table A.1 continued from previous page

Articles	Discover	Present	Search	Identify	Compare	Summarize	Trends	Outliers	Features	Distribution	Dependency	Correlation	Similarity	Network Data	Spatial Data
(KRENN; KEPLER, 2018)		x	x		x							x		x	
(CABALLERO et al., 2018)	x	x	x	x					x					x	
(OBERHAUSER; POGOLSKI; MATIC, 2018)		x	x	x					x		x			x	x
Total	33	42	37	16	13	10	5	8	35	4	7	17	7	38	5
%	76.74	97.67	86.05	37.21	30.23	23.26	11.63	18.60	81.40	9.30	16.28	39.53	16.28	88.37	11.63

Table A.2: Data extracted regarding aspects "How?" for the visualization-related analysis based on Munzner's framework (MUNZNER, 2014): each mark indicates that the paper, referenced in the line, describes visual representations and/or interactive features that can be mapped to the framework's concept in the corresponding column. Totals and percentages for each column are presented; the dark bar represents the number of papers with the corresponding feature.

	How																
	Encode									**Manipulate**			**Facet**			**Reduce**	
	Arrange			**Map**													
Articles	**Express**	**Separate**	**Arrange Network**	**Hue**	**Saturation**	**Luminance**	**Size**	**Shape**	**Motion**	**Change**	**Select**	**Navigate**	**Juxtapose**	**Partition**	**Superimpose**	**Filter**	**Aggregate**
(JÚNIOR et al., 2018)	x	x		x		x	x	x		x	x		x				
(ECKLEDER et al., 2009)				x						x							
(REIJERS et al., 2011)				x						x							
(KUMMER; RECKER; MENDLING, 2016)				x													
(EMENS; VANDERFEESTEN; REIJERS, 2016)				x		x			x	x		x	x			x	
(HIPP; MUTSCHLER; REICHERT, 2012)			x					x		x		x				x	
(STORCH; LAUE; GRUHN, 2013)				x				x		x				x	x	x	
(HIPP et al., 2014)	x	x		x				x		x	x	x	x			x	
(IVANCHIKJ; FERME; PAUTASSO, 2015)	x	x		x				x		x				x			

Table A.2 continued from previous page

Articles	Express	Separate	Arrange Network	Hue	Saturation	Luminance	Size	Shape	Motion	Change	Select	Navigate	Juxtapose	Partition	Superimpose	Filter	Aggregate
(GULDEN; ATTFIELD, 2016)	x	x		x				x		x	x	x	x				
(AWAD; WESKE, 2010)				x													
(LAUE; AWAD, 2011)			x	x				x							x		
(WITT et al., 2015)			x	x	x									x		x	
(CORRADINI et al., 2017)			x	x						x		x					
(ONGGO; KARPAT, 2011)			x					x		x	x	x	x				
(MUELLER-WICKOP et al., 2011)			x	x													x
(JOSCHKO; WIDOK; PAGE, 2013)								x									
(KOSCHMIDER; KRIGLSTEIN; ULLRICH, 2013)			x	x						x	x	x	x				
(HIPP et al., 2015)			x	x			x	x									
(MERINO et al., 2016)			x					x									
(SALNITRI; DALPIAZ; GIORGINI, 2017)				x				x									
(EFFINGER; SPIELMANN, 2010)			x					x		x		x					
(STROPPI; CHIOTTI; VILLARREAL, 2011)								x									

Table A.2 continued from previous page

Articles	Express	Separate	Arrange Network	Hue	Saturation	Luminance	Size	Shape	Motion	Change	Select	Navigate	Juxtapose	Partition	Superimpose	Filter	Aggregate
(KABICHER; KRIGLSTEIN; RINDERLE-MA, 2011)			x	x											x		
(KRIGLSTEIN; RINDERLE-MA, 2012)			x	x						x	x			x	x		
(KABICHER-FUCHS; KRIGLSTEIN; FIGL, 2012)			x	x				x		x	x	x	x		x	x	
(KOLB; REICHERT, 2013)			x	x							x	x		x		x	x
(KRIGLSTEIN; WALLNER; RINDERLE-MA, 2013)			x	x			x			x				x	x		x
(KRIGLSTEIN; RINDERLE-MA, 2013)	x	x		x				x									
(REICHERT, 2013)			x	x				x		x	x	x				x	x
(CORDES; VOGELGESANG; APPELRATH, 2015)			x	x											x		
(GALL et al., 2015)				x	x			x									
(PERALTA et al., 2015)				x				x		x	x		x				
(PINI; BROWN; WYNN, 2015)	x			x			x	x			x						
(GUO; BROWN; RASMUSSEN, 2012)				x	x	x	x	x	x			x			x		
(HOLZMÜLLER-LAUE et al., 2013)			x						x		x	x	x				x
(KATHLEEN; ROSS; KRIGLSTEIN, 2014)			x	x	x	x	x	x									

Table A.2 continued from previous page

Articles	Express	Separate	Arrange Network	Hue	Saturation	Luminance	Size	Shape	Motion	Change	Select	Navigate	Juxtapose	Partition	Superimpose	Filter	Aggregate
(JOŠT et al., 2017)	x				x						x	x					
(POLDERDIJK et al., 2018)			x	x				x		x	x	x					
(CORRADINI et al., 2018)								x	x	x	x		x				
(KRENN; KEPLER, 2018)			x	x						x	x					x	
(CABALLERO et al., 2018)				x			x	x		x	x	x	x		x		
(OBERHAUSER; POGOLSKI; MATIC, 2018)			x	x						x	x	x					
Total	7	5	22	34	5	4	7	25	4	23	18	17	11	6	9	9	5
%	16.28	11.63	51.16	79.07	11.63	9.30	16.28	58.14	9.30	53.49	41.86	39.53	25.58	13.95	20.93	20.93	11.63

APPENDIX B — QUESTIONNAIRE

This Appendix presents the questionnaire (see Table B.1) used to gather data from a random sample of participants with academic and business background, and some level of experience in the business process modeling task.

Table B.1: Questions composing the survey questionnaire, comprised by four sections. The sample images, displayed in questions 13 to 16, were based on four different modeling tools (Bizagi, Bonita, Camunda, and Signavio, respectively), and the respondents did not know which tools these images represent. Respondents informally reported a 10-minutes average time to answer the questionnaire.

Id	Question *Response format {Options}*
	Section 1: User profile *General data*
1	Age *Number*
2	Are you linked to the academic, private or public sector? *Multiple-choice {Academic (e.g.: Professor, Researcher, Student, ...), Private (e.g.: Business Professional, Designer, Developer, ...), Public (e.g.: Public job at Federal, National levels, ...), None, Other (Justify | Text)}*
3	Your main occupation is *Multiple-choice {None, Student, Researcher, Software Developer, Process Analyst, Other (Justify | Text)}*
	Section 2: User profile *Experience with process modeling*
4	For how long have you known business process modeling? *Single-choice {I do not know, Less than 2 years, Between 2 and 4 years, More than 4 years}*
5	For how long have you known BPMN (Business Process Model and Notation)? *Single-choice {I do not know, Less than 2 years, Between 2 and 4 years, More than 4 years}*
6	When was the last time you modeled a business process? *Single-choice {I never modeled, This week, Last week, Last month, Last semester or latter}*
7	Which modeling tool did you use? [You may select more than one] *Multiple-choice {None, I modeled using paper and pen (or similar), Bizagi, Camunda, Bonita, Signavio, Oracle BPM, Other (Justify | Text)}*
8	What was the process modeling language/notation you used? [You may select more than one] *Multiple-choice {None, BPMN, EPC, Petri Nets, YAWL, UML AD, Other (Justify | Text)}*
9	You consider your level of experience with business process modeling

Table B.1 continued from previous page

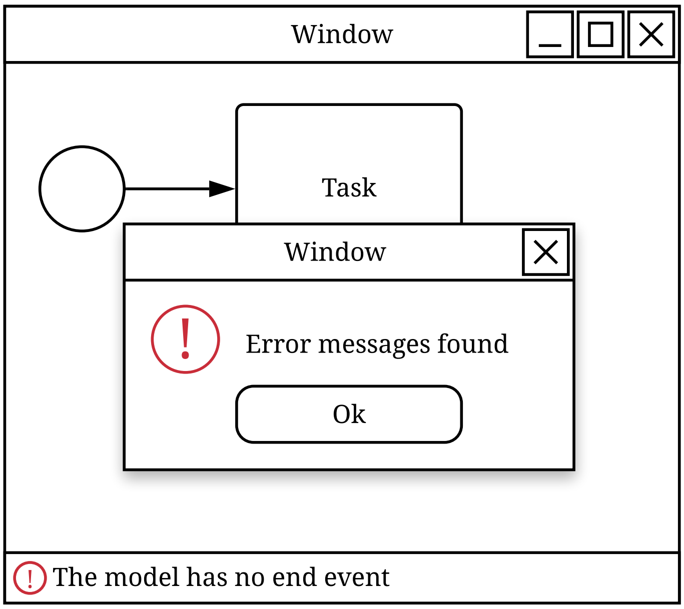
Id	**Question** *Response format {Options}*
	Single-choice {I never modeled, Very low, Low, Moderate, High, Very high}
10	How many years of professional experience (employee/trainee) with business process modeling do you have? *Single-choice {None, Less than 2 years, Between 2 and 4 years, More than 4 years}*
11	Your process modeling learning took place in *Multiple-choice {Academic, Business, Public, None, Other (Justify | Text)}*
	Section 3: About feedback on problems within process model
12	How often do you have difficulty in trying to identify problems in a process model? *Single-choice {I never modeled a process, None, Very low, Low, Moderate, High, Very high}*
13	How satisfied are you with this way of providing feedback about the process model problems? *Sample image (based on Bizagi)* 
14	How satisfied are you with this way of providing feedback about the process model problems? *Sample image (based on Bonita)*

Table B.1 continued from previous page

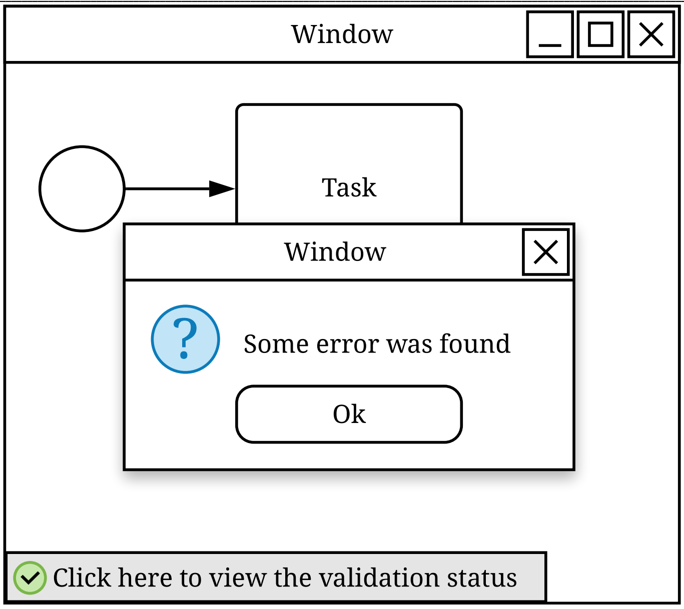
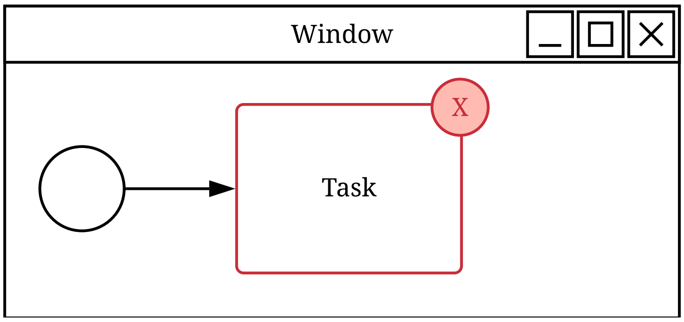
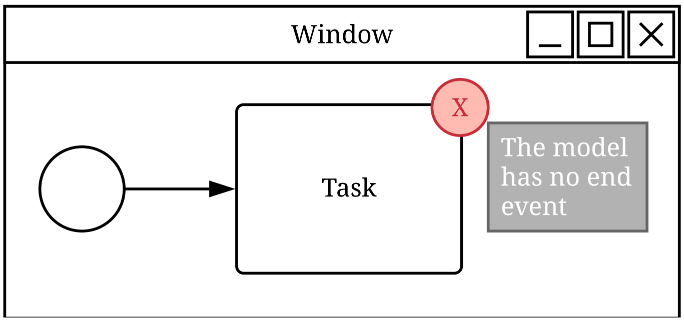
Id	**Question** *Response format {Options}*
	
15	How satisfied are you with this way of providing feedback about the process model problems? *Sample image (based on Camunda)* 
16	How satisfied are you with this way of providing feedback about process model problems? *Sample image (based on Signavio)* 

Table B.1 continued from previous page

Id	**Question** *Response format {Options}*
17	How do you think modeling tools should provide feedback on problems in process models? *Text*
18	What kind of information do you think modeling tools should present about problems in process models? *Text*
	Section 5: Satisfiability regarding feedback on problems within process model
19	I always analyze the feedback that the tool provides *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
20	I consider the feedback from the tool that I use unnecessarily complex *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
21	I consider the feedback from the tool that I use well explanatory *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
22	I consider that the feedback requires a person with technical knowledge to be properly used *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
23	I consider that the feedback is very well integrated with the modeling tool *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
24	I consider that the feedback have many inconsistencies *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
25	I consider that the feedback allow people to learn quickly about modeling *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
26	I consider the feedback confusing *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*

Table B.1 continued from previous page

Id	**Question** *Response format {Options}*
27	I consider I felt myself confident when trying to correct process models based on the feedback *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*
28	I consider I needed to learn a lot of new things before I could understand and use the feedback *Single-choice {I never modeled a process, Strongly disagree, Disagree, Neutral, Agree, Strongly agree}*

APPENDIX C — INTERVIEW

This Appendix presents the detailed interview script used for conducting each interview.

Table C.1: Interview script used to conduct the interviews and gather data to complement questionnaire subjective questions (i.e., questions 18 and 19). The white rows are read to the interviewee.

Interview information *Interviewee name, interview data, time, and duration*
Introduction
I would like to thank you for answering the questionnaire. The purpose of this semi-structured interview is to complement the data obtained through the questionnaire.
This interview takes place in the context of my MSc in Computing thesis at PPGC-UFRGS with focus on "visualization of problems within business process models" under the supervision of Professor Lucinéia Heloisa Thom and co-advisor Professor Carla Maria Dal Sasso Freitas.
This interview should take less than 15 minutes, I will be typing all your answers, and your responses will be kept confidential. In other words, we won't link your name in our report to any answer you provide us.
At any time you may stop the interview if you feel it is necessary.
Interview
How do you think modeling tools should provide feedback about problems in process model?
What kind of information do you think modeling tools should present about problems in process model?
Conclusion
Thank you for your participation! Would you like to add something else?
Many thanks again for your participation. As soon as possible, I will share the results with the academic community.

APPENDIX D — SCIENTIFIC CONTRIBUTIONS

This Appendix presents the publications of the author during his studies towards the M.SC. Degree in Computing at PPGC-UFRGS. There are three papers directly related to this dissertation. During the MSc project we also co-supervised an undergraduate work:

- Vinicius Stein Dani, Carla Maria Dal Sasso Freitas, Lucinéia Heloisa Thom. **Ten Years of Visualization of Business Process Models: A Systematic Literature Review** (2019). *Computer Standards & Interfaces* (CS&I).
  Status: Published. Qualis: B1;
- Clemilson Luís de Brito Dias, Vinicius Stein Dani, Jan Mendling, Lucinéia Heloisa Thom. **Process Modeling Problems: An analysis of BPMN 2.0 tools behavior** (2019). *1st Value and Quality of Enterprise Modelling Workshop, 17th International Conference on Business Process Management (BPM 2019).*
  Status: Accepted. Qualis: A2;
- Clemilson Luís de Brito Dias (author), Vinicius Stein Dani (co-advisor), Lucinéia Heloisa Thom (advisor). **Análise de Comportamento de Ferramentas de Modelagem de Processos com Base em Anti-Padrẽs** (2018)
  Undergraduate dissertation.

The author also collaborated in two other publications, whose first authors are colleagues of the BPM Research Group:

- Valter Helmuth Goldberg Júnior, Vinicius Stein Dani, Diego Toralles Avila, Lucinéia Heloisa Thom, José Palazzo Moreira de Oliveira, Marcelo Fantinato. **An Interface Prototype Proposal to a Semiautomatic Process Model Verification Method Based on Process Modeling Guidelines** (2018). *Springer Book Chapter.*
  Status: Published. Qualis: N/A;
- Ana Cláudia de Almeida Bordignon, Lucinéia Heloisa Thom, Thanner Soares Silva, Vinicius Stein Dani, Marcelo Fantinato. **Natural Language Processing in Business Process Identification and Modeling: A Systematic Literature Review** (2018). *XIV Simpósio Brasileiro de Sistemas de Informação (SBSI 2018).*
  Status: Published. Qualis: B2.

Finally, the author also collaborated with colleagues of another group, in a paper related to a course he attended to with them:

- Gabrielle Almeida de Souza, Laura Amaya Torres, Vinicius Stein Dani, David Steeven Villa, Abel Ticona Larico, Anderson Maciel, Luciana Nedel. **Evaluation of Visual, Auditory and Vibro-Tactile Alerts in Supervised Interfaces** (2018). *20th Symposium on Virtual and Augmented Reality (SVR 2018).*
  Status: Published. Qualis: B2.