TÉCNICO LISBOA

Enterprise-wide Business Process Integration from Multiple Concerns

Joana Filipa Pereira Ribeiro

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisor: Prof. Pedro Manuel Moreira Vaz Antunes de Sousa

Examination Committee

Chairperson: Prof. Francisco João Duarte Cordeiro Correia dos Santos
Supervisor: Prof. Pedro Manuel Moreira Vaz Antunes de Sousa
Member of the Committee: Prof. José Luís Brinquete Borbinha

December 2020

Acknowledgments

I would like to thank my beloved parents, Vanda Pereira and Francisco Ribeiro, and my dear brother, Vasco Ribeiro, for the tremendous support over these last five years and for always believing in me. To my grandparents, whose wisdom I long to achieve, for being the most perfect role models I could look up to. To my godparents, because I was fortunate enough to be blessed with an extra set of parents to walk me through life, for their unconditional love and kindness. To my uncles and aunts, who took me to the wildest adventures growing up and provided with the most astonishing childhood. To my little cousins, who bring such joy and light into my life. To my sister-cousins, for always cheering me up and sharing life-changing laughing sessions with me.

I would also like to thank my rock and pillar in life, João Campos, for being my safety net and always catching me when I fall. For always being there for me through thick and thin, and reassuring me whenever I needed. Thank you for being by my side every step of the way, and hopefully you'll continue to do so for the rest of my life.

To the friends who got me through this bumpy, yet unforgettable, 5-year walk - Sara Esperto, Tiago Romero, Carlos Carvalho, Mafalda Mendes and Mafalda Morgado - for the unconditional support, kindness and friendship that I will never forget.

And to my best friends, for their understanding and support throughout all these years. For always having a kind word to say, for facing by my side the adventures of growing up and for sharing with me some of the most memorable moments of my life. Because whatever you do in this life, it's not legendary unless your friends are there to see it.

I would also like to acknowledge my dissertation supervisors Prof. Pedro Manuel Moreira Vaz Antunes de Sousa and Prof. Sérgio Luís Proença Duarte Guerreiro for their insight, support and sharing of knowledge that has made this Thesis possible.

To each and every one of you – Thank you.

Abstract

Business processes are the core asset of an organization and they deeply impact its functioning on every aspect. Different stakeholders within a company have distinct concerns, aspirations and points of view regarding a certain business process, perceiving it in contrasting manners. Business Process Modeling aims to portray the way organizations conduct their business processes through abstract descriptions, therefore the production of different models for the same business process stems from the existence of different stakeholder-specific views, which often lack accuracy and consistency. As such, this work will develop an approach to integrate views from different stakeholders, by asking them a set of questions in a form format. From their answers, relations of resemblance and composition between the activities of each business process will be extracted and used to construct a Consolidated Model that gathers the knowledge of all the stakeholders. Thus, a solution for integrating multiple concerns is detailed with the goal of making the business process models valuable for the organization and decrease ambiguity issues. A demonstration in a bank context as a real-world motivating scenario was used to show how the solution unfolds when integrating the respective views in order to successfully build a Consolidated Model.

Keywords

Business Process; Business Process Modeling; Stakeholder-specific Process View; View Integration; Consolidated Model; Form Inquiry.

Resumo

Os processos de negócios são o principal ativo das organizações e têm um impacto profundo no seu bom funcionamento em todos os aspectos. Diferentes partes interessadas dentro de uma empresa têm preocupações, aspirações e pontos de vista distintos em relação a determinado processo de negócio, apreendendo-os de maneiras contrastantes. A Modelação de Processos de Negócios visa retratar a forma como as organizações conduzem os seus processos de negócios através de descrições abstratas, e por essa razão a produção de diferentes modelos para o mesmo processo de negócios decorre da existência de diferentes visões específicas das partes interessadas, que muitas vezes carecem de precisão e consistência. Como tal, este trabalho desenvolverá uma abordagem para integrar vistas de diferentes partes interessadas, fazendo-lhes um conjunto de perguntas em um formato de formulário. A partir das suas respostas, serão extraídas relações de semelhança e composição entre as atividades de cada processo de negócio, que serão posteriormente utilizadas para a construção de um Modelo Consolidado que reúne o conhecimento de todas as partes interessadas. Assim, uma solução para integração de múltiplos interesses é detalhada com o objetivo de adicionar valor à organização e diminuir os problemas de ambiguidade. Uma demonstração em contexto bancário é utilizada como um cenário real para mostrar como a solução se desdobra ao integrar as respectivas visões para construir com sucesso um Modelo Consolidado.

Palavras Chave

Processo de Negócio; Modelação de Processos de Negócio; Vista do Processo Específica às Partes Interessadas; Integração de Vistas; Modelo Consolidado; Formulário de Inquérito.

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

BPM Business Process Management

BPR Business Process Reengineering

IT Information Technology

BPMN Business Process Model and Notation

EA Enterprise Architecture

ADM Architecture Development Method

1

Introduction

Contents

This Chapter - **Introduction** - presents the motivation for this work, describing the problem posed and the objectives we aim to achieve, which is based on the following premise: business processes are present in many endeavors in life [7], and represent one of the core resources of organizations. However, regarding the same business process, stakeholders usually have distinct concerns which results in the creation of separate models for the same business process, causing consistency problems or imposing challenges for keeping the multiple models consistent [8].

1.1 Motivation

Business Process Modeling can be used to improve the perception and understanding of business processes, but the way things are done within an organization can be perceived very differently according to the role and position of the person describing it. The different stakeholders, internal or external to the organization, have different aspirations, concerns and points of view regarding a certain business process. That is why a business process can be described and modeled in completely different ways according to the chosen point of view, and this can be the source of management problems and inconsistencies, since due to this fact organizations often tend to produce several models for the same process. Therefore, the production of the business process models often lacks accuracy and keeping them consistent is a demanding task.

Consequently, the goal of achieving a common understanding using Business Process Modeling is difficult to accomplish in the majority of the cases. Even with the wide and powerful environment of notations and tools used to represent and analyze models [9], a common understanding between stakeholders who have different views over the same process is difficult to achieve.

There are two main reasons that may be in the source of these issues that represent the motivation for this work [9, 10]:

- Different process stakeholders belong to different organizational areas, hence they have different concerns and focus on different perspectives of a business process, just like a business process often crosscuts intra and inter organizational boundaries and tends to cross multiple organizational units.

- The specification of a business process is intrinsically tied to its design team, which means that a business process model is a representation of the modeler's perspective regarding a given process. So, different teams will always achieve different specifications for the same process and the assessment to make sure if they are equivalent is complex.

The task performed by the organization of managing multiple process diagrams that are a representation of the same business process may lead to inconsistencies, such as heterogeneous schemes for naming its activities and entities, usage of different modeling styles and process hierarchies with arbitrary depth and level of detail. Also, these inconsistent models are not only prejudicial for the users and stakeholders understanding but also may lead to misleading interpretations of the process content or ignoring relevant information [11].

Another variable encountered in this problem is change. There is continuous pressure on organisations to adopt new technologies, be competitive and revise strategy for their survival [12], thus the importance of Change Management, that has been defined as 'the process of continually renewing an organisation's direction, structure, and capabilities to serve the ever-changing needs of external and internal customers' [13]. Change is constant within an organization and its business actors due to their learning nature and adaptation actions, which means that the executed tasks are being constantly updated and therefore the business processes may suffer alterations. Moreover, when modeling a business process, there is a need to find a balance when producing models: if the business process model is too complex - very detailed and little granularity activities - it will be impossible to keep updated and will become obsolete very quickly, hence becoming useless. If the business process model is too simple - few detail and very high granularity activities - it will also be useless because there is no detail about the way the process unfolds and which specific activities it comprises. Therefore, it is important to make proper use of Business Process Modeling in order for the produced models to be useful for the organization.

Summing up, the task of Business Process Modeling must cope not only with the multiple views and concerns of the different organizational stakeholders, but also try to keep up with the ongoing changes and capture the relationships at many levels within the organization, with the goal of producing business models and keeping them accordingly updated, providing the organization with actual added value.

1.2 Problem Description

The problem this thesis will work on is the integration of different business process models that represent the views from the different stakeholders into a centralized model. It stems from the fact that a business process model does not portray the concerns of every stakeholder, therefore the activity of business process modeling may lead to multiple and sometimes conflicting process models for the same business process. However, the lack of a systematic way for the representation of the stakeholders' heterogeneous process concerns results in a lack of consistency between views and all the issues that arise with that.

A 2017 survey [14] on *Business Process Variability Modeling* gathered the different existing approaches that model the families of business process variants and provided a comparative evaluation amongst them. A common feature present in all approaches is the fact that they are able to support the representation of a family of business processes variants through a single model and each variant can be derived using model transformations. This survey presented the wide variety of approaches to customizable process models, which is a consolidation of process variants, and noted two underdeveloped areas. Firstly, the lack of effective methods and tool support to aid users in the creation, use, maintenance and specifically in the customization of the models. Secondly, the question of how to guide users during the customization of customizable process models has had few solving attempts.

These two shortcomings are an interesting nuance to the broader problem previously described that was related to the lack of clarity and difficulty in keeping the models updated, since they focus on how to help and deal with the stakeholders. Solving the identified shortcomings may lead to making the production and maintenance of the process models simpler and more effective, thus trying to increase the probability of keeping the models updated so that they can actually be helpful for the organization.

To better illustrate the problem posed, we present an illustrative example with the intention of promoting the readers' understanding of the problem. The *Car Repairing Process* [1] is a well known example to illustrate problems of this nature, therefore this scenario will be used throughout this document. The example showcases the same business process but in the perspective of two completely different stakeholders, and then a Consolidated Model that gathers the knowledge of both views. Figure 1.1 presents the stakeholder-specific views for the Facilities and HR department, showcasing the relevant activities for each one.

Figure 1.2 shows a possible Consolidated Model, which represents an enriched model resulting from the integration of knowledge from the two available distinct views.

Having examined both Figures, the more detailed and specific problem this work will try to solve arises. It consists of discussing how one can go from having completely different views, depicted in Figure 1.1, over the same process and end up with a Consolidated Model that comprises all the activities, depicted in Figure 1.2. This discussion will entail how to find the relations between activities needed to build the Consolidated Model, given that the broad process happens in the same organization for all departments and respective stakeholders, so to this extent it is expected the existence of a relation between them.

Therefore, this work will focus on the problem of View Integration in order to achieve a Consolidated Model so that it is possible to tackle the organizational issues previously mentioned.

1.3 Objectives

Given the motivation and the problem description from this work, we will now describe the goals and objectives we aim to achieve. As previously stated, the same business process may have different perspectives regarding each stakeholder's concerns, which may lead to inconsistencies, that we will try to diminish by integrating the different stakeholder-specific views into a centralized Consolidated Model.

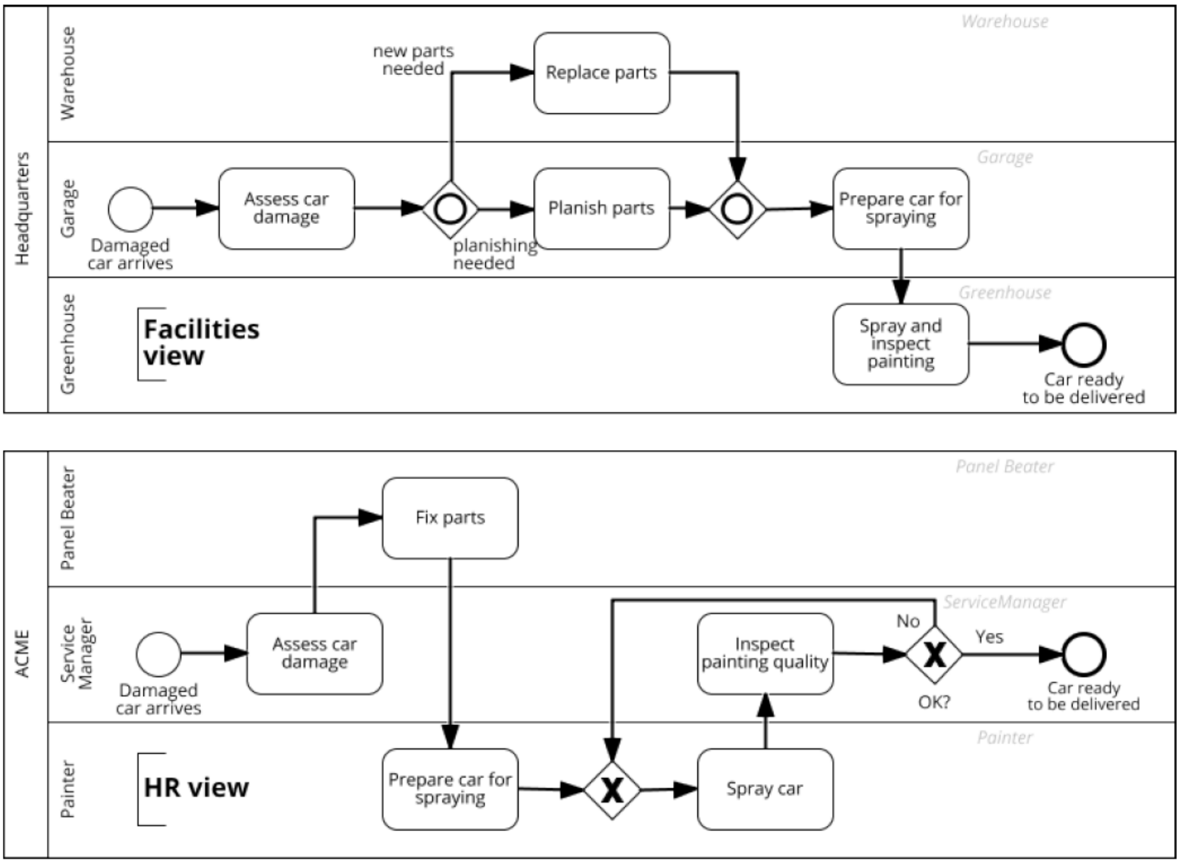

Figure 1.1: Facilities and HR View, adapted from [1]

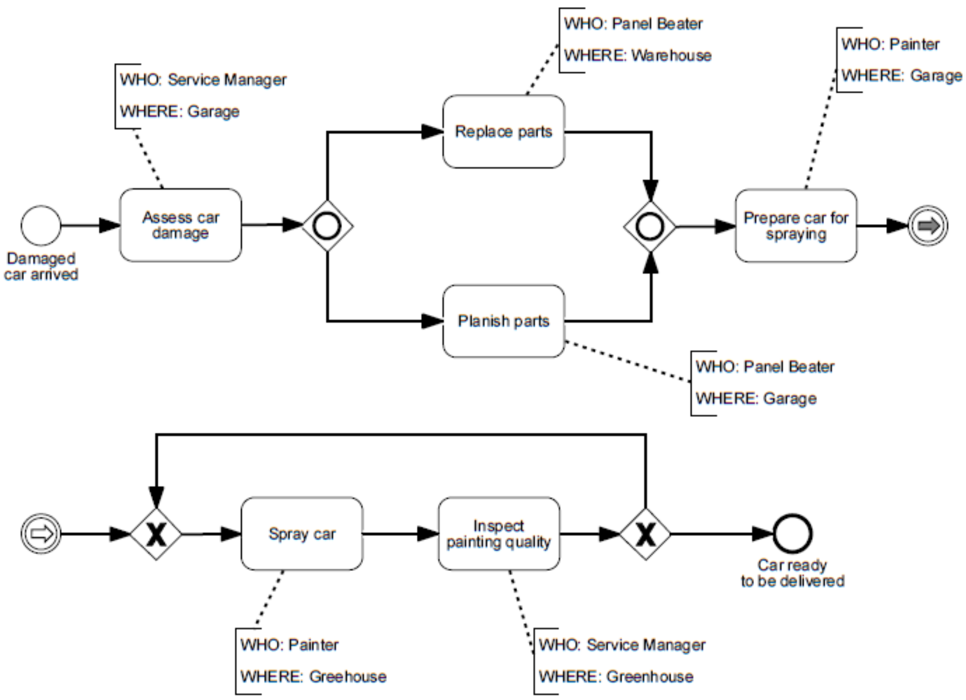

Figure 1.2: Consolidated Model of the Car Repairing Process, adapted from [1]

Goals: This thesis aims to integrate the different stakeholder views into a centralized Consolidated Model in order to find relations of resemblance or composition between the activities of the different views that allow for their integration and further enrichment.

To do so, the present work will discuss what is the best way to fill out the data structures needed to generate the Consolidated Model, defining the most effective approach to integrate different views from distinctive stakeholders. Through this taxonomy enrichment and creation of new ways of inquiring, we aim to help the different stakeholders to express their concerns in a structured and effective manner that will ease the production and management of the models. Hence, we will be able to aid the organization in maintaining a consistency between views by building a Consolidated Model that will be beneficial and add value for them. Afterwards, a demonstration in a fictitious and in real-world motivating scenario is performed in order to show how the solution works by integrating the respective views in order to successfully build a Consolidated Model.

1.4 Research Methodology

This work will follow the *Design Science Research Methodology (DSRM)* proposed by Peffers et. al [2] to conduct the research. The DSRM is an iterative research methodology that focuses on the creation and validation of artefacts that address a research problem. It is divided in the following six phases:

1. **Problem Identification and Motivation**: describe the research problem while justifying the importance of the solution.
2. **Definition of the Objectives for a Solution**: derive the main objectives from the problem identification, taking into consideration what is possible and feasible.
3. **Design and Development**: create the artefact after determining its desired functionality and architecture.
4. **Demonstration**: demonstrate how to use the artefact to solve one or more instances of the problem. This can be achieved through experimentation, simulation, case study, proof or other appropriate activity.
5. **Evaluation**: observe and measure how well the artefact supports a solution to the problem. This involves comparing the solution's objectives with the demonstration results.
6. **Communication**: communicate the problem and its importance, the artefact, its functionality and novelty, the intricacies of its design and its effectiveness to relevant audiences.

The mapping between the process activities and our work is presented in Figure 1.3.

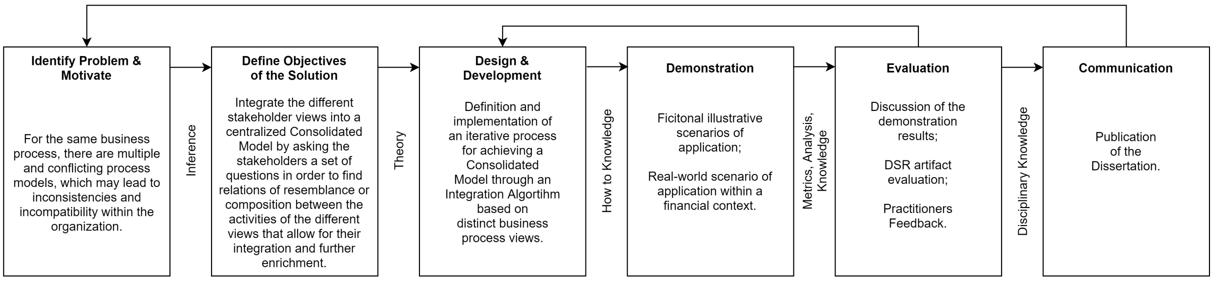

Figure 1.3: DSRM process steps mapped to the phases of this research work, adapted from [2]

1.5 Document Organization

The remainder of this report is structured and mapped with the DSRM in Figure 1.3 as follows.

The present Chapter 1 - Introduction - provides a context explaining the scope of this work, including an introduction, the motivation for this work, the objectives, the problem description and the chosen research methodology. This Chapter corresponds to the first two steps of the DSRM.

Chapter 2 - Background - refers to the background for this work, providing a brief review of the disciplines of Business Process Management (BPM), Business Process Modeling, Enterprise Architecture (EA), among others.

Chapter 3 - Related Work - refers to the *state-of-the-art*, regarding relevant work performed to try to solve the same or similar problems the current work is also trying to solve and some pertinent work that has not been discussed yet and can be incorporated in our solution. It also introduces some limitations and problems that emerged from previous ideas and it ends with a brief analysis of what are the contributions of each of them.

Chapter 4 - Proposed Solution - is where we propose and describe our solution, explaining where it will stand and how it will tackle the identified inefficiencies and issues in order to solve the research problem. This Chapter presents the initial design and ideas regarding the solution, which corresponds to a part of the third step of the DSRM.

Chapter 5 - Demonstration - demonstrates the use of the developed solution in two illustrative scenarios, a well-known and published example and a real-world example within financial context. This Chapter corresponds to the fourth step of the DSRM.

Chapter 6 - Evaluation - describes and substantiates how the solution was evaluated and validated and discusses the results. In this section, some guidelines to follow when evaluating the solution are presented. This Chapter corresponds to the fifth step of the DSRM.

Chapter 7 - Conclusion - finalizes the document and presents the final remarks of this work, summarizing the research work by identifying its advances and limitatios, given that the latter turned into suggestions for future work. This Chapter corresponds to a part of the sixth step of the DSRM.

2

Background

Contents

This Chapter - **Background** - discusses some concepts that are the basis of this research, and that allowed us to develop the present work. It starts with some context about the scope of this research, followed by the introduction of topics such as Business Process Management, Business Process Modeling and Enterprise Architecture.

2.1 Scope

The scope of this work lays on the areas of knowledge regarding business processes, from the different modeling issues organizations face, to managing their lifecycle in a valuable manner. Since we are dealing with organizations, focusing on effectively achieving its current and future objectives is also relevant. Therefore, concepts concerning Enterprise Architecture will also be discussed.

Business processes have a direct impact in the quality perceived by the market of the products and services offered. Business processes, being one of the core resources of organizations, shape the way of working within an organization, by determining tasks, jobs and responsibilities. Business processes integrate systems, data and resources within and across organizations and determine their potential to adapt to new circumstances and to comply with a fast-growing number of requirements [3]. Also, they deeply influence the revenue potential and the cost profile of an organization [3].

2.2 Business Process Management

In an historical overview perspective, one can say that BPM stems from an evolution of ideas in what concerns work organization [3]. Starting with the labor division by Adam Smith and later by Frederick Taylor which led to workers becoming specialists on a specific part of a given business process and consequently to the creation of the *manager* class to supervise them [15]. Then, process thinking entered the picture with the famous case of Ford's acquisition of a financial stake in Mazda in the 80s which transformed into a well-known case study, where the excessive paperwork and the resources it consumed were identified as a problematic performance issue. Later on, Michael Porter's work [16] led to the topic of the importance of aligning the strategic goals of an organization with its business processes, arguing that the strategy of a given organization was closely tied with their value chains. Ultimately, Davenport and Short [17] triggered the concept of Business Process Reengineering (BPR), which focuses on the redesign of core business processes to achieve improvements in productivity. Since then, BPR has been adopted by several organizations and can be perceived as a subset of techniques to be used in the scope of BPM [3].

Nowadays, many businesses are trying to use Information Technology (IT) to improve and manage their business processes. While some of these systems entail incremental process change, others require more far-reaching business process redesign. To cope with these changes, many organizations in this day and age are turning to BPM that provides a variety of tools and methodologies to analyze existing processes, design new processes and optimize them. BPM is a management methodology which was originally defined in 2004 and attempts to systematically translate a firm's strategies into operational targets.

BPM was defined by Dumas et. al [3] as the "art and science of overseeing how work is performed in an organization to ensure consistent outcomes and to take advantage of improvement opportunities", that depending on the business goals and thus business objectives of the organization, can vary a lot. The goal of BPM [18] is adding value to the organization, its customer and stakeholders, hence it aims at managing business processes in the most effective way [3].

2.2.1 Business Process Management Lifecycle

Usually, companies that make use of BPM go through five steps: identify processes for change, analyze existing processes, design the new process, implement the new process and finally continuous measurement [19]. These five steps correspond to the *BPM Lifecycle*, depicted in Figure 2.1, that aims to implement continuous business process improvement into an organization. After an initial step of Process Identification, it consists of Process Discovery, Process Analysis, Process Redesign, Process Implementation and Process Monitoring and Controlling [3]. The five phases of the BPM Lifecycle are described as follows [3]:

- **Process Identification** - For this first phase, a business problem is posed, and the relevant corresponding processes are identified, delimited and related to each other. The outcome of this phase is a new or updated process architecture that provides an overall view of the processes in an organization and their relationships.

- **Process Discovery** - This phase can also be called Process Modeling. Here, the current state of all the relevant processes is documented, usually as one or many AS-IS process models.

- **Process Analysis** - In this phase, all the problems associated with the AS-IS process are identified and documented. The outcome of this phase is a structured collection of issues that are prioritized in terms of their impact and effort required to solve them.

- **Process Redesign** - This phase can also be called Process Improvement. This phase aims to identify changes to the process that would help to address the issues previously identified and permit the organization to meet its performance objectives. The Process Redesign phase is deeply connected with the Process Analysis phase, since as new change options are proposed, they are analyzed using process analysis techniques. The outcome of this phase is a TO-BE process model, redesigned with the most promising change options.

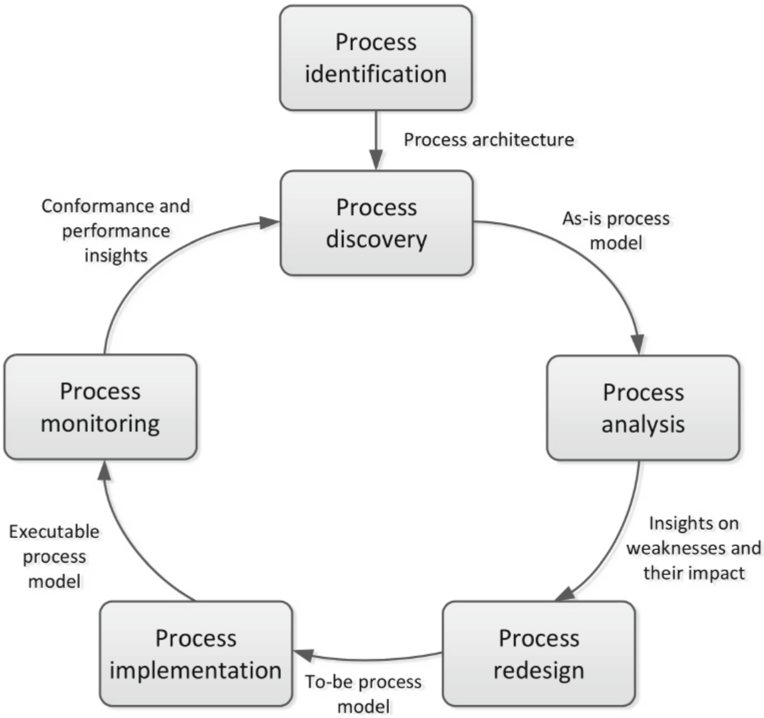

Figure 2.1: BPM Lifecycle [3]

- **Process Implementation** - In this phase, the changes required to shift from the AS-IS to the TO-BE process are executed. This phase covers two aspects: organizational change management and process automation.
- **Process Monitoring and Controlling** - Finally, once the redesigned process is running, this phase monitors how well the process is performing according to its performance measures and objectives. When errors are found, corrective actions are undertaken. In the case of the appearance of new issues, the cycle must be repeated on a continuous basis.

2.2.2 Business Process Management Challenges

One of the greatest challenges posed by BPM is overcoming the organizational culture, since often the organization's way of working is completely embedded within an organization and its employees and an attempt to change it may be a difficult barrier to overcome [19]. However, even if it results in employees resisting change, it is vital for the organization to be kept updated. In that line of thought, BPM is a methodology that can aid organizations to optimize their performance and become more effective and efficient, since it plays a central role at operational, organizational and technological levels [10]. The current work will focus on the early stages of BPM, specially in Process Discovery and Process Analysis.

However, not with the goal of redesigning the processes itself or, in other words, the way of working of the organization, but instead redesigning and enriching the views of the different stakeholders about the same process.

When using BPM methodologies, the complexity of the business process models often increases and it becomes more difficult to manage them. Therefore, it is very useful for an organization to define its Business Process Architecture [20], which can be seen as tool that provides a structured overview of all the processes, their relationships and their boundaries as we will discuss up next.

2.2.3 Business Process Architecture

A Business Process Architecture provides a structured overview and representation of the processes that exist within an organisation and explicits their relationships. It can also be seen as a tool to design a structure for the organizational business processes, in order to aid the organization in their maintenance. As aforementioned, the complexity of the business process models can get out of hand, and is greatly influenced by the level of complexity of the organization itself. To deal with such a panoply of processes with different levels of detail, it is useful to classify them according to different criteria.

For instance, the Porter's Value Chain model [16] divides processes into two categories:

- **Core Processes** - Processes that are essential in adding value and creating a competitive advantage.
- **Support Processes** - Processes that enable the execution of the core processes.

Whereas the relationships between an organisation's business processes can be one of three different types [3]:

- **Sequence** - A logical or temporal sequence between two processes.
- **Decomposition** - A process is specified in greater detail in several subprocesses.
- **Specialization** - A process that can have several variants.

A Business Process Architecture can be divided into the following three levels [3], as depicted in Figure 2.2:

- **Level 1: Process Landscape** - This level shows the main processes on a very abstract level. Each of the elements of the process landscape model points to a more concrete business process on level 2.
- **Level 2: Abstract Process Models** - This level shows the processes at a finer degree of granularity, but still in a quite abstract way. Each element on level two points further to a process model on level 3.

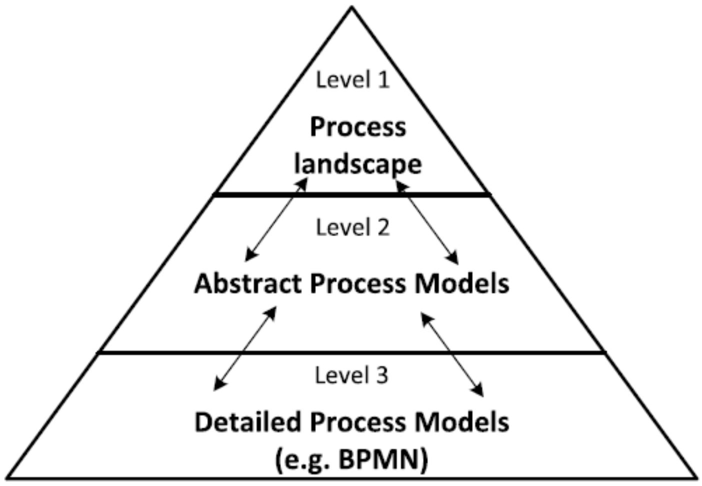

Figure 2.2: Process Architecture [3]

- **Level 3: Detailed Process Models** - This level shows the detail of the processes including control flow, data inputs and outputs, and assignment of participants.

2.3 Business Process Modeling

Business Process Modeling can be seen as the human activity of creating a business process model [21]. A business process model is the product of defining a business process through a Business Process Modeling Language or Notation that provides a predefined collection of components and relationships to direct the business process modeling mission [21].

Designing the processes of an organization, whether they are the firsts to be developed at an initial stage of the life of the company or at a more advanced state of redesigning existing ones by applying BPM, leads to the creation of business process models, which can represent various aspects, views or perspectives of the organization. The most common views include the strategic or organizational view, processes view, information view, application view and technology view [22]. The business process view is one of the most critical enterprise models since it provides an accurate representation of the way the organization functions. However, it is very difficult for the organizations to make use of such models as a repository of organizational knowledge with universal usage because keeping them updated and aligned with organizational reality is a tremendous challenging task.

Business Process Modeling is an approach to portray the way organizations conduct their business processes through abstract descriptions. Through Business Process Modeling, business process views are developed and they represent a fundamental prerequisite for organizations wishing to engage in business process improvement or BPM initiatives. Business processes are designed to achieve specific

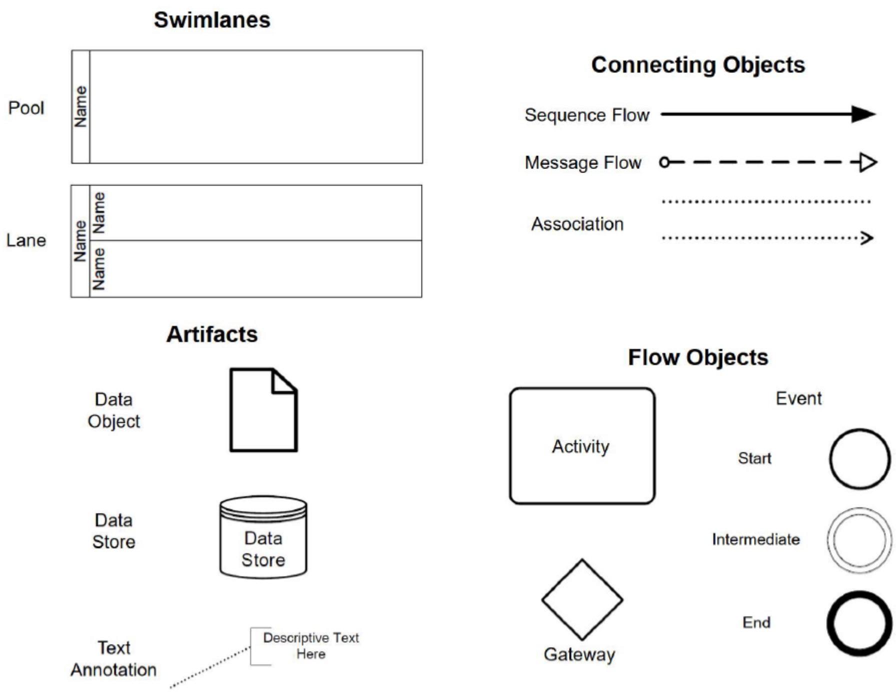

Figure 2.3: Overview of the BPMN Core Elements, adapted from [4]

goals and the task of Business Process Modeling has as top three perceived benefits process improvement, understanding and communication across the different stakeholder groups [23].

Furthermore, the main purpose of Business Process Modeling is the construction of a concise and unambiguous representation of a business, either representing the current way of working, the *AS-IS* or the future one *TO-BE* that the organization hopes to achieve, both crucial to the BPM lifecycle. Current business process analysis tools allow several process modeling notations to be used and most of them allow a series of analysis to be done upon business process blueprints [9].

2.3.1 Business Process Model and Notation

There is a panoply of modeling languages, but for this work we will make use of Business Process Model and Notation (BPMN) [4]. The first version of BPMN was presented as the standard business process modeling notation in 2004 [4]. Ever since then, BPMN has become widely supported in the industry, being thorough evaluated by the academic community.

The original goal of BPMN was to provide a notation that is understandable by all business users, regardless of their function. Another factor driving its growth is that business process models created by business people have traditionally been theoretically isolated from the representations of processes needed by systems designed to implement and execute those processes. Hence, the original process models needed to be manually translated to executions models, which could lead to errors and difficulty in understanding its evolution. Therefore, it was also the goal of BPMN to create a bridge from a visual notation to execution languages.

BPMN 2.0 consists of three diagrams: the business process diagram, conversation diagram, and choreography diagram. We will only be focusing on the business process diagrams specified in BPMN, which can include private business processes, which are internal to a specific organization, and public processes, that represent interactions between two or more business entities.

The BPMN language is grouped into four basic categories of elements [4], as depicted in Figure 2.3, that can be described as following:

- **Flow Objects**: Contain events, activities, and gateways.
  - **Events**: Are defined as something that happens in a process, and on one hand the way the process responds to this, if it is a catching event, or on the other hand how the process generates a signal that something has occurred, if it is a throwing event. Events are either start events, intermediate events, or end events.
  - **Activities**: Are divided into process, subprocess, and tasks. They denote the work that is done within a company.
  - **Gateways**: Are used for determining branching, forking, merging, or joining of paths within the process.
- **Connecting Objects**: Are used for connecting the flow objects.
  - **Sequence Flow**: Defines the execution order of the activities within a process.
  - **Message Flow**: Indicates a flow of messages between business entities prepared to send and receive them.
  - **Associations**: Used to associate both text and graphical non-flow objects.
- **Swimlanes**: Used to denote a participant in a process and represent a graphical container for a set of activities taken on by that participant.
  - **Pool**: Represents a business process participant.
  - **Lane**: Sub-partitioning of a Pool, representing a sub-category of the process participant represented in the Pool.
- **Artifacts**: Contain data objects, data stores, groups, and annotations.
  - **Data Object**: Provide information on resources required or produced by activities, and usually represent documents used in a process, both in physical and digital form.
  - **Data Store**: Used to store information in a more permanent way.

– **Group**: Visual aid used for documentation or analysis purposes.

– **Annotation**: Used to add additional information about the model.

2.4 Enterprise Architecture

There are several definitions of Enterprise Architecture (EA), which are dependent on organisational and application aspects [24]. For the Open Group [25], EA consists of defining and understanding the different elements that make up the enterprise and how those elements are inter–related. It can also be defined as a ”a coherent whole of principles, methods, and models that are used in the design and realization of an enterprise's organizational structure, business processes, information systems, and infrastructure” [26] and its purpose is to capture the essentials of the business while providing an holistic view of the enterprise. EA can be positioned in the context of managing the enterprise and assist in coping with the complexity of ensuring the alignment between the business and IT. This can be very useful to understand the purpose and context of the activities when trying to discover process similarity.

In the course of this document, various EA concepts are mentioned and used to define the proposed solution. Therefore, the definition of the terms relevant for the understanding of the remaining of this document, which is based on the International Standard ISO/IEC/IEEE 42010:2011 [27], is the following:

- **Stakeholder** - “individual, team, organization, or classes thereof, having an interest in a system”.
- **Concern** - “interest in a system relevant to one or more of its stakeholders”.
- **Architecture** - “fundamental concepts or properties of a system in its environment embodied in its elements, relationships and in the principles of its design and evolution”.
- **Architecture View** - “work product expressing the architecture of a system from the perspective of specific system concerns”.
- **Architecture Viewpoint** - “work product establishing the conventions for the construction, interpretation and use of architecture views to frame specific system concerns”.
- **Architecture Description** - “work product used to express an architecture”.

There are many EA Frameworks available, and each organization should pick whichever better fits its needs and take relevant pieces of each one of them to implement onto their processes according to their own business objectives. We emphasise TOGAF [25] and the Zachman Framework [28], but there are many others. The Zachman Framework is the one closely related with our work, therefore we will discuss it thoroughly in 3 in parallel with the literature that used it as a baseline.

TOGAF [25] is a process-driven set of supporting tools for developing an enterprise architecture. The set of steps required to develop the architecture is described by the Architecture Development Method (ADM). This EA Framework was designed to support the four architecture domains that commonly constitute an enterprise architecture:

- **Business Architecture** - Defines the organizations' strategy, governance and business processes.
- **Application Architecture** - Plans the systems to be deployed, their interactions and their relationships to the core business processes of the organization.
- **Data Architecture** - Describes the organization's logical and physical data assets and data management resources.
- **Technology Architecture** - Describes the logical software and hardware capabilities that are required to support the deployment of core applications.

3

Related Work

Contents

This Section - **Related Work** - corresponds to the first three steps of the DSRM process, namely "Identify Problem Motivate", "Define Objectives of a Solution" and "Design Development". It presents a literature gathering and review of several contributions regarding the topics related with solving this problem, such as the Zachman Framework, Annotations, View Integration, View Generation, Process Equivalence and Process Mining, and ends with an analysis of its relevance to the proposal of this work.

3.1 Zachman Framework

This Subsection describes and analyses the contributions that used Zachman Framework to support their research, starting with a brief contextualization of the Framework itself.

The Zachman Framework was initially proposed by John A. Zachman but has suffered many updates ever since. The usage of some type of logical architecture for defining the interfaces and the integration of all the components of the system was triggered by the increasing scope of design and complexity of information system implementations [6]. The Zachman Framework is a framework for enterprise architecture that proposes a six by six matrix-like representation to classify descriptive representations relevant for describing an enterprise.

Sousa et. al [5] described a method to infer business activities in order to facilitate the consistent representation of business processes, that relies on using a number of properties derived from the dimensions of the Zachman Framework. We retrieved the definitions for the dimension columns from that work and present them in Table 5.1. The definitions for the rows are retrieved from [6] and are presented in Table 3.2.

The Zachman Framework is a schema that consists in the intersection between two historical classifications [28]:

- The fundamentals of communication found in the primitive interrogatives: What, How, When, Who, Where and Why as columns - check Table 5.1.
- The ratification transformations of an abstract idea into an instantiation: Identification, Definition, Representation, Specification, Configuration and Instantiation as rows – check Table 3.2.

Furthermore, the cells of the matrix are an intersection between the interrogatives and the transformations, or in other words, the product abstractions and the stakeholders perspectives, respectively, resulting in the framework classification. Looking at the matrix as a whole, it would necessarily constitute the total set of descriptive representations that have relevance for describing an enterprise. More specifically, the Zachman Framework is an ontology, which means it represents a theory of existence of a structure of vital components for which explicit expressions are necessary or mandatory for operating the subject. This framework can be seen as a metamodel, the basis for the enterprise architecture [28]

Table 3.1: Zachman Framework Dimension Columns [5].

Dimension	Focus	Purpose
What	Data	The enterprise's information and its way of usage.
How	Function	The process of translating the mission of the organization into its business, definitions and operations.
Where	Network	The geographical distribution of the organization's activities and artifacts.
Who	People	Who is related with the major artifacts of the organization: business processes, information and IT.
When	Time	How each artifact relates and evolves with the timeline.
Why	Motivation	The translation of goals into actions and objectives.

Table 3.2: Zachman Framework Dimension Rows [6]

Abstraction Level	Perspective	Stakeholder Example
Identification	Executive	Planner
Definition	Business Management	Owner
Representation	Architect	Designer
Specification	Engineer	Builder
Configuration	Technician	Implementer
Instantiation	Enterprise	User

and provides a taxonomy for relating the concepts that describe the real world to the concepts that describe an information system and its implementation.

Sousa et. al [9] analyzed the Zachman Framework and proposed a rule for activity decomposition, in which each activity α can be decomposed into two or more distinct discrete activities if and only if one the conditions stated in Table 3.3 is satisfied [5]. In [9], a Business Process Modeling approach based on the aforementioned Zachman Framework Dimensions was proposed, in which the authors stated a possibility of reduction of the number of different blueprints through the use of dimensional process equivalence. In this approach, the authors made use of the Zachman's Framework to recursively characterize the business processes. This approach has some similarities to the one used in data modeling, in which two entities are equivalent if they have the same properties, regardless of their name. So, it considers some properties of the framework itself [29], such as classification, recursiveness and cell uniqueness. Classification states that every artifact of the organization can be uniquely classified, recursiveness means that the framework can be applied to further specify the contents of each cell and lastly cell uniqueness means that each cell must be described with the sufficient level of detail so that it accomplishes its purpose [9].

The authors in [9] use the aforementioned rule to define business process equivalence using ac-

Table 3.3: Criteria from activity decomposition [5]

Dimension	Criteria
What	α is composed by two or more activities which receive/create different data entities.
How	α is composed by two or more activities which are processed using different applications.
Where	α is composed by two or more activities which occur in different locations.
Who	α is composed by two or more activities which are managed by different business actors.
When	α is composed by two or more activities which are performed in distinct periods of time.
Why	α is composed by two or more activities which exist to serve different purposes.

tivity equivalence. Firstly, each notation element is mapped into one of the six Zachman Framework dimensions and for each dimension, a hierarchical tree of concepts is defined. Then, the authors define activity equivalence in the following way: ”An activity (A) is dimensional equivalent to another (A') when they have no different when, what, where, who and why” and by using it and acknowledging that the activities are children of a given process, a process (P) is dimensional equivalent to another (P') if all their children are dimensional equivalent. Colaço and Sousa [30] went a step further and made an association between Business Process Modeling notations and the Zachman Framework dimensions, stating that an Event corresponds to When, an Activity to How, a Gateway to Why, a Swimlane to Where or Who, and a Data Object to What, based on the relationships identified in [10].

Nevertheless, the Zachman Framework does not specify techniques for process modelling or decomposition so this results in the framework being independent of specific methodologies, but it is a good starting point for our work.

This is a good starting point to extract a baseline for this work, because topics such as activity decomposition and equivalence, business process equivalence, business process definition, among others were discussed and present great relevance for the construction of the proposed solution. The association between the BPMN elements and the Zachman Framework Dimensions is also quite interesting, since it represents the core associations that will constitute part of the first group of questions that will have to be posed to the stakeholders in order to obtain information to construct the Consolidated Model.

3.2 Organizational Taxonomy

This Subsection describes and analyses the contributions that discuss Organizational Taxonomy to deal with the problem of lack of communication and understanding between the stakeholders within an organization.

Pereira et. al [31] state that an Organizational Taxonomy defines a controlled vocabulary which aims to be understandable by all the process stakeholders and hence provide a common language for communication. It consists of an hierarchical collection of terms that help to structure, classify and represent all the concepts and relationships of a business process while enabling a common agreement in the community to use the same terms in the same form. It should be noted that the definition of the classification structure varies according to the organization.

Based on the Zachman Framework, the authors present a taxonomy based on the categorization of each concept instantiation of a business process, as depicted in Figure 3.1.

Figure 3.1 depicts the organizational taxonomy which is a representation of the hierarchical classification of the concepts used to represent business processes. With a clear taxonomy, in theory all the people within the organization would use the same concepts to represent the same things, and thus

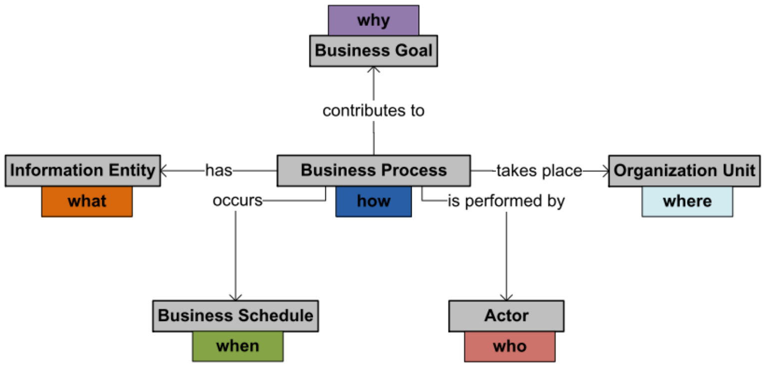

Figure 3.1: Organizational Taxonomy

decrease inconsistencies.

The concepts present in Figure 3.1 are the core notions one should have when trying to describe an organization, which are Business Process, Business Goal, Information Entity, Business Schedule, Actor and Organization Unit. The authors define a business process as a set of connected activities (how) which consumes and produces tangible or intangible artefacts (what), is performed by people or systems (who), contributes to achieving goals (why), takes place in a specific location (where) and during a specific period of time (when). These concepts are once again related with the Zachman Framework Dimensions.

For each concept present in Figure 3.1, a taxonomy based on the categorization of each concept instantiation in the classification structure is created. This classification structure is not steady and directly depends on the functioning of the organization. This is why there is a need to define the ontology, i.e. a formal explicit description of the concepts, its relationships and properties, within a domain of discourse, that should be applied to each concept.

The authors proposed a controlled vocabulary grounded on six dimensions of inquiry that is organized as a taxonomy that allows for the hierarchical creation of an ontology that describes the specific domain of the organization. This proposal is of great relevance for this work, because it is not only connected with the Zachman Framework Dimensions but addresses a very important issue already discussed, the lack of common language stakeholders and the process designers, resulting in a significant gap between different modelling perspectives.

3.3 Annotations

This Subsection describes and analyses some contributions that used Annotations, explaining its benefits and the several fields where they are used.

Annotations are an extent to an entity providing information to add extra detail or observations. They are very useful and have specific uses in a wide range of areas such as biology, programming languages, book publishers market, among others. In biology, Stein [32] ensures the importance of annotations by explaining that they fill the gap between sequence and biology of the organism. In book publishers, often the annotated edition of a book containing marginal comments to explain or clarify some aspects is published. In the law domain, the publishers disclose books with annotated versions of legislation that provide extra information about its interpretation. In programming languages such as Java, annotations can be made to classes, methods, variables, parameters and Java packages and added to the source code, being used to provide supplement information about the given program. Regarding software modeling such as UML, the annotations are used as comments to explain the different elements and their functionalities, and are usually attached to the modeling element.

Yang et. al [33] discuss virtual learning communities (VLCs) as a prime example of collaborative learning. VLCs are knowledge based social entities where the collaborative effort made by all the people involved is vital for their continuous growth. The collaborative learning is achieved through intensive interactions and exertion of knowledge effect. One of the several capabilities of the participants to manage their knowledge is, among many others, responding or adding comments to messages or articles posted by others. This capability meets one of the three perspectives provided by the ontology enabled annotation and knowledge management system proposed by the authors: personalized annotation.

Yang et. al [33] also state that annotations can benefit the learning process in what regards attention, discussion, organization and indexing. Attention since annotations catch the students' eye and help them focusing on annotated concepts, discussion since it eases the argument over assignments based on each aspect in an efficient manner, organization since it reminds the students of important concepts and helps them create their knowledge based on annotation and finally indexing that eases the personalized knowledge discovery in view of information retrieval.

Becker-Kornstaedt [34] states that using annotations avoids lost of experience gained in process execution and states that the majority of knowledge gained by the process performers is often lost if not captured immediately and there is a lack of adequate methods to capture this type of knowledge. The authors present the concept of an electronic process handbook that supports users to incorporate their personal experience through annotations. Process handbooks are created by process engineers and are used in most organizations for handling process models. The authors state three benefits in recording the annotations systematically in the context where the experience was gathered: the experience gained can be specified as an annotation in the needed context, the experience can be integrated into the process description itself to provide support in the next time the process is executed and finally the analysis of the experience can be used across projects for project improvement, which is specially important for model maintenance. Therefore, the lessons learned from executing processes are applied in the incremental change aimed by systematic process improvement in the form of annotations. However, as previously discussed, process models are often outdated and sometimes even obsolete due to the process models or the process itself, so they are usually not optimally suited to support process performance. Thus, a mechanism to systematically capture annotations is needed. The authors propose a electronic process handbook extension that allows to attach annotations, making it possible to provide supplementary process information to others. By using this concept, process performers can systematically capture process knowledge and process engineers can incorporate it into process models for process model maintenance.

Castela and Tribolet [35] created a continuously updated business process model that uses the concept of annotation to create interactions with the stakeholder to both make explicit and communicate their concerns about the processes and to discuss the existing business processes representations. Their work was aimed to prove that the organizational actors, provided with a process and a supporting tool, could act as active updaters and modelers of the business process models of the organization, by stating their concerns regarding their work and comparing existing representations with actually executed activities. They state that the organizational actors could not only continuously monitor their activities but also to propose updates to its representation.

The authors [35] establishes a process to continuously update the AS-IS process model, called *PROASIS*, in a collaborative way through its stakeholders that can suggest corrections and updates to the process. This is an attempt to ensure the alignment with the organizational reality and to provide the basis for the development and maintenance of organizational self-awareness. For the stakeholders to do so, it is proposed an annotation mechanism that can be either textual or graphical. The fundamental concepts used in the construction of the *PROASIS* include organizational resource, organizational actors, roles, activity, business processes, orthogonal dimensions of representation, context, enterprise model and organizational unit.

The *PROASIS* is executed by the organizational actors that share a common representation of business processes and use annotations as a way to build updating proposals that aim to perform corrective maintenance to either correct or increase the detail of the model with the final goal of aligning it with the reality perceived by each organizational actor. It is produced as follows: after making an annotation on a modeling element, a negotiation with the actors that share the same context is performed in order to clarify the original purpose of the annotation, where all the actors either agree or disagree with the annotation made. Then, an intermediate negotiation activity is performed, called review, and finally the annotation should be evaluated by the actors that possess some degree of responsibility. In the end, if the evaluation ends in approval, the change requested could be incorporated in the new version of the process model.

In the survey [14] previously discussed in this work, *Element Annotation* is one of the four mechanisms considered to handle the problem of process customization, being the other three Node Configuration, Activity Specialization and Fragment Customization. The way that the approaches in the Element Annotation capture variability is exactly via annotations attached to elements of the customizable process model, who link an element in the customizable process model to an element in a domain model. The main approaches that fall in this group rely on graphical annotation of model elements with properties of the application domain.

The considerable importance of Annotations in many fields led us to incorporate it as a stakeholder aid in the proposed solution, by allowing them to freely write about the model in a specific question posed for that purpose. This way, we can identify if there is a type of question missing, if a question is being misunderstood and overall give freedom to the stakeholder to express his specific concerns about a given activity that maybe we were not aware of its relevance or it is difficult to pose in a more standard question format. Annotation will also be used to showcase the relations found between activities in the Consolidated Model.

3.4 View Integration

This Subsection describes and analyses the contributions that discuss View Integration. View Integration is the process that aims to combine different specific views of the same business process into a consolidated one.

Colaço and Sousa [30] proposed a method for integrating business process models into a single Consolidated Model. The *View Integration Method* is supported by a business process repository and works as follows. It starts with the identification of a modeling need, followed by the modeling of the view of a given process. Then, that view is uploaded into the repository and the view elements are classified by the stakeholders which allows for the creation of an organisational taxonomy. Since this method is iterative, this sequence of steps is repeated in the context of another view of the same process which will allow for a to become more detailed in each iteration. The process ends with the generation of the Consolidated Model by the repository based on the information introduced by the stakeholders.

Navathe and Schkolnick [36] present a conceptual framework for logical database design which was later improved by Navathe and Gadgil [37], which is divided into four main phases: View Modeling, View Integration, Schema Analysis and Mapping, and lastly Physical Schema Design and Optimization. We emphasize the first two, since they are also present in our approach. The authors [37] specifically focus on the second step, View Integration, which is defined as the phase where the user views are combined into a global model of the data and any conflicts in the process are presented for resolution. This work uses the Navathe and Schkolnick model [36] as a vehicle for modeling user views and discusses the several problems related to the view integration process and proposes an approach to the development of a software system for automating the construction of integrated views.

Navathe and Gadgil [37] came to a few interesting conclusions, and we highlight the most important ones, given the problem description of the current work. It is stated that the success of view integration is largely dependent on getting as explicit input from the stakeholders as possible, given that with a higher level of detail the model will be more likely to better represent the reality. However, this can be a tough task for the designer and the authors pose a problem that is deeply related with the current work: it is not obvious whether the designer should do an ad-hoc analysis of user views to detect equivalences and then specify them on its own or whether he should expect that the different areas are aware of the differences in their data necessities. Also, in the conflict resolution area, it is declared that considerable human involvement is necessary since ultimately the responsibility to resolve the issues is up to the users and the management. Lastly, a very compelling conclusion that we can take from this paper is that a good approach for performing view integration is to have the used machine dealing with a large number of integration alternatives and present them to the user in order to get some feedback.

Mendling and Simon [38] also discuss this topic and focus on the fact that the need to consolidate the knowledge of the different stakeholders in the design of business processes is utilized only to a limited extent by the existing modeling methodologies. This shortcoming is addressed in this paper where a method for business process design by view integration is proposed.

The authors state [38] that the task of capturing and designing real-world process semantics implies a considerable complexity and calls for a structured approach. View integration is a classical technique for database design and its main idea is to firstly identify the different views on every person's data, then each of them is interviewed and the view is documented in a separate input schema where the matching parts are identified and based on these matches, the integrated schema is derived as a merge of the input schemas. In the context of Business Process Modeling, this approach has not been much successful due to two main reasons. Firstly, the conceptual differences among process models and data models lead to non direct application of the database schema integration, and secondly the specific techniques for behavior integration have been defined for Petri nets but not for conceptual languages. Therefore, an analogous approach to data schema integration which addresses the specific details of business process models is needed. The authors also discuss process equivalence, which we will get into more detail in Subsection 3.6.

As this work will focus on how to get from different stakeholder-specific views to a Consolidated Model, the topic of View Integration is extremely important and was very helpful by providing guidance for the proposed solution.

3.5 View Generation

This Subsection briefly discusses View Generation because it has relevance for the proposed solution, since after the generation of the Consolidated Model, the stakeholder-specific views will be generated once again, through activity enrichment since relevant detail can be added. Even if the generation itself cannot be produced due to time constraints, a relation between the specific views and the Consolidated Model will be defined.

Artur Caetano et. al [8] proposed a process view generator that produces diagrams according to different concerns. This view generation capability is provided by a tool that generates dynamic views from a business process repository that works as a knowledge base. The generator is composed of three main logical components: the repository model, the controller and the viewer. The repository model is based on the six Zachman Dimension communication questions as independent concerns for the decomposition of a business process. The controller specifies the viewpoint used to produce the view, since it manages not only the dimensions but also the level of detail of the generated view. The viewer component allows for producing multiple visualizations based on the same model and on the same viewpoint specified in the controller, presenting the results extracted from the repository.

Pichler and Eder [39] also presented their work on this topic by introducing a model and an architecture that allows to capture arbitrary process perspectives that can be further used for generating process views. Their work is based on part of the problem previously described, which consists in the difficulty for the various stakeholders to get focus in their areas of interest. The concept of process views arises once again emerges as an attempt to solve this problem. The authors state that most process view-related research publications solely focus on control flow issues but they do not show how view-relevant control flow elements are selected corresponding to specified characteristics, such as behavior (control flow), function, information (data), organization, and operation. The authors have a more analytical purpose in what concerns generation of business processes, based on queries which formulate combinations of constraints on diverse perspectives.

Cardoso and Sousa [1] found a knowledge gap in the work developed by Artur Caetano et. al [8], which was the fact that it only focused on describing a conceptual tool without formally defining the algorithms that support the generation of the views. Moreover, it does not apply the problem to a specific process modelling notation, like BPMN. The authors tried to fill this gap by developing a generation algorithm for generating stakeholder-specific views in BPMN.

The view generator proposed by Artur Caetano et. al [8] was presented and can be very useful for our solution. Even though view generation is out of the scope of this work, there is still a step of view enrichment after the production of the Consolidated Model that benefits from this insight.

3.6 Process Equivalence

This Subsection describes work regarding Process Equivalence, a topic that portrays interest for our solution in what concerns the production of the Consolidated Model by finding equivalence when comparing activities.

Many authors have worked on different notions of equivalence, but few regarding the comparison of process models. Thus, most equivalence outcomes are binary, and this is not helpful since in real organization processes, sometimes there is need to define a degree of equivalence. Also, not all parts of a process have the same importance for the organization, therefore it should be defined which parts of the process will be compared because some of them being equivalent may not mean the whole process is.

Aslst et. al [40] proposes a new way of comparing processes, based on their behavior, thus quantifying equivalence. The authors propose to compare two processes on the basis of some event log containing typical behavior, called exemplary behavior. This exemplary behavior can be obtained through real process executions, user-defined scenarios or simulating one of the two models. They used Petri Nets to implement this approach, but state that it could be applied to other models with executable semantics, such as BPMN, and use concepts as fitness, precision and recall to develop their solution.

Mendling and Simon [38] specifies a method for business process design by view integration starting from two views of a process as input, using EPCs. The business process designer has to identify semantic relationships in terms of equivalence and sequence between functions and events of the different models. Thus, a Merge Operator is introduced and works as following. Firstly, it taked two EPCs of the same business process plus a set of identified semantic relationships as input and produces an integrated EPC. Then each pair of nodes which describes the same real-world events is merged into a single node and the former input and output arcs are joined and split with and-connectors. Finally, for each pair of nodes in the sequence relationship and and-split is inserted after the predecessor node by an arc to a new and-join before the successor node. The authors [38] also identifies restructuring reduction rules to simplify the integrated process model so that the EPC only contains the absolute necessary structure.

In what concerns Process Equivalence, the comparison considering behavior [40] can be interesting and we will also consider the merge operator, the formalized semantic relationships between different process models and the reduction rules defined by Mendling and Simon [38] in our solution approach.

3.7 Process Mining

This Subsection goes over Process Mining techniques, since Process Mining is able to produce models from information, which is related with our topic.

Process Mining is inserted both in the fields of computational intelligence and data mining, and also process modeling and analysis. It consists in discovering, monitoring and improving real processes by extracting knowledge from event logs available in Information Systems [41]. All Process Mining techniques' starting point is an event log which is a sequential recording of events in which each event refers to an activity and is related to a particular process instance [41].

There are three main types of Process Mining: discovery, conformance checking and enhancement. Discovery takes an event log and produces a model without priori information. Conformance Checking compares an existing process model with an event log of the same process. Enhancement improves an existing process model using information about the actual process recorded in an event log. All three tecniques can take as input an event log, but only Conformance Checking and Enhacement can take a model. Discovery outputs the model, Conformance Checking outputs a diagnostics and Enhacement outputs a new model [41]. The authors [41] propose 6 Guiding Principles to prevent users from making mistakes when applying process mining: GP1 - Event data should be treated as first-class citizens, GP2 - Log extractions should be driven by questions, GP3 - Concurrency, choice and other basic control-flow constructs should be supported. GP4 - Events should be related to model elements, GP5 - Models should be treated as purposeful abstractions of reality and GP6 - Process Mining Should be a Continuous Process. Despite the applicability of process mining there are some important challenges that need to be addressed, such as cleaning event data, dealing with complex event logs that have diverse characteristics, improving the representational bias used for process discovery, cross-organization mining, among others [41].

The problem of only being able to handle some cases of routing constructs and thus producing unsound models that do not cover all the traces in the event log suffered a solving attempt by Liesaputra et. al [42]. The authors propose a technique based on using pattern recognition called *Maximal Pattern Mining (MPM)* that can handle loops, duplicate tasks, non-free choice constructs and long-distance dependencies. Hence, the discovered models are easier to understand.

Guido Schimm developed a mining tool [43] suitable for discovering hierarchically structured workflow processes. His work was related to mine workflow models from workflow logs, in a block-oriented representation language. The main advantage the author poses is that using a block-oriented is the direct relation with process algebra and thus with Sequence, Parallel, Alternative, and Loop building blocks.

Regarding Process Mining, using an Enhacement technique from the three available ones described in [41] may be very useful because it can take as input a model and output a new model based on information about the actual process, which we will be able to gather from the stakeholders. This type of input and output are very convenient for our work since they are the same type as we expect to have in the solution we will try to develop. Pattern Recognition [42] to detect duplicate activities can also be useful for our solution.

4

Proposed Solution

Contents

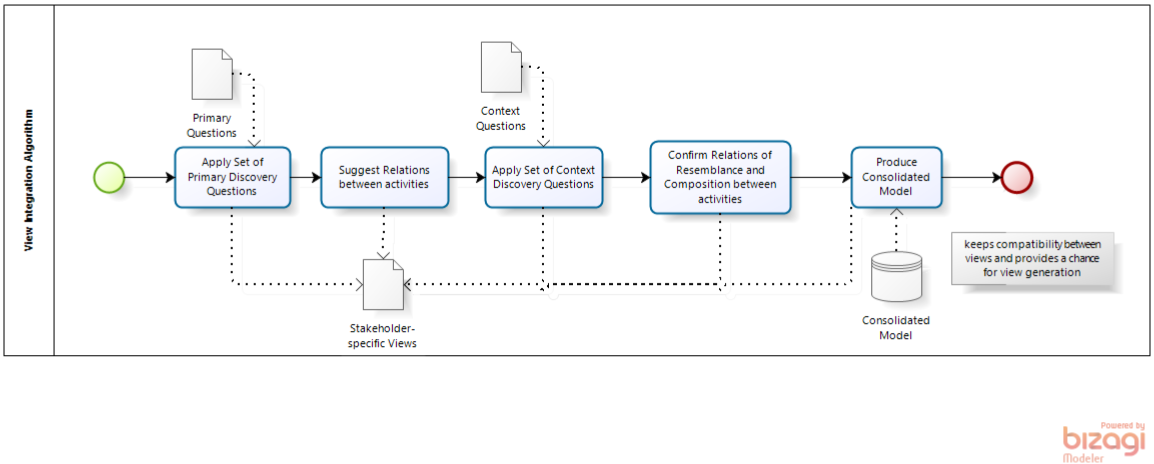

Figure 4.1: View Integration Algorithm

This Chapter - **Proposed Solution** - covers the third activity of the DSRM, Design Development, as it describes the proposal to address the research problem. We will start by briefing explaining our solution through a Big Picture 4.1 overview, and then we will get into the details of each building block of the solution in the Architecture and Design 4.2 section. Finally, we will discuss the View Integration Algorithm that is the basis of our solution in the section Integration Algorithm 4.3.

4.1 The Big Picture

In this section, we will discuss the proposed solution at a broader angle, allowing for an end-to-end understanding of our proposal. We aim to explain the different building blocks of the solution and how they communicate with each other to achieve the goals described in Section 1 that we set ourselves to accomplish in order to bridge the problem also described in Section 1, that we shall recall.

The problem this thesis aims to resolve is how to integrate the different stakeholder-specific views into a centralized Consolidated Model. Therefore, let us recall the goals for this work:

> *Goals*: This thesis aims to integrate the different stakeholder views into a centralized Consolidated Model in order to find relations of resemblance or composition between the activities of the different views that allow for their integration and further enrichment.

Figure 4.1 unfolds how the proposed solution works, showing the view integration algorithm as a process for better understanding. The activities that appear in this process compose the three main building blocks of this solution:

1. **Relations Discovery Questions**: Series of questions to be asked to the stakeholders in a form format, to gather the knowledge they possess on activities of process views in the midst of being merged.

2. **Relation between Activities**: Connections of resemblance and composition between activities from different views to be discovered upon analysing the answers provided by the stakeholders concerning the set of questions presented.

3. **Production of the Consolidated Model**: Creation of the Consolidated Model, based on the Relations between Activities found, allowing for a compact view of the business process.

The methodology behind this solution works as follows. For a given business process of a given organization, the starting point is the different stakeholder-specific views available for that same business process. The following step is to find business process equivalence in order to have a possibility of view integration, which we implemented through the business processes' constituent activities equivalence. In order to discover equivalent activities, a set of Relations Discovery Questions regarding their properties, characteristics and context were posed for the stakeholders to answer. Given their answers, relations between activities will be extracted with the goal of finding resemblance and composition connections between them. Finally, considering the relations found, it is possible to produce a Consolidated Model that integrates all of the stakeholder-specific views and also offers a possibility of enriching them.

Summing up, we aim to integrate the stakeholder-specific views by discovering resemblance and composition relations between activities of different views, enabling the construction of a Consolidated Model and the enrichment of the stakeholder-specific views. Through the application of this methodology, we intend to solve the aforementioned problem and fulfil the goal we set ourselves to achieve.

4.2 Architecture and Design

In this subsection, we will discuss the Architecture and Design of our solution, addressing the assumptions made, the input and output of the solution, its building blocks and the algorithm used to implement them.

4.2.1 Assumptions

For the solution proposed by this work, we made the following assumptions:

- A business process model should be available for every stakeholder-specific view that wishes to endorse in the integration process.

- The business process models should be represented in BPMN [4].

- There is a function that allows to load the algorithm with the activities belonging to the business process models.
- The view element classification algorithm by Colaço and Sousa [30] works seamlessly and correctly classifies all elements.
- There is a sweeper algorithm whose job is to go through the business process models and fulfil the form's answer options with the corresponding BPMN elements.

4.2.2 Input/Output

The stakeholder-specific views work both as the input and output of the Consolidated Model, being the starting and the ending point of the solution. Therefore, the first step is to load them into the Process Repository.

As previously stated, the same business process may have different perspectives regarding each stakeholder's concerns, which may originate different views for that same business process. Therefore, the stakeholder-specific views operate as input, which means they represent the specific view that each particular stakeholder has from a given *AS-IS* business process of the organization.

With the goal of integrating stakeholder-specific views, after receiving them as input, the solution proceeds to its following steps of inquiring the stakeholders and establishing relations between activities from their answers. The production of the Consolidated Model in accordance with the relations found is the step where it is possible to become aware of the business process model as a whole. Afterwards, the same stakeholder-specific views who initiated this mechanism, can be enriched with the same relations used to build the Consolidated Model, being them also the output of the solution.

The stakeholder-specific views that are the input for this solution, the Consolidated Model produced at the end and the enriched stakeholder-specific views also produced at the end are all represented in BPMN [4].

A secondary structure, named Secondary Relation Structure, embodied by an organizational taxonomy tree, is also part of the output of this solution. Not as a main component but as a complementary structure that enables the stakeholders to become aware of the type of relation a pair of activities has. This supplement will be further detailed in 4.2.4.

4.2.3 Set of Relations Discovery Questions

Having received the stakeholder-specific views as input, the following step corresponds to the application of the first building block of this solution.

The first building block will be a set of Relations Discovery Questions that will be presented to each stakeholder in the format of a form, with the goal of building a Consolidated Model and thus enriching the stakeholder-specific views. Some of the questions that compose the set of Relations Discovery Questions were chosen to gather information on the most relevant properties to characterize an activity, thus allowing for a comparison among them based on the answers provided for each property. The remaining questions are focused on the context of the business process as whole, comparing activities that belong to different stakeholder-specific views. We will now discuss these two groups of questions, what characterizes them and how they are meant to function collectively.

The set of Relations Discovery Questions is divided into two groups, the **Primary Questions**, which is divided into **Early Questions**, **Detail Questions** and **Responsibility Questions**, and the **Context Questions**. Namely, the answers gathered for the first group influence the questions of the second group. These two groups differ on its constitution and functioning and have distinct goals, being the combination of the two the building block referring to the set of Relations Discovery Questions:

- **Primary Questions** 4.2.3.A - Questions posed for a single activity at a time, concerning its characteristics, properties and immediate environment.

- **Context Questions** 4.2.3.B - Questions posed for a pair of activities at a time, not just looking at the activities by themselves but together with the flow and the context of the process as a whole.

The goal of this first building block is to extract information about the activities, thus discovering relations of resemblance and composition between them. Our approach comprises new ways of inquiring and the extraction of evidence goes beyond the well-known six dimensions that represent the foundation for this work.

It ought to be noted that the questions will be asked for the activities in particular, not for the process as a whole. Therefore, we will be evaluating process equivalence through activity equivalence, through the comparison of activities. We will use Process Equivalence and Process Mining techniques to perform the comparison between two activities at a time and detect duplicates. At the end, there will be an open-box question for the stakeholder to freely write, in order to allow him to add his personal annotations, in case some details were not taken into consideration in the Form, given the proved importance annotations have in process customization [35].

We will implement the set of Relations Discovery Questions through Google Forms [44], by asking each stakeholder to answer the form with the chosen questions and thus produce the Consolidated Model. A form is a simple interface for accessing and manipulating data, that will be filled with information provided by the stakeholders and afterwards will be associated with a taxonomy that will allow us to construct the Consolidated Model.

4.2.3.A Primary Questions

As previously stated, these questions are aimed to be posed for an individual activity at a time, considering its features and direct environment. The Primary Questions are composed by three different sets of questions, each of them focusing on a specific trait of the activities within a business process that will lead us to important conclusions, **Early Questions**, **Detail Questions** and **Responsibility Questions**, that we will now dive into.

The questions in this set have multiple choice answers, which compels the stakeholders to choose among the available options in most cases, instead of freely writing their opinion. The exception to this rule will be explained in the next subsection, Early Questions. Given that the options are limited and usually short, their comparison is pretty straightforward, but we will discuss this with more detail in the Section Integration Algorithm 4.3.

Early Questions

The Early Questions are based in the Six Interrogatives that are also deeply related with the Zachman Framework, both discussed in 3. These Six Interrogatives - Who, When, Where, What, Why and How - characterize the main features an activity possesses and define the basis of this building block.

These questions have a direct connection with the work of Colaço and Sousa [30], with a slight nuance. Their work focused on View Integration but with a more detailed and thorough approach to the classification of the elements represented in the models. They performed a view element classification based on the instance of the lane in which a given activity was inserted in. That is the reason behind the fact that from the lead-off interrogatives, only three - Who, Where and What - were transferred to this algorithm and labeled as dimensions. Hence, the view elements classification created an association between the possible instances of the lanes within the BPMN and the corresponding dimension.

Once these dimensions are identified and the elements that belong to each one of them are categorized through the application of a classification process of the view elements, they are utilized to create a taxonomy, which allows for a more complex and detailed organization of the different elements of each dimension.

This concept of taxonomy integrated with the dimensions was further developed by Diogo Cardoso [1], who created an hierarchical structure, the taxonomy tree. That structure is associated with each dimension and allows the existence of various levels of detail. The remainder of his work focuses on the tweaking of the dimensions and respective levels of detail allowing for the creation of different views of a business process.

From both their works, we can conclude that the concept of hierarchical taxonomy can be obtained using the user input by performing a classification process of the view elements. Therefore, we also applied this concept to our work, creating a secondary structure in the form of an hierarchical taxonomy tree that works as an addition to the solution's main purpose.

As in the work of Colaço and Sousa [30], the Interrogatives were shortened to a smaller set. In the case of this work, they were abbreviated to the four main ones - Who, When, Where, What - that can be easily mapped out to an hierarchy of concepts extracted from the stakeholder-specific view model in BPMN. The Who, Where and What were directly transferred from their work, however we decided to add the When. This extension is due to the fact that the main goal of this set of Relations Discovery Questions was to include all the Six Interrogatives, nevertheless the Why and How allowed for extensive and detailed answers that made the task of comparing the answers to this set a bit more complicated and prone to misclassification. However, the When options are a limited set that the stakeholder who is answering only has to choose from, so we decided to include this dimension in the Early Questions set.

The set of **Early Questions** is the following:

1. Who performed the activity?

2. Where did the activity take place?

3. When does this activity occur?

4. In what does this activity consist of?

For the set of Relations Discovery Questions presented above, the corresponding options of answers are mapped with the elements of the business process models, hence we assume that an algorithm based on the view element classification by Colaço and Sousa [30] placed the BPMN elements in the correct options for the Early Answers. As a result of this mapping, the options for question 1 are the lanes identified as *Who*, the options for question 2 are the lanes identified as *Where* and the options for question 4 are the lanes identified as *What*, all in the view element classification algorithm. Finally, the options for question 3 are the elements identified with *Timer Events* by the sweeper algorithm.

Given the answer options that will provide us with information on the four aforementioned dimensions, the stakeholder may choose one of them or add a new option, choosing "Other" instead of one of the available answers. When choosing this answer, the stakeholder can freely write what he believes to be the proper answer, but it is mandatory to place it in the existent hierarchy for that dimension, stating whether it *contains* or *is contained* a certain element within that dimension. Later on, the person responsible for this integration process, let us name him the *process guardian*, will approve and validate this new answer. For the purpose of this work, we assume that the answer is valid if the one providing it is able to fit it into the existing hierarchy. This will allow for a more detailed relation of resemblance.

Detail Questions

Having discussed the Early Questions that endorse the main properties of an activity, let us discuss the second set of Relations Discovery Questions, the Detail Questions, that endorse second level features and result of the execution of the activity. The Detail Questions are based on the immediate environment surrounding the activity. All the information about what the activity produces, triggers and updates is retrieved with these questions, together with information about the execution of the activity itself.

As explained in the Assumptions section 4.2.1, this solution is built to work with models designed in BPMN. Therefore, we used to our advantage some BPMN specific-concepts to retrieve information about the views through this set of Relations Discovery Questions. Concepts such as activity, gateway, event, artefact, data store and data object, that were previously described in the Background 2, were used both for the questions and the options of answer. However, in order for this set of questions to be understood by all the stakeholders regardless of their position and knowledge, instead of utilizing the concepts themselves, we used their definitions and their roles within a business process model to describe them.

For instance, Question 1 and 2 aim to discover the antecedent and consequent activities, overshadowing the details that are too tight with the modeling task, meaning that we aim to compare trigger and triggered activities only and not events or gateways. Question 3 aims to discover if the activity needs computing support or if it is manual. Questions 4 and 5 regard discovering the used Data Stores, whereas Questions 6 and 7 are related with unveiling the Data Objects. Question 8 regards discovering information about the Events, Question 9 about artefacts and finally Question 10 deals with the level of granularity of the activity.

The goal of the Detail Questions is to unveil characteristics of activities based on what changes with their execution, using more technical details of the activities. We aim to find resemblance in activities who have similar immediate environments, based on the premise that if two activities receive the same input and output the same result they are more likely to have a degree of resemblance.

The set of **Detail Questions** is the following:

1. What activity triggered this activity to be executed?
2. What activity does this activity trigger?
3. Which computing system/tool is used in the execution of this activity?
4. Does the activity require the use of information that is produced out of the scope of this process (or other executions of this process)?

5. Does this activity require the storing of information in a more permanent way?
6. Does the activity produce any documents, either in physical or digital form, that can only be assessed during the execution of the process?
7. Does the activity require any documents, either in physical or digital form, that can only be assessed during the execution of the process?
8. Does the activity depicts any kind of visual representation of data or notes?
9. Does something else occur during the course of the process (e.g. a message, a timeout)?
10. What is the level of granularity of this activity?

Once again, we assume the existence of a sweeper algorithm that fills the answer options with the BPMN-specific elements, since we assume all process views are represented in this modeling language. The goal of this set of questions is to take advantage of that, by characterizing activities concerning their near environment.

For questions 1 and 2, the answer options are the set of activities of that business process view. The answer options for question 3 are the available *Systems* or the option *manually*, for question 4 and 5 the available *Data Stores*, for questions 6 and 7 the available *Data Objects*, for question 8 the available *Artefacts* and for question 9 the available *Events*. Regarding question 10, the answer options are *low*, *medium* and *high*, the tree options a stakeholder can choose from when classifying the level of granularity of an activity. This question will be used specifically for the investigation of composition relations.

Responsibility Questions

Having discussed both the Early Questions and the Detail Questions, the Responsibility Questions are the remaining set of Primary Questions. The Responsibility Answers are targeted to the person responsible for a given activity, being the one who usually overviews the process specific to that view. Therefore, if the person answering the form is not responsible for the activity, these questions will not be posed.

In the case of the person answering these questions is in fact responsible for the activity, we aim to extract the last bit of information about it, taking advantage of the fact that the person providing the answers has a higher level of accountability about the execution of that activity than a regular stakeholder. We rest this on the argument that a stakeholder with a responsibility role is over an activity more likely to have context-awareness over it and answer more accurately.

The set of **Responsibility Questions** is the following:

1. Are you responsible or part of team responsible for the execution of this activity?
2. Do you agree with the name given to this activity?
3. Do you work with anyone outside your team for the execution of this activity?
4. Does this activity always unwind the same way?
5. Is this activity a mandatory activity to guarantee a good functioning of the process?
6. Could this activity be set aside while the process remained with good functioning?
7. Does this activity usually happen always at the same time?
8. Does it usually take the same amount of time to execute this activity?
9. Does this activity usually happen always at the same place?
10. Does this activity usually unfold in a standard manner?
11. Does this activity happen always with the same purpose?

This set of questions has as answer options a limited set of two: *yes* or *no*. The goal of this set is to take advantage of possible extra insight the person who is responsible for the activity may have.

4.2.3.B Context Questions

Having covered the details and goals of the Primary Questions, we will now discuss the second major set of questions, the Context Questions.

We consider lack of broad context and severe individuality in the **Primary Questions** a major drawback for achieving an effective Set of Relations Discovery Questions that would allow us to discover relevant relations between activities. Therefore, the remaining questions of the first building block of this solution, named **Context Questions Questions** 3.3, will take into account the context and the flow of the process. The rationale behind it is to compare a pair of activities at the same time, instead of just comparing their features separately. By doing this, we can ask questions regarding activities from two models at the same time to identify relations and hence reaching a Consolidated Model.

The Context Questions are only applicable if the activities being compared offer a possibility of resemblance after undergoing the Primary Questions. In other words, the stakeholders answer the set of Primary Questions about all the activities, and for the pairs of activities when a possibility of similarity is found, they answer the set of Context Questions regarding that pair, that can be identified as a resemblance candidate pair. Similarly, for the sets of activities when a possibility of composition is found based on the answers of the Primary Questions, the stakeholders answer the set of Context Questions regarding that set, that can be identified as a composition candidate set.

The set of **Context Questions** is the following:

1. Do both activities have similar/equivalent names?
2. Do both activities occur simultaneously?
3. Are both activities executed by the same person?
4. Do both activities occur at the same location?
5. Do both activities have a similar precedent activity?
6. Do both activities have a similar consequent activity?
7. Do both activities have the same purpose?
8. Do both activities endorse the same regulations?
9. Do both activities need the same set of requirements to execute?
10. Do both activities use the same IT support system to execute?
11. Is activity A a possible composition of activities B, C and D?

This set of questions has as answer options a limited set of two: *yes* or *no*, just like the Responsibility Questions do. The goal of this set is to corroborate if the resemblance candidate pairs are in fact similar or not, by asking questions 1-10, and if the composition candidate sets portray a situation of composition, by asking question 11.

4.2.4 Relations

The second building block of this solution is the relationships found between activities. There are two types of relationships we aim to encounter in this solution: *Resemblance* and *Composition*. The goal of finding these associations is to identify connection points between activities so that it is possible to join them into the Consolidated Model through them. As we have previously stated, we aim to explore process equivalence through activity equivalence, and finding these relations of *Resemblance* and *Composition* is how we execute such task.

Resemblance

The *Resemblance* relation between activities is identified by means of the following procedure: after gathering the answers for all of the Primary Questions, a comparison of answers is launched. This comparison task consists in comparing the *strings* that correspond to the answers provided by the stakeholders in the form, and it embodies four phases. If a pair of activities successfully passes through the first three, it becomes a Resemblance Candidate Pair and goes on to the fourth phase, where it will be confirmed if the pair is actually similar.

Firstly, the Early Answers are compared, and only if all four of them match the Detail Answers are compared. If all of the Detail Answers match, the Responsibility Answers are compared, and a pair of activities is set as Resemblance Candidate Pair if all the Responsibility Answers are *yes*, with a error margin of two answers being *no*. Finally, for each Resemblance Candidate Pair, the set of Context Questions is asked and the answers will determine if the pair is actually similar or not. Once again, we employ the error margin of among all the Novel Answers answered as *yes*, two answers being *no* to be considered similar.

As previously mentioned, there is a **Secondary Relation Structure**, whose goal is to hierarchically structure the level of resemblance between activities. There are two different instances of resemblance in our solution:

- **Full Resemblance**: The pair of activities matched exactly all the Early Answers, so they are as similar as this algorithm can encounter.
- **Partial Resemblance**: The pair of activities did not match exactly all the Early Answers.

For the case of **Full Resemblance**, the algorithm just outputs the pair of activities and the elements of the dimensions in which they overlap. However, for the case of **Partial Resemblance**, which occurs every time a stakeholder chooses the option *Other* for at least one of the questions, he has to place the new element within the Secondary Relation Structure, stating whether it *contains* or it *is contained* in a given existent element. Thereby, it is possible to unveil other levels of resemblance that are not just being completely similar in all dimensions. In both cases, it can be perceived at what level the pair of activities is similar, which provides more detailed insight about them and is further applied in the construction of the Consolidated Model.

Composition

The *Composition* relation between activities is identified by means of the following procedure: after gathering the answers for all of the Early Questions, a comparison of answers is launched, in which if half of the answers are similar for a given set of activities, that set is considered a Composition Set Candidate. Then, for the activities that belong to that set, the answer for question 10 - *What is the level of granularity of this activity?* - of the Detail Questions is analyzed. If within the Composition Set Candidate, at least one activity with the answer *low* or *medium* - which we will name mother-activity - and one or more than one with the answer *high* - which we will name children activities - can be tracked, the latter are saved as composing the former. Finally, as it happens in the *Resemblance* relation, the Composition Set Candidate is confirmed within the Context Questions set, particularly with question 11. Upon an affirmative answer for question 11, the Composition Set Candidate is confirmed as a Composition Set and that information leverages the construction of the Consolidated Model. For this type of relation, the Detail Questions, excluding question 10, and Responsibility Questions are ignored, since we believe that for discovering Composition Sets, the information retrieved from those two sets of questions is not relevant.

4.2.5 Consolidated Model

After gathering all the relevant information from the different stakeholders through the Set of Questions 4.2.3 and discovering the Relations 4.2.4 between activities, is it possible to construct a Consolidated Model. The Consolidated model is presented graphically, in a BPMN based manner, but with a few annotations stating the relations found.

Since the construction of the Consolidated Model is set upon the relations found, the activities are represented in the following way:

- Activities with **no relation** found - exactly as they are.
- Activities with ***Resemblance*** **relation**:
  - **Full Resemblance** - the one with the more complex name is shown as the main one in the Consolidated Model, and the remaining one is represented above.
  - **Partial Resemblance** - the one with the more complex name is shown as the main one in the Consolidated Model, and the remaining one is represented above.
- Activities with ***Composition*** **relation** - the activity of the set classified as the mother-activity, labeled with *medium* or *low* appears as the main one in the Consolidated Model, and the children activities - labeled with *high* granularity - appear above.

The stakeholder-specific views that worked as the starting point of this solution can also be enriched with the information discovered throughout this process. For each view, if a given activity belongs to a relation of Resemblance or Composition, the corresponding pair or set, accordingly, will also be presented in that view. This way, we will not only make available the Consolidated Model where evidence about the process as a whole can be perceived, but we will also enrich the stakeholder-specific views that were the input of this solution.

4.3 Integration Algorithm

After discussing the building blocks from which this solution arises, let us examine the algorithm behind it. The implementation of this algorithm was executed in *Python 3* and all of the code is publicly available on Github [1].

This algorithm is asynchronous, therefore it relies on receiving different inputs at different stages of its execution, according with its intermediate results, to achieve the final outcome. The starting point, as previously explained, are the stakeholder-specific views, which are transcribed to this algorithm in the form of a set of activities per view and their corresponding BPMN elements.

integrateSTKViews

The function *integrateSTKViews* is the main function of this solution, being the translation of the application of the building blocks that are the pillar of this solution. We will overview this function as a big picture of the algorithm, and then dive into each function to understand the detail behind them.

Given the set of views to be integrated, the algorithm will find view equivalence through the comparison of two views at a time. The first step of this algorithm is to generate and ask the Primary Questions, corresponding to *generatePrimaryQuestions* and *askPrimaryQuestions* accordingly. Then, the answers to those questions will be compared in *comparePrimaryAnswersResemblance* and in *comparePrimaryAnswersComposition* in order to discover relation between them and consequently potential Resemblance Pair and Composition Sets. For each pair or set of candidates, the Context Questions will be generated and asked, in *generateNovelQuestionsResemblance* and *askNovelQuestionsResemblance*, and in *generateNovelQuestionsComposition* and *askNovelQuestionsComposition* correspondingly. Provided with the answers to those questions, it is possible to construct the Consolidated Model in *produceConsolidatedModel*. Finally, the Secondary Relation Structure can also be generated in *gen-*

[1] https://github.com/joanafpribeiro/thesis

erateResemblanceStructure.

Algorithm 4.1: *integrateSTKViews*

```
Result: consolidatedModel
for each pair of views do
    firstViewQuestions, secondViewQuestions ←
     generatePrimaryQuestions(firstView, secondView);
    firstViewAnswers, secondViewAnswers ←
     askPrimaryQuestions(firstViewQuestions, secondViewQuestions, firstView, secondView);

    similarActivitiesPairs ←
     comparePrimaryAnswersResemblance(firstViewAnswers, secondViewAnswers);
    compositionActivitiesSets ←
     comparePrimaryAnswersComposition(firstViewAnswers, secondViewAnswers,
    firstView, secondView);
    novelQuestionsSimilar ← generateNovelQuestionsResemblance(similarActivitiesPairs);
    novelQuestionsComposition ←
     generateNovelQuestionsResemblance(similarActivitiesPairs);
    novelAnswersSimilar ← askNovelQuestionsResemblance(novelquestionssimilar);
    novelAnswersComposition ←
     askNovelQuestionsComposition(novelquestionscomposition);
    consolidatedModel ←
     produceConsolidatedModel(novelanswerssimilar, novelanswerscomposition);
    resemblanceStructure ←
     generateResemblanceStructure(consolidatedModel, firstView, secondView);
    if Consolidated Model is not Null then
        outputSimilarPairs();
        outputCompositionSets();
    else
        Fail
    end
end
```

generatePrimaryQuestions

This function is the place where the Primary Questions will be generated. A list with the predefined questions 4.2.3.A is initially loaded into the Process Repository, and in the algorithm, in this function, they will be produced according to the activities from a given view, which we access through the function *getActivitiesfromBPMN*. The view element classification algorithm together with the sweeper algorithm provide the options for each Primary Question, finding the correspondence between the elements/dimensions and the questions. Then, the answer options are appended to the corresponding question. At the end, the result is a set of Primary Questions particularly created for this view, as well as the options for each one of them.

askPrimaryQuestions

This function inquires the stakeholders with the Primary Questions - Early, Detail and Responsibility - and retrieves their answers. The result of this function is the set of Primary Answers.

comparePrimaryAnswersResemblance

Given the set of Primary Answers of the two views being compared, this function compares both sets and outputs Resemblance Candidate Pairs. The logic behind this function translates what has already been detailed in 4.2.4 in the paragraph discussing Resemblance.

The first step is to divide the Primary Answers into Early Answers, Detail Answers and Responsibility Answers. Then, the Early Answers for both views are compared and the pairs in which all four of them match are saved in a list of *similarPairs*. This comparison unfolds by comparing the strings that mimic the options chosen by the stakeholders. Afterwards, a similar process occurs for the Detail Answers, but before saving the pairs in which all the answers match, it is checked if they correspond to the ones previously saved after comparing the Early Answers. Only the ones who do are saved in *similarPairs*, which means that this list holds the pairs of activities that have similar Early and Detail Answers. Finally, the same comparison process happens for the Responsibility Answers, once again with the error margin of two negative answers for a given pair. Once again, the pairs that surpass this threshold are checked to verify if they are present in *similarPairs*. Only the ones who are get to be part of the final set of Resemblance Pairs, saved in *similarPairs*. The result of this function is the set of Resemblance Candidate Pairs, containing the pairs of activities who have similar Early, Detail and Responsibility Answers, for the two views being compared.

comparePrimaryAnswersComposition

Given the set of Primary Answers of the two views being compared, this function compares both sets and outputs Composition Candidate Sets. The logic behind this function translates what has already been detailed in 4.2.4 in the paragraph discussing Composition. Firstly, this function compares the Early Answers provided. The criteria regarding the Early Answers in order for a set of activities being considered a Composition Candidate Set is that half of them match. For the sets who fulfill that requirement, a filter is applied before moving onto the next stage. Since for this relation the result of the comparison is not saved as a pair of activities belonging to the two views being compared, this filter assures that in each set not all the activities pertain to the same view. The following step is to analyse the answer to the question regarding the granularity of the activity, which is part of the Detail Questions. For each set, it is evaluated if there are different levels of granularity assigned to their activities: there should be at least one activity assigned with *low* or *medium* and at least one activity assigned with *high*. The sets in which that condition is verified are considered the Composition Candidate Sets. Their activities are represented in the following manner: [low/medium - high], meaning that the activities classified with *low* or *medium* - the mother-activities - are a possible composition of the activities classified with *high* - the children activities.

generateNovelQuestionsResemblance

This function is executed only in the scope of the Resemblance relation. It receives the Resemblance Pairs resulting from the *comparePrimaryAnswersResemblance* function, and generates the Context Questions accordingly. Similarly to the *generatePrimaryQuestions* process, the standard questions were previously loaded into the Process Repository and in this moment they are generated for the specific pairs of activities found. The answer options for this set of questions are simply *yes* or *no*, so the final step consists in appending them to the questions. The result of this function is a set of Context Questions particularly created for this view, as well as the options for each one of them.

generateNovelQuestionsComposition

This function is executed only in the scope of the Composition relation. It receives the Composition Sets resulting from the *comparePrimaryAnswersComposition* function, and generates the Context Questions accordingly. Similarly to the *generatePrimaryQuestions* and the *generateNovelQuestionsResemblance* process, the standard questions were previously loaded into the Process Repository and in this moment they are generated for the specific sets of activities found. For the particular case of the Composition Relation, there is only one Context Question just to confirm the Composition Sets. The answer options for this question are simply *yes* or *no*, so the final step consists in appending them to the question. The result of this function is a set of Context Questions particularly created for this view, with only one question, as well as its answer options.

askNovelQuestionsResemblance

This function inquires the stakeholders with the Context Questions regarding the Resemblance relation, in order to confirm if the Resemblance Candidate Pairs are actually similar or not. This function receives the set of questions from the *generateNovelQuestionsResemblance*, inquires the stakeholder and retrieves their answers. The result of this function is the set of Novel Resemblance Answers regarding the Resemblance relation.

askNovelQuestionsComposition

This function inquires the stakeholders with the Context Questions regarding the Composition relation, in order to confirm if the Composition Candidate Sets are actually a composition of activities or not. This function receives the set of questions from the *generateNovelQuestionsComposition*, inquires the stakeholder and retrieves their answers. The result of this function is the set of Novel Composition Answers regarding the Composition relation.

produceConsolidatedModel

This function receives the answers from the complete set of Context Questions, from both the *askNovelQuestionsResemblance* and the *askNovelQuestionsComposition*. The goal of this function is to confirm if the Resemblance Candidate Pairs and Composition Candidate Sets are actually similar and a composition, respectively. It starts by analysing the received answers regarding the Resemblance relation, in which if the answers are all affirmative for a given pair, with a tolerable error of two negative answers, the pair is considered similar. A similar process occurs for the Composition relation, but in this situation there is only one answer to confirm. If that answer is affirmative for a given set, the set is considered a composition. With the final Resemblance Pairs and Composition Sets found, it is possible to generate the Consolidated Model. This Model is a conjunction of activities from different stakeholder-specific views, with the addition of making available the relations between them that have been previously found. A graphic representation of the Consolidated Model is presented in the Demonstration 5 for better understanding. If no relation is found, there is no production of such a model.

generateResemblanceStructure

This function produces the Secondary Relation Structure that allows the stakeholders to perceive the dimensions in which a pair of similar activities match. It gathers all the information retrieved about this type of relation and outputs it, which allows for the stakeholders to not only understand which activities are equivalent, but also what dimensions make assemble their equivalence. As previously explained, there are two cases of resemblance, either Full or Partial. In the case of Full Resemblance, this function joins the activities with the respective corresponding dimension and the name of the specific element of that dimension. Therefore, it will become visible for everyone exactly which element a given pair of similar activities hold in common. The graphical representation for this case will be a straight line with all the aforementioned information, given that Full Resemblance pairs match in all dimensions. In the case of Partial Resemblance, which happens when the stakeholder chooses the option *Other* instead of one of the available ones, is it important to recall that it is demanded that the stakeholder places hierarchically the element they believe to be the right answer. Therefore, that placement will be exhibited in this structure. Hence, for this instance, the structure will be a tree-like hierarchical one. In one hand, if the placement option chosen was *contains*, the element provided by the stakeholder will be above the one he described to be contained in. On the other hand, if the placement option chosen was *is contained in*, the element provided by the stakeholder will be below the one he described to contain the former. The case of the Composition relation is visually similar to this one. On a similar note as for the Consolidated Model, a graphic representation of this structure is presented in the Demonstration 5 for better understanding.

5

Demonstration

Contents

This chapter - **Demonstration** - corresponds to the fourth activity of the DSRM with the exact same name: Demonstration. We will demonstrate the usage of our solution by applying it to one fictitious scenario and a real-world scenario. The goal of both demonstrations is to ease the understanding of this research and improve the reader's comprehension on how it can add value to an organization.

For both scenarios, the demonstration unfolds as follows:

1. Identify the Business Process and the respective stakeholder-specific views.
2. Apply the Set of Questions in a Form format to the stakeholders.
3. Apply the View Integration Algorithm to discover Relations between activities.
4. Build the Consolidated Model and enrich the stakeholder-specific views.

5.1 Illustrative Scenario 1 - Car Repairing Process

For the Illustrative Scenario 1, we will be recalling the Car Repairing Process example that has been previously described in Section 1. This example has been analyzed in the literature and published in the proceedings of the ninth EEWC [1]. We will make use of this process by applying our View Integration Algorithm to the available views and evaluate if it is possible to reach the relations presented in the Consolidated Model.

For the Car Repairing Process, the two available views are the HR View and the Facilities Views, designed by the HR Department and the Facilities Department, respectively. Given that they portray distinct departments, they also have diverging perspectives. This problem is aligned with what we aim to solve with the proposed solution, which stems from the existence of different stakeholder-specific views and the need to integrate them into a Consolidated Model.

Figure 1.1 displays the HR View, which is focused on the employees and organisational units that execute the activities. Figure 1.1 also shows the Facilities Views that targets the location where the process activities are performed. Consequently, the two views have a very distinct point of view depending on the characteristics of the department.

Following the methodology of our solution, the first step is to load both views onto the Process Repository, so that the View Integration Algorithm can be launched. Then, the Set of Questions is generated for the activities that belong to both the HR and the Facilities Views. The surveys were produced in Google Forms and will be made available for the Illustrative Scenario 2, since it is more complex than the one produced for the current scenario. Given the fact that this scenario is fictitious, and therefore there are no real stakeholders to whom ask the questions, for the purpose of this demonstration, the answers to the survey will delineate a happy path of responses, meaning the answers will impersonate a default scenario featuring no exceptional or error conditions.

After gathering the answers from the stakeholders and loading them into the Process Repository, the algorithm proceeds to find Relations of Resemblance and Composition between activities. Table 4.2.4 depicts the relations found between activities of the HR and Facilities Views.

Table 5.1: Relations found between activities from Facilities and HR Views

Facilities View	HR View	Relation
Assess Car Damage	Assess Car Damage	Resemblance
Replace parts	Fix parts	Resemblance
Prepare Car for Spraying	Prepare Car for Spraying	Resemblance
Spray and Inspect Painting	Spray Car + Inspect Painting Quality	Composition

A relation of Resemblance was found between *Assess Car Damage* from the Facilities View and *Assess Car Damage* from the HR View. This occurred due to all similar answers regarding the Early Questions for both activities, since the difference between them was the dimension of the lane in which they were inserted, which is being accounted for in the algorithm. Regarding the Detail Questions, besides the similar consequent and antecedent activities, there was little to no information about other BPMN elements which led to almost none questions answered. The Responsibility Questions were all answered affirmatively, since this is a theoretical example regarding an happy path. The Novel Questions were also all answered affirmatively due to the same reasons, confirming the Resemblance Relation.

In a similar manner, a relation of Resemblance was also found between *Replace parts* from the Facilities View and *Fix parts* from the HR View, and between *Prepare Car for Spraying* from the Facilities View and *Prepare Car for Spraying* from the HR View.

A relation of Composition was found between *Spray and Inspect Painting* from the Facilities View and *Spray Car* and *Inspect Painting Quality* from the HR View. This occurred due to the fact that they had at least half of coinciding Early Answers and *Spray and Inspect Painting* was classified with *low* granularity whereas *Spray Car* and *Inspect Painting Quality* were classified with *high*, which allowed for a possibility of composition. Once again, the Novel Answers were all affirmative, confirming the Composition Relation.

Analyzing the relations found, one could produce the Consolidated Model present in Figure 1.2, hence enrich the stakeholder-specific views by adding to each activity the detail associated with the relations found. Consequently, we can conclude that our algorithm was able to find the relations needed to produce such model, therefore being successful for this scenario.

For this case, all the Resemblance Relations found were Full Resemblance Relations, therefore the Secondary Relation Structure for all of them is simply a straight line and does not exhibit an hierarchy, as depicted in Figure 5.1, 5.2 and 5.3. On the other hand, the Composition Relation does, showing the mother-activity above the children-activities, as depicted in Figure 5.4. Since the Resemblance is Full for all cases, there was no need to show the matching dimensions, because it would be all of them.

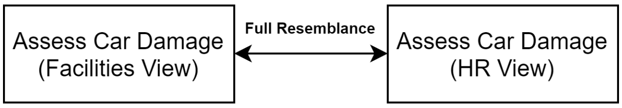

Figure 5.1: Full Resemblance 1

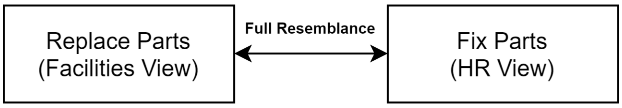

Figure 5.2: Full Resemblance 2

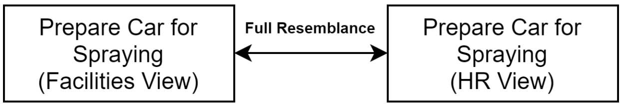

Figure 5.3: Full Resemblance 3

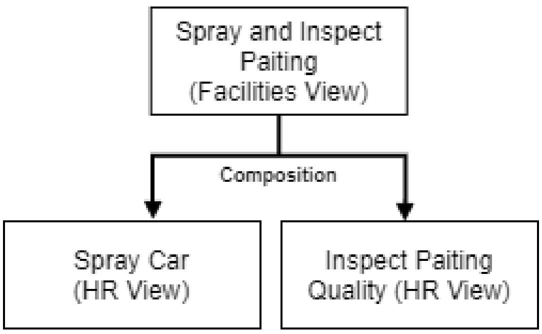

Figure 5.4: Composition 1

5.2 Illustrative Scenario 2 - Analysis, Decision and Granting of Credit

For the Illustrative Scenario 2, we will be dealing with a real-world scenario. For this case, we will be analyzing the business process of Analysis, Decision and Granting of Credit of Caixa Geral de Depósitos [45]. Esperto and Sousa [46] produced stakeholder-specific views for this process, regarding the IT View, the Audit View and the Risk View. We will be using the models produced by them as a black-box, just as it would happen if this solution was being tested within an organization with their existing models. We are aware that the models may contain a few errors and inconsistencies, but we will be working with them AS-IS to stay truthful to our purpose with this demonstration.

Similarly to Illustrative Scenario 1, given that these views portray distinct departments, they also have differentiating perspectives, which is once again aligned with what we aim to solve with the proposed solution. Following the same reasoning applied in the first scenario, from this starting point we applied our solution with the goal of integrating those three views into a possible Consolidated Model.

Figure 5.5 displays the IT View, which is responsible for the architecture, hardware, software and networking of the technology within the company. Figure 5.6 shows the Risk View, which concerns the identification, evaluation, and prioritization of risks within the company. Figure 5.7 depicts the Audit View, which regards the identification, evaluation, and prioritization of risks within the company. These three departments are considered different stakeholder groups that have completely distinct concerns, but all of them play an important role in the process of Analysis, Decision and Granting of Credit. Consequently, they portray distinct points of view depending on the characteristics of each department.

After having identified the stakeholder-specific views, a survey with the specific questions regarding the activities that belong to those views is generated. This survey was produced in Google Forms, and regards the IT View [1], the Audit View [2] and the Risk View [3]. Afterwards, the survey regarding the Novel Question [4] was posed. The stakeholders proceeded to answer all the surveys.

Subsequently, all the answers were gathered and the View Integration Algorithm was applied. In accordance with the algorithm, the following Relations of Resemblance and Composition between activities presented in Table 5.2, Table 5.3 and in Table 5.4 were found.

Examining the relations found, one could produce the Consolidated Model depicted in Figure 5.15. The Secondary Relation Structure for each relation is presented in Figures 5.8, 5.9, 5.10, 5.11, 5.12, 5.13 and 5.14.

As you can corroborate by the Consolidated Model produced, it is possible to find some relations between the activities of the different views represented as a form of annotations. Therefore, it is possible

[1] https://forms.gle/aYqqrzJz8tiVMbpH6
[2] https://forms.gle/58CoTmhDvKog9xAZ9
[3] https://forms.gle/x99oefK96rS4Lnx68
[4] https://forms.gle/8VrvwUm847Jv4r1U7

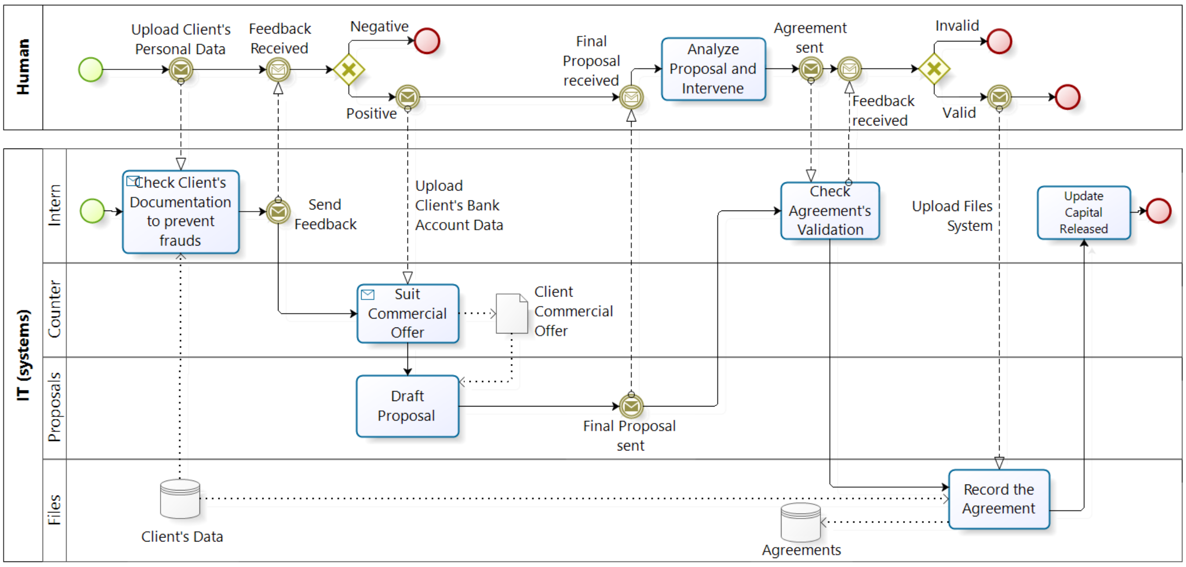

Figure 5.5: IT View

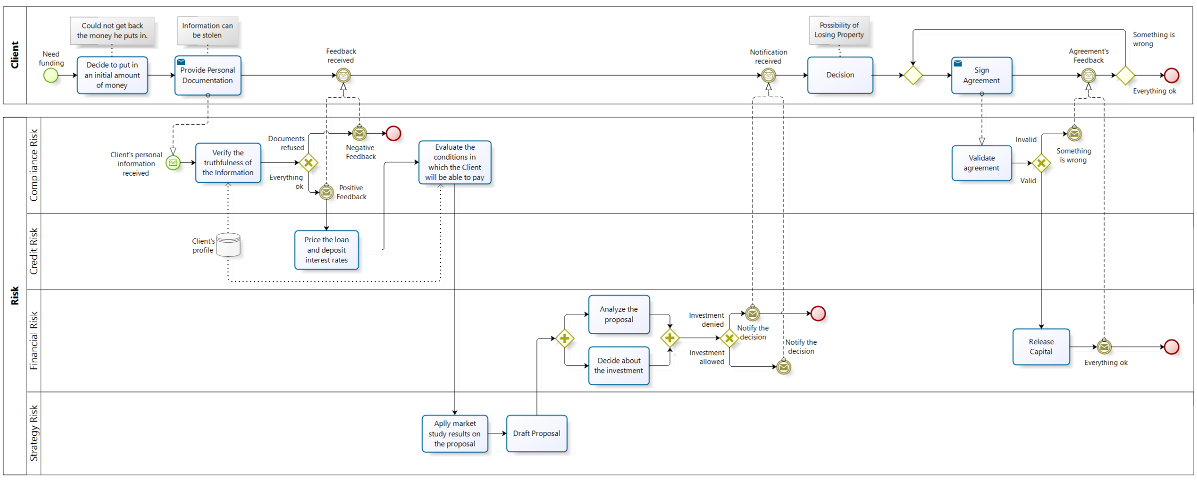

Figure 5.6: Risk View

Table 5.2: Relations found between activities from IT and Risk

IT View	Risk View	Relation
Draft Proposal	Draft Proposal	Resemblance
Analyze the Proposal and Intervene	Analyze the Proposal	Resemblance
Check Agreement's Validation	Validate Agreement	Resemblance
Update Capital Release	Release Capital	Composition

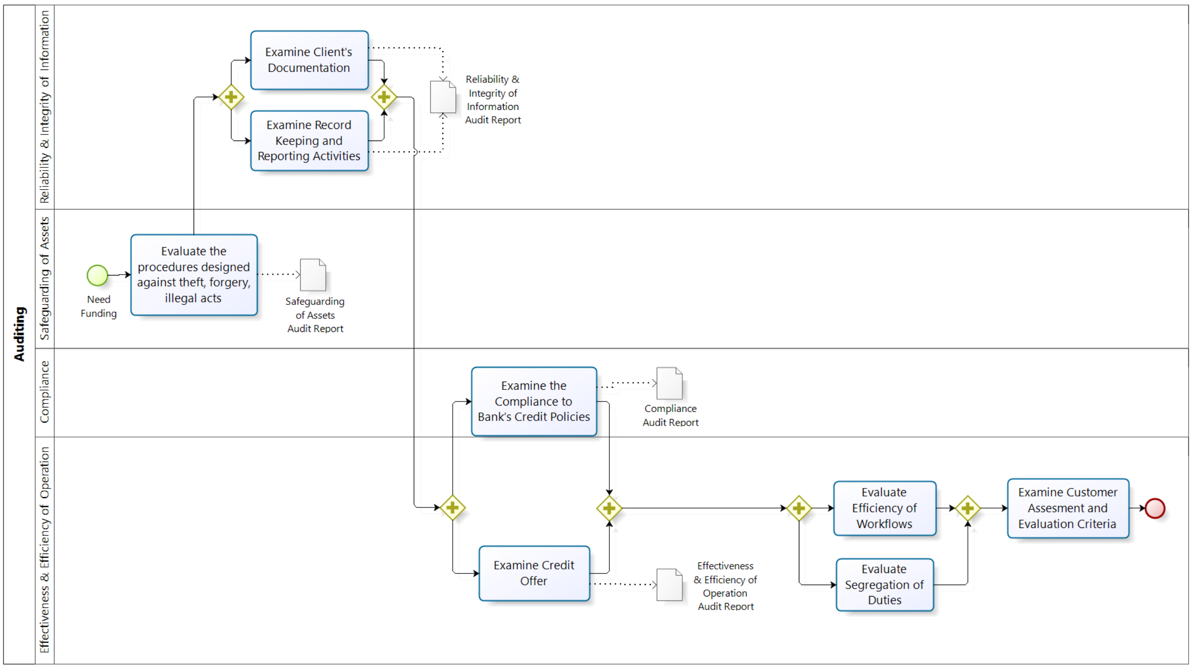

Figure 5.7: Audit View

Table 5.3: Relations found between activities from IT and Audit

IT View	Audit View	Relation
Check Client's Documentation to Prevent Frauds	Examine Client's Documentation	Composition

Table 5.4: Relations found between activities from Risk and Audit

Risk View	Audit View	Relation
Verify Truthfulness of Information	Examine Client's Documentation	Composition
Evaluate the Conditions in which the client will be able to pay	Examine the record keeping and reporting act	Resemblance

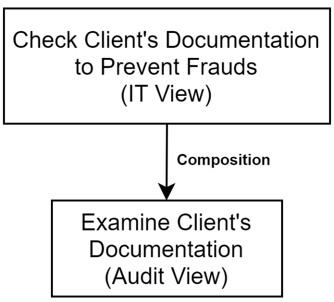

Figure 5.8: Composition 1

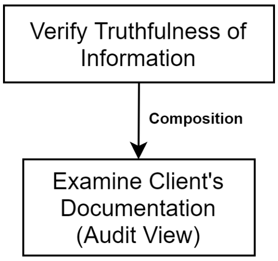

Figure 5.9: Composition 2

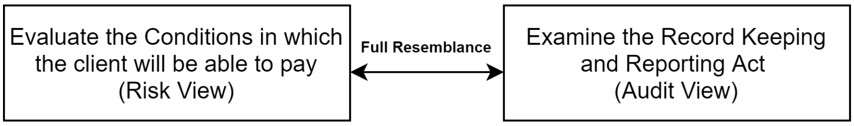

Figure 5.10: Full Resemblance 1

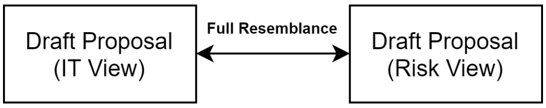

Figure 5.11: Full Resemblance 2

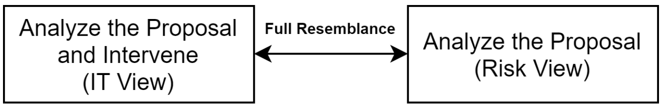

Figure 5.12: Full Resemblance 3

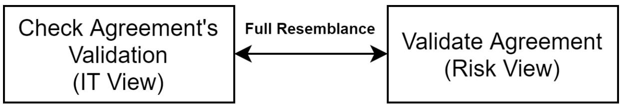

Figure 5.13: Full Resemblance 4

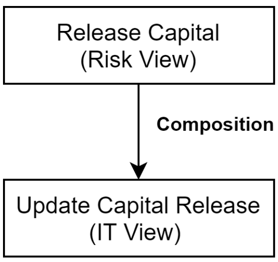

Figure 5.14: Composition 3

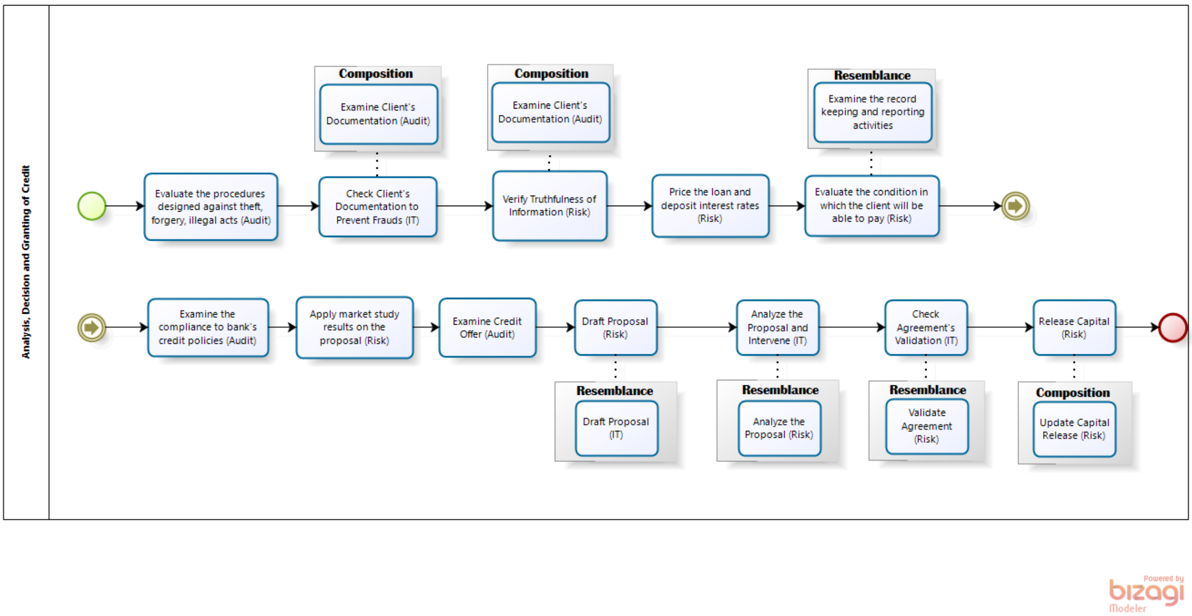

Figure 5.15: Consolidated Model for Analysis, Decision and Granting of Credit Process

to integrate them at a certain level into a Consolidated Model, hence allowing the organization to have a business process model for the Analysis, Decision and Granting of Credit that collapses knowledge about the IT, Audit and Risk department. Consequently, we can conclude that our solution is successful for this scenario.

6

Evaluation

Contents

This chapter - **Evaluation** - corresponds to the fifth activity of the DSRM with the same name: Evaluation. This chapter aims to explore how successful the proposed solution artefact presented in Chapter 4 4 actually is in solving the problem posed, and discuss the Demonstration 5 results comparing them to the research goals.

6.1 Evaluation Methods

To attain the established objectivesof this research work, we developed the View Integration Algorithm and used two different scenarios to demonstrate its results, which will now be assessed.

The DSRM presents some interesting guidelines to evaluate our solution. The artifact evaluation approach proposed by Prat et al. [47] will be used together with the four research principles introduced by Osterle et al [48]. Prat et al. [47] proposed an hierarchy of evaluation criteria for IS artifacts and a model of generic evaluation methods composed of six dimensions: criterion, evaluation technique, form of evaluation, secondary participants, level of evaluation, and relativeness of evaluation. The system dimensions defined by the authors are goal, environment, structure, activity and evolution. For each of them, there is evaluation criteria and sub-criteria. To assess our solution, the following criteria was chosen: Goal - **Efficacy** and **Technical Feasibility**, Environment - Consistency with People - **Usefulness**, **Understandability**, **Ease of Use** and **Operation Feasibility**.

The solution artifact will also be assessed against the four basic principles that DSR in IS must comply introduced by Osterle et al [48]: **Abstraction** - each artifact must be applicable to a class of problems, **Originality** - Each artifact must substantially contribute to the advancement of the body of knowledge, **Justification** - Each artifact must be justified in a comprehensible manner and must allow for its validation and **Benefit** - Each artifact must yield benefit, either immediately or in the future, for the respective stakeholder groups.

Considering the criteria described above, the evaluation will be two-fold, firstly by analysing the Demonstration results and then by discussing the stakeholders' opinions on the usefulness of the solution.

The solution artefact will be considered successful if we are able to reach a unique view, the so called Consolidated Model, that allows to keep a known relation between the different views. We will consider this work a success if we can identify correspondence between at least a pair of activities for every pair of views. This will allow for the organization stakeholders to have a consolidation path to generate views in the future, upon necessity or thorough investigation on that business process. If the solution is successful, we consider it effective thus achieving the Goal criteria aforementioned. Otherwise, we will try to reason out and understand where the shortcomings of our solution lay.

If we have a successful solution, another test that can be performed to verify the accuracy of the solution is switching the order of the views upon the application of the solution. This test will be successful if that happens, meaning the Consolidated Model should be independent of the order of the view entrance.

We also gathered feedback with the stakeholders to assess the Utility, Understandability, Ease of Use and Operation Feasibility of our solution.

6.2 Demonstration

This evaluation method is based on the **Demonstration** Chapter 5, where the proposed solution was put to practise with two Illustrative Scenarios, the first being a fictitious one and the second a real-world one.

The results of the Demonstration were described in detail in Chapter 5 5, however let us recall them. Either in Illustrative Scenario 1 - **Car Repairing Process** - or in Illustrative Scenario 2 - **Analysis, Decision and Granting of Credit** - the View Integration Algorithm was able to find relations of both Resemblance and Composition between activities of different views. The results show that with such relations, it is possible to withdraw a consolidation path and produce a Consolidated Model. Hence, it was possible to integrate views through their activities with distinct levels of abstraction and concerns.

Given the results achieved, we can argue about the **Goal** criteria, regarding the efficacy and technical feasibility of the solution artefact:

- **Efficacy** - defined as ”the degree to which the artefact achieves its goal considered narrowly, without addressing situational concerns” which translates to our context to the degree in which it is possible to integrate business process views into a Consolidated Model. This is verified in both scenarios of the Demonstration, where the solution provides a View Integration Algorithm capable of achieving the integration of distinct business process views, hence we consider the solution to be effective.

- **Technical Feasibility** - defined as ”evaluates, from a technical point of view, the ease with which a proposed artefact will be built and operated”, which measures the difficulty in developing a tool that supports the proposed solution. We believe that the AS-IS solution, with the surveys in Google Forms and the algorithm implemented in Python could be transferred to a plug and play tool with moderate ease.

Regarding the functioning of the algorithm, we also performed a couple of tests. For both Illustrative Scenarios, we exchanged the order in which the stakeholder-specific views were processed, to check if the relations between activities remained the same and were all found in every case. This test showed to be successful since regardless of the order of the view processing, the algorithm always found the same relations, therefore generating the same Consolidated Model. We can conclude that our solution is independent of the order of the view entering and processing.

6.3 Stakeholder Feedback

In order to argue about the **Environment - Consistency with People** criteria, regarding the usefulness, understandability, ease of use and operation feasibility of the solution artefact, we inquired the stakeholders to gather their opinions on the subject through a survey [1]. This stakeholder feedback was gathered in the scope of Illustrative Scneaio 2 - Analysis, Decision and Granting of Credit of CGD. We asked the stakeholders to classify statements according with their level of agreement in a scale of 1 to 5, 1 being ”Don't Agree” and 5 being ”Strongly Agree”. We came to following conclusions:

- **Usefulness** - defined as ”the degree to which the artefact positively impacts the task performance of individuals”, which defines how the solution positively impacts the work of the stakeholders. The solution has been proven as useful since the premise of this work was to build the Consolidated Model for the exact purpose of aiding in keeping the business views consistent. According to the survey, the stakeholders also considered the solution useful with a score of 4 in a scale from 1 to 5.
- **Understandibility** - defined as ”the quality of comprehensible language or thought”, which according to the survey, the stakeholders considered the solution understandable with a score of 4 in a scale from 1 to 5.
- **Ease of Use** - According to the survey, the stakeholders considered the solution easy to use with a score of 4 in a scale from 1 to 5.
- **Operation Feasibility** - defined as ”the degree to which management, employees, and other stakeholders, will support the proposed artefact, operate it, and integrate it into their daily practice”, which according to the survey, the stakeholders considered the solution operational feasible with a score of 4 in a scale from 1 to 5.

According to the O¨sterle et al. [48], DSR in IS must comply with four basic principles in which our solution will be assessed against:

- **Abstraction** - Each artifact must be applicable to a class of problems - The proposed solution is applicable to the class of view integration problems, where the different business process models are specified in BPMN, taking into account the assumptions made for the well functioning of the algorithm.

[1] https://forms.gle/xwDHXnaTGHCb6NK46

- **Originality** - Each artifact must substantially contribute to the advancement of the body of knowledge - The proposed solution tackles new ways of inquiring the stakeholders in order to assist the view integration process for in the area of business process design, which to our understanding is innovative and lacks contributions.
- **Justification** - Each artifact must be justified in a comprehensible manner and must allow for its validation - Each building block of the proposed solution is based on the literature review and emerged from the attempt on solving the problem described, later validated through a demonstration.
- **Benefit** - Each artifact must yield benefit, either immediately or in the future, for the respective stakeholder groups - The proposed solution provides a method to achieve view integration and supply the organization with a Consolidated Model on a given business process. This will allow the organization to maintain different views at the same time in a centralized manner, which has many benefits such as decreasing the chance of inconsistencies and spreading a common understanding on the business process.

7

Conclusion

Contents

This chapter - **Conclusion** - presents a final review of our work, stating its contributions, limitations and what we believe should be part of the next steps regarding the future work about this topic.

7.1 Contributions

The problem this work tried to solve is the integration of different business process models that represent the views from the different stakeholders into a centralized model. It served the purpose of discussing the difficulties that organizations face in coping with different stakeholder-specific views for the same business process, and how keeping them coherent and consistent is very troublesome. Consequently, let us recall the objectives defined for this research work:

> *Goals*: This thesis aims to integrate the different stakeholder views into a centralized Consolidated Model in order to find relations of resemblance or composition between the activities of the different views that allow for their integration and further enrichment.

Our contribution to this topic stands on the following topics:

- The addition of new and different types of questions for retrieving relations between activities.
- The production of the View Integration Algorithm.
- The integration problem being centered on stakeholder input.

The main contribution of this research work was the View Integration Algorithm, since it was based on the gathering of information directly from the stakeholders in an innovative manner. The adding of new questions, regarding the context and the environment of the business process, together with the inquiring to a pair of activities at the same time is a step further from the approaches identified in the literature and presents a novelty feature.

Our contribution to this problem is grounded on constructing a Consolidated Model that gathers the knowledge of the stakeholders and is able to find similarities through process equivalence discovery, through Form inquiring.

The proposed solution artefact also aims to contribute to making the stakeholder's tasks easier, which is why the solution is highly stakeholder dependent, since the information provided by them is the core of the View Integration Algorithm.

In order to demonstrate the functionality of the prototype of the View Integration Algorithm in practise, it was tested to support two distinct illustrative scenarios. According to the results, we can state that the goals for this research work were accomplished, since in both scenarios it was possible to integrate the available views by finding relations between activities and build a Consolidated Model. Therefore, we considered this solution to be successful.

7.2 Limitations

The main limitations identified in the proposed solution are the following:

- It is assumed that the solution is asynchronous, which requires an input providing as the solution runs.
- It is assumed the good functioning of a few functionalities needed for the execution of the solution.

These drawbacks make the solution dependent on a process manager to oversee its execution and provide input as it runs, but all of them can be fixed with not much trouble.

7.3 Future Work

After analysing the research proposal and encountering the aforementioned limitations, we believe that the future work regarding contributions for this topic should assess the following aspects. For instance, if the comparison algorithm to discover activity equivalence should be more fine-grained, if the set of questions posed should be enlarged or shortened, and evaluate if there are other types of relations that could be helpful to identify in order to solve the problem.

For the particular case of the proposed solution artefact, a future work contribution should try to turn it synchronous. This way, it could work as a plug and play feature in an organizational system, where the stakeholder-specific views are loaded and the survey with the set of questions is filled directly in the system. Thereby, the solution would become centralized and the Consolidated Model would be saved and easy to access. Also, connecting the solution to a BPMN-generating tool comes hand in hand with the first suggestion, since it would make it more automatic.

Bibliography

[1] D. Cardoso and P. Sousa, “Generation of stakeholder-specific bpmn models,” 2019.

[2] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.

[3] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*, 2013.

[4] G. Aagesen and J. Krogstie, “Bpmn 2.0 for modeling business processes,” in *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*, 2015, pp. 219–250.

[5] P. Sousa, C. Pereira, R. Vendeirinho, A. Caetano, and J. Tribolet, “Applying the Zachman Framework Dimensions to Support Business Process Modeling,” in *Digital Enterprise Technology*, 2007, pp. 359–366.

[6] J. A. Zachman, “Framework for information systems architecture,” *IBM Systems Journal*, vol. 38, no. 2, p. A Framework for Information Systems Infrastrucure, 1999.

[7] R. K. L. Ko, “A computer scientist’s introductory guide to business process management (BPM),” *Crossroads*, vol. 15, no. 4, pp. 11–18, 2009.

[8] A. Caetano, C. Pereira, and P. Sousa, “Generation of Business Process Model Views,” *Procedia Technology*, pp. 378–387, 2012.

[9] C. M. Pereira and P. Sousa, “Business process modelling through equivalence of activity properties,” in *ICEIS 2008 - Proceedings of the 10th International Conference on Enterprise Information Systems*, 2008, pp. 137–146.

[10] C. M. Pereira, A. Caetano, and P. Sousa, “Ontology-driven business process design,” in *IFIP Advances in Information and Communication Technology*, 2011, pp. 153–162.

[11] A. Caetano, A. R. Silva, and J. Tribolet, “Business Process Decomposition – An Approach Based on the Principle of Separation of Concerns,” *Enterprise Modelling and Information Systems Architectures - An International Journal*, vol. 5, no. 1, pp. 44–57, 2010.

[12] A. Anyieni, “Organisational Change : A Critical Review of the Literature,” *The International Journal of Professional Management*, vol. 11, no. 2, pp. 1–6, 2016.

[13] J. W. Moran and B. K. Brightman, “Leading organizational change,” *Journal of Workplace Learning*, vol. 12, no. 2, pp. 66–74, 2000.

[14] M. La Rosa, W. M. Van Der Aalst, M. Dumas, and F. P. Milani, “Business process variability modeling: A survey,” 2017.

[15] W. M. P. van der Aalst, “Business Process Management: A Comprehensive Survey,” *ISRN Software Engineering*, 2013.

[16] K. E. Weil, “PORTER, Competitive advantage, creating and sustaining superior performance,” *Revista de Administração de Empresas*, 1985.

[17] T. H. Davenport and J. E. Short, “New industrial engineering: Information technology and business process redesign,” *IEEE Engineering Management Review*, 1998.

[18] W. M. Van Der Aalst, A. H. Ter Hofstede, and M. Weske, “Business process management: A survey,” 2003.

[19] K. C. Laudon and J. P. Laudon, *Management Information Systems Managing the Digital Firm*, 1968.

[20] R. Dijkman, I. Vanderfeesten, and H. a. Reijers, “The Road to a Business Process Architecture: An Overview of Approaches and their Use,” *Architecture*, vol. 350, no. July, pp. 1–17, 2011.

[21] J. Mendling, “Foundations of Business Process Modeling,” *Handbook of Research on Modern Systems Analysis and Design Technologies and Applications, Mahbubur Rahman Syed and Sharifun Nessa Syed, Eds. Hershey, PA: IGI Global, 2008, ch. 15, pp. 189–222.*, 1998.

[22] J. Schekkerman, *How to Survive in the Jungle of Enterprise Architecture Framework: Creating or Choosing an Enterprise Architecture Framework*. Trafford, 2003.

[23] M. Indulska, P. Green, J. Recker, and M. Rosemann, “Business process modeling: Perceived benefits,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, pp. 458–471.

[24] A. K. Jallow, P. Demian, C. J. Anumba, and A. N. Baldwin, “An enterprise architecture framework for electronic requirements information management,” *International Journal of Information Management*, 2017.

[25] The-Open-Group, *TOGAF® Version 9.1* , 2011.

[26] M. Lankhorst, M. E. Iacob, H. Jonkers, L. Van Der Torre, H. A. Proper, F. Arbab, F. S. De Boer, M. Bonsangue, S. J. Hoppenbrouwers, G. V. Van Zanten, L. Groenewegen, R. Van Buuren, R. J. Slagter, J. Campschroer, M. W. Steen, A. W. Stam, R. J. Wieringa, P. A. Van Eck, D. Krukkert, H. W. Ter Doest, D. Van Leeuwen, P. Fennema, J. Jacob, H. Bosma, M. J. Cuvelier, P. G. Penders, S. F. Bekius, and W. P. Janssen, *Enterprise architecture at work: Modelling, communication, and analysis*, 2005.

[27] International Organization Of Standardization, "ISO/IEC/IEEE 42010:2011 - Systems and software engineering – Architecture description," *ISOIECIEEE 420102011E Revision of ISOIEC 420102007 and IEEE Std 14712000*, 2011.

[28] "The concise definition of the zachman framework," https://www.zachman.com/about-the-zachman-framework, accessed: 2019-11-26.

[29] W. Inmon, J. Zachman, and J. Geiger, "Data stores, data warehousing and the zachman framework: Managing enterprise knowledge," 01 1997.

[30] J. Colaço and P. Sousa, "Consolidation of business process models using an incremental and multi-stakeholder approach," 2017.

[31] C. M. Pereira, A. Caetano, and P. Sousa, "Using an organizational taxonomy to support business process design," *Advances in Enterprise Information Systems II - Proceedings of the 5th International Conference on Research and Practical Issues of Enterprise Information Systems, CONFENIS 2011*, pp. 147–154, 2012.

[32] L. Stein, "Genome annotation: From sequence to biology," 2001.

[33] S. J. Yang, I. Y. L. Chen, and N. W. Shao, "Ontology enabled annotation and knowledge management for collaborative learning in virtual learning community," 2004.

[34] U. Becker-Kornstaedt and R. Reinert, "A concept to support process model maintenance through systematic experience capture," in *ACM International Conference Proceeding Series*, 2002, pp. 465–468.

[35] N. Castela and J. Tribolet, "As-is representation in enterprise engineering: Processes and tools for its dynamic updating," 2011.

[36] S. B. Navathe and M. Schkolnick, "View representation in logical database design," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1978, pp. 144–156.

[37] S. B. Navathe and S. G. Gadgil, "METHODOLOGY FOR VIEW INTEGRATION IN LOGICAL DATABASE DESIGN." in *Very Large Data Bases, International Conference on Very Large Data Bases*, 1982, pp. 142–164.

[38] J. Mendling and C. Simon, "Business process design by view integration," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, pp. 55–64.

[39] H. Pichler and J. Eder, "A generic perspective model for the generation of business process views," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, pp. 477–482.

[40] W. M. Van Der Aalst, A. K. De Alves Medeiros, and A. J. Weijters, "Process equivalence: Comparing two process models based on observed behavior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, pp. 129–144.

[41] W. Van Der Aalst, A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, A. Burattin, J. Carmona, M. Castellanos, J. Claes, J. Cook, N. Costantini, F. Curbera, E. Damiani, M. De Leoni, P. Delias, B. F. Van Dongen, M. Dumas, S. Dustdar, D. Fahland, D. R. Ferreira, W. Gaaloul, F. Van Geffen, S. Goel, C. Günther, A. Guzzo, P. Harmon, A. Ter Hofstede, J. Hoogland, J. E. Ingvaldsen, K. Kato, R. Kuhn, A. Kumar, M. La Rosa, F. Maggi, D. Malerba, R. S. Mans, A. Manuel, M. McCreesh, P. Mello, J. Mendling, M. Montali, H. R. Motahari-Nezhad, M. Zur Muehlen, J. Munoz-Gama, L. Pontieri, J. Ribeiro, A. Rozinat, H. Seguel Pérez, R. Seguel Pérez, M. Sepúlveda, J. Sinur, P. Soffer, M. Song, A. Sperduti, G. Stilo, C. Stoel, K. Swenson, M. Talamo, W. Tan, C. Turner, J. Vanthienen, G. Varvaressos, E. Verbeek, M. Verdonk, R. Vigo, J. Wang, B. Weber, M. Weidlich, T. Weijters, L. Wen, M. Westergaard, and M. Wynn, "Process mining manifesto," in *Lecture Notes in Business Information Processing*, 2012, pp. 169–194.

[42] V. Liesaputra, S. Yongchareon, and S. Chaisiri, "A process mining technique using pattern recognition," in *CEUR Workshop Proceedings*, 2015, pp. 57–64.

[43] G. Schimm, "Process miner - A tool for mining process schemes from event-based data," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2002, pp. 525–528.

[44] "Google forms," docs.google.com/forms, accessed: 2020-12-02.

[45] "Cgd," www.cgd.pt, accessed: 2019-12-26.

[46] S. Esperto and P. Sousa, "Generation of business process views in finance context," 2017.

[47] N. Prat, I. Comyn-Wattiau, and J. Akoka, “Artifact evaluation in information systems design-science research - A holistic view,” in *Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014*, 2014.

[48] H. Österle, J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, and E. J. Sinz, “Memorandum on design-oriented information systems research,” 2011.