ENTERPRISE-WIDE BUSINESS PROCESS INTEGRATION FROM MULTIPLE CONCERNS

Joana Ribeiro
Instituto Superior Técnico, Universidade de Lisboa
Lisbon, Portugal
joana.f.p.ribeiro@tecnico.ulisboa.pt

***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.**

***Index Terms*—Business Process; Business Process Modeling; Stakeholder-specific Process View; View Integration; Consolidated Model; Form Inquiry.**

I. INTRODUCTION

Business Processes are designed to achieve specific goals and the task of Business Process Modeling has the goal of improving both the understanding and communication across the different stakeholder groups, which are two of its top three perceived benefits, together with process improvement [1].

Business Processes are the core asset of organizations, since they have a direct impact in the quality perceived by the market of the products and services offered. They shape the way of working within an organization, by determining tasks, jobs and responsibilities, and influencing its revenue potential and cost profile [2].

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.

Therefore, the goal of achieving a common understanding is difficult to achieve in the majority of the cases. Not only the business modeling languages lack a way for the stakeholders to express their concerns towards a given process, but also there are usually multiple conflicting diagrams found for the same business process and there is a lack of mechanisms to deal with domain changes and with the integration of requirements from multiple stakeholders. There is a lack of clarity and rigor in the production of the *AS-IS* and the *TO-BE* processes, which are crucial for the BPM lifecycle.

There are two main reasons that may be in the source of these issues that represent the motivation for this work [3], [4]:

- 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.

Giving these difficulties, keeping coherence and consistency between the different existing business process models that are a representation of the same business process can be very troublesome. Therefore, the task performed by the organization of managing them 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 [5], which could lead to different problems within the organization.

Therefore, the present work will focus on the problem of keeping consistency between Business Process Views, by building a model that centralizes all the information about them. Thus, this work aims to centralize the different stakeholder views into a Consolidated Model by asking the stakeholders a Set of Questions, in order to find a relation between the activities of the different views, so that it is possible to enrich them or produce new ones if they were nonexistent. 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, thus adding value to the organization.

II. Background

A 2017 survey [6] 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. 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 with the lack of clarity and difficulty in keeping the models updated, since they focus on how to help and deal with the stakeholders.

Business Process Management was defined by Dumas et. al [2] 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 Business Process Management [7] is adding value to the organization, its customer and stakeholders, hence BPM aims at managing Business Processes in the most effective way [2].

The **BPM Cycle** [2] provides a structured view of how a given process can be managed, and represents the concepts, methods, techniques and tools that compose the BPM discipline and were crucial to keep in mind when developing this research. It comprises the following phases: Process Identification, Process Discovery, Process Analysis, Process Redesign, Process Implementation, Process Monitoring and Controlling.

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** [8], which can be seen as tool that provides a structured overview of all the processes, their relationships and their boundaries.

Business Process Modeling was defined by Mendling [9] as "the human activity of creating a business process model", which is the result of specifying a Business Process through a Business Process Modeling Language, that offer a predefined set of elements and relationships for guiding the task of Business Process Modeling [9]. There are a panoply of modeling languages, but for this work we will make use of BPMN, Business Process Model and Notation [10], which is a modeling language whose goals were to provide a notation that is understandable by all business users and to create a bridge from a visual notation to execution languages.

Enterprise Architecture is 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" [11] 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. Among all of the EA Framework, we emphasize the Zachman Framework, which is the one closely related with our work, that will be explained up next.

III. Related Work

The **Zachman Framework** [12], initially proposed by John A. Zachman, is a framework for enterprise architecture that proposes a six by six matrix-like representation to classify descriptive representations relevant for describing an enterprise.

The Zachman Framework is a schema that consists in the intersection between two historical classifications [13], the fundamentals of communication found in the primitive interrogatives: What, How, When, Who, Where and Why as columns and the ratification transformations of an abstract idea into an instantiation: Identification, Definition, Representation, Specification, Configuration and Instantiation as rows.

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. This framework can be seen as a metamodel, the basis for the enterprise architecture [13] 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 [4] analyzed the Zachman Framework and proposed a rule for activity decomposition and equivalence. 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.

In [4], a Business Process Modeling approach, similar to the one used in data modeling which two entities are equivalent if they have the same properties, 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.

Colaço and Sousa [14] 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 [3].

Organization Taxonomy is another very relevant topic for this research, given that a clear taxonomy would increase the chances of all the people within the organization using the same concepts to represent the same things, and thus decreasing inconsistencies. Pereira et. al [15] 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 manner.

Based on the Zachman Framework [13], the authors present a taxonomy based on the categorization of each concept instantiation of a Business Process. The authors define a business process as a set of connected activities (How - Business Process) which consumes and produces tangible or intangible artefacts (What - Information Entity), is performed by people or systems (Who - Actor), contributes to achieving goals (Why - Business Goal), takes place in a specific location (Where - Organization Unit) and during a specific period of time (When - Business Schedule).

This proposal is of great relevance for this work, because it is not only connected with the Zachman Framework Dimensions but it 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.

Another meaningful topic is **View Integration**, which is the process that aims to combine different specific views of the same business process into a consolidated one. Navathe and Schkolnick [16] present a 4-step conceptual framework for logical database design which was later improved by Navathe and Gadgil [17]. The authors in [17] 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. Navathe and Gadgil [17] concluded 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. They have also concluded 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 [18] 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, proposing a method for business process design by view integration, analogous to a data schema integration approach.

Colaço and Sousa [14] 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, these 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.

The topic of **View Generation** also has relevance for the proposed solution, since after the construction of the Consolidated Model, the stakeholder-specific views will be generated once again, through activity enrichment since relevant detail can be added.

Artur Caetano et. al [19] proposed a Process View Generator, composed by three main logical components: the repository model, the controller and the viewer, 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.

Cardoso and Sousa [20] found a knowledge gap in the work developed by Artur Caetano et. al [19], 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.

Pichler and Eder [21] 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, having a more analytical purpose in what concerns generation of business processes, based on queries which formulate combinations of constraints on diverse perspectives.

Finally, **Process Equivalence** and **Process Mining** techniques portray interest for our solution in what concerns the production of the Consolidated Model by finding equivalence when comparing activities and the ability to discover and improve processes, respectively.

Aslst et. al [22] 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.

Mendling and Simon [18] specify a method for Business Process design by view integration starting from two views of a process as input, using EPCs [9]. 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 and reduction rules are introduced and to join similar EPC nodes.

Process Mining consists in discovering, monitoring and improving real processes by extracting knowledge from event logs available in Information Systems [23]. From the different existing types, using an Enhacement technique from the three available ones described in [23] 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 [24] to detect duplicate activities or using a mining tool [25] suitable for discovering hierarchically structured workflow processes can also be useful for our solution.

IV. Proposed Solution

A. The Big Picture

The Proposed Solution will now be presented at a broader angle, allowing for an end-to-end understanding. Figure 1

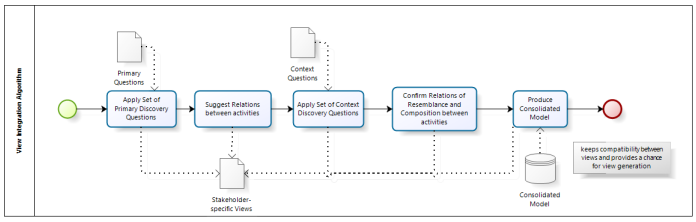

Fig. 1: Proposed Solution Big Picture

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.

B. Architecture and Design

We made the following assumptions regarding the Proposed Solution:

- 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 [10].
- 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 [14] 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.

C. Input/Output

The stakeholder-specific views work both as the input and output of the Consolidated Model, therefore the first step is to load them into the Process Repository.

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 [10].

A Secondary Relation Structure, the 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 resemblance a pair of activities has.

D. 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:

- **Primary Questions** - Questions posed for a single activity at a time, concerning its characteristics, properties and immediate environment. Includes the **Early Questions**, **Detail Questions** and **Responsibility Questions**
- **Context Questions** - 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 know 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 [26]. The addition of new and different types of questions, some even regarding two activities at a time, has a novelty feature and represents our contribution to this research topic.

We will implement the set of Relations Discovery Questions through Google Forms [27], by asking each stakeholder to answer the form with the chosen questions and thus produce the Consolidated Model.

*1) **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 chose among the available options in most cases, instead of freely writing their opinion. Given that the options are limited and usually short, their comparison is pretty straightforward.

The **Early Questions** are based in the Six Interrogatives discussed in, that are also deeply related with the Zachman Framework discussed in III. 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 [14], with a slight nuance. They 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.

As in the work of Colaço and Sousa [14], 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-specif 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 chose 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?

The corresponding options of answers are mapped with the elements of the business process models, hence we assume that a sweeper algorithm based on the view element classification by Colaço and Sousa [14] 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 bet 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 answers is valid if the one providing it is able to fit it into the existing hierarchy.

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, this solution is build to work with models designed in BPMN. Therefore, we used 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, 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 functioning, instead of utilizing the concepts themselves, we used their definitions and their roles within a business process model.

For instance, Question 1 and 2 aim to discover the antecedent and consequent activities, overshadowing the details that are to tight with the modeling activity, 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 and finally Question 9 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.

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.

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 Questions that would allow to discover relevant relations between activities. Therefore, the remaining questions of the first building block of this solution, named **Context Questions Questions**, 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 two 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.

E. 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 this 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.

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 stakeholder 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 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.

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.

F. Consolidated Model

After gathering all the relevant information from the different stakeholders through the Set of Relations Discovery Questions and discovering the Relations 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 alterations.

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 will 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.

G. View Integration Algorithm

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

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* functions 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 *generateResemblanceStructure*.

V. Demonstration

We will demonstrate the usage of our solution by applying it to a real-world scenario, with the goal of easing the understanding of this research and improve the reader's comprehension on how it can add value to an organization. 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.

For this demonstration, we will be analyzing the business process of Analysis, Decision and Granting of Credit of Caixa Geral de Depósitos [28]. Esperto and Sousa [29] produced stakeholder-specific views for this process, regarding the IT View, the Audit View and the Risk View. 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.

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

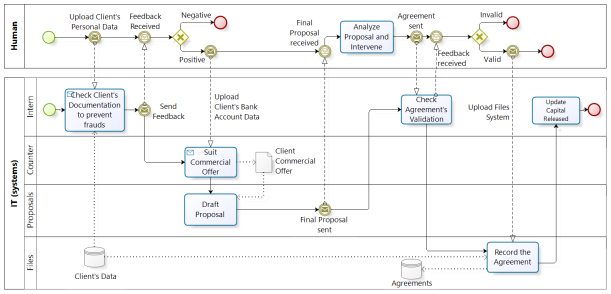

Fig. 2: IT View

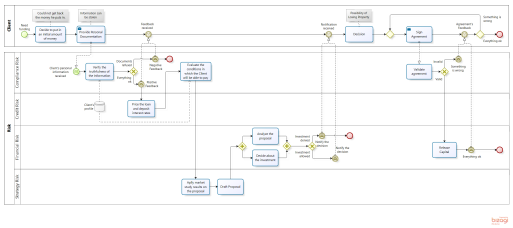

Fig. 3: Risk View

Figure 2 displays the IT View, which is responsible for the architecture, hardware, software and networking of the technology within the company. Figure 3 shows the Risk View, which concerns the identification, evaluation, and prioritization of risks within the company. Figure 4 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 [2], the Audit View [3] and the Risk View [4]. Afterwards, the survey regarding the

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

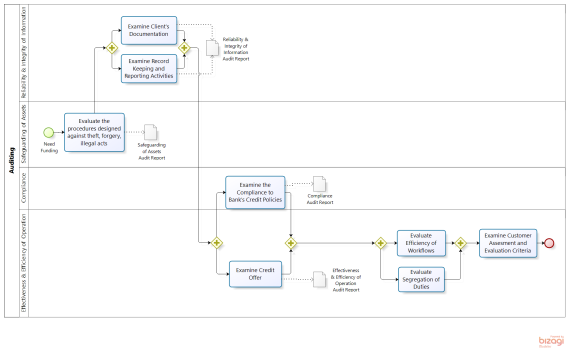

Fig. 4: Audit View

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

Fig. 5: Table with Relations found between Risk and Audit

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

Fig. 6: Table with Relations found between IT and Risk

Novel Question [5] 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, Table 6 and in Table 7 were found.

Examining the relations found, one could produce the Consolidated Model depicted in Figure 8. The Secondary Relation Structure for each relation is omitted, since it is just a visual representation of Figures 5, 6 and 7.

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 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.

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, making this solution successful.

[5]https://forms.gle/8VrvwUm847Jv4r1U7

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

Fig. 7: Table with Relations found between IT and Audit

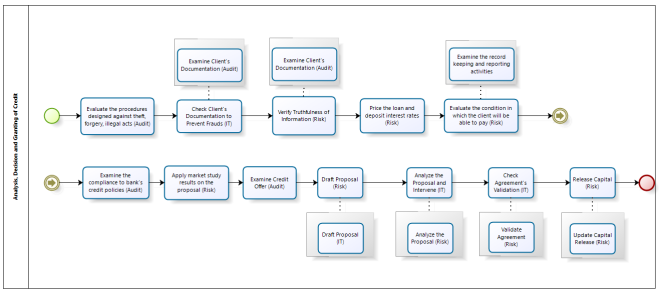

Fig. 8: Consolidated Model for Analysis, Decision and Granting of Credit Process

We inquired the stakeholders in order to evaluate concepts such as Usefulness, Understandibility Ease of Use Operation Feasibility, as defined by Prat et al. [30]. The stakeholders considered the solution useful with a score of 4, understandable with a score of 4, easy to use with a score of 4 and operational feasible with a score of 4, all in a scale of 1 to 5.

According to the O¨sterle et al. [31], 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.
- **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.

VI. Conclusions and Contributions

Our contribution to this topic stand on the following topics:

- The addition of new and different types of questions.
- 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.

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

- It is assumed that the solution only works for business process models represented in BPMN.
- 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.

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.

REFERENCES

[1] Marta Indulska, Peter Green, Jan Recker, and Michael Rosemann. Business process modeling: Perceived benefits. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 458–471, 2009.

[2] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Fundamentals of Business Process Management*. 2013.

[3] Carla Marques Pereira, Artur Caetano, and Pedro Sousa. Ontology-driven business process design. In *IFIP Advances in Information and Communication Technology*, pages 153–162, 2011.

[4] Carla Marques Pereira and Pedro Sousa. Business process modelling through equivalence of activity properties. In *ICEIS 2008 - Proceedings of the 10th International Conference on Enterprise Information Systems*, pages 137–146, 2008.

[5] Artur Caetano, António Rito Silva, and José Tribolet. Business Process Decomposition – An Approach Based on the Principle of Separation of Concerns. *Enterprise Modelling and Information Systems Architectures - An International Journal*, 5(1):44–57, 2010.

[6] Marcello La Rosa, Wil M.P. Van Der Aalst, Marlon Dumas, and Fredrik P. Milani. Business process variability modeling: A survey, 2017.

[7] Wil M.P. Van Der Aalst, Arthur H.M. Ter Hofstede, and Mathias Weske. Business process management: A survey, 2003.

[8] Remco Dijkman, Irene Vanderfeesten, and Hajo a Reijers. The Road to a Business Process Architecture: An Overview of Approaches and their Use. *Architecture*, 350(July):1–17, 2011.

[9] 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.

[10] Gustav Aagesen and John Krogstie. Bpmn 2.0 for modeling business processes. In *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*, pages 219–250. 2015.

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

[12] John A. Zachman. Framework for information systems architecture. *IBM Systems Journal*, 38(2):A Framework for Information Systems Infrastrucure, 1999.

[13] The concise definition of the zachman framework. https://www.zachman.com/about-the-zachman-framework. Accessed: 2019-11-26.

[14] João Colaço and Pedro Sousa. Consolidation of business process models using an incremental and multi-stakeholder approach. 2017.

[15] Carla Marques Pereira, Artur Caetano, and Pedro 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*, pages 147–154, 2012.

[16] Shamkant B. Navathe and Mario Schkolnick. View representation in logical database design. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 144–156, 1978.

[17] Shamkant B. Navathe and Suresh G. Gadgil. METHODOLOGY FOR VIEW INTEGRATION IN LOGICAL DATABASE DESIGN. In *Very Large Data Bases, International Conference on Very Large Data Bases*, pages 142–164, 1982.

[18] Jan Mendling and Carlo 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)*, pages 55–64, 2006.

[19] Artur Caetano, Carla Pereira, and Pedro Sousa. Generation of Business Process Model Views. *Procedia Technology*, pages 378–387, 2012.

[20] Diogo Cardoso and Pedro Sousa. Generation of stakeholder-specific bpmn models. 2019.

[21] Horst Pichler and Johann 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)*, pages 477–482, 2010.

[22] W. M.P. Van Der Aalst, A. K. De Alves Medeiros, and A. J.M.M. 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)*, pages 129–144, 2006.

[23] Van Der Aalst et. al. Process mining manifesto. In *Lecture Notes in Business Information Processing*, pages 169–194, 2012.

[24] Veronica Liesaputra, Sira Yongchareon, and Sivadon Chaisiri. A process mining technique using pattern recognition. In *CEUR Workshop Proceedings*, pages 57–64, 2015.

[25] Guido 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)*, pages 525–528, 2002.

[26] Nuno Castela and José Tribolet. As-is representation in enterprise engineering: Processes and tools for its dynamic updating. 2011.

[27] Google forms. docs.google.com/forms. Accessed: 2020-12-02.

[28] Cgd. www.cgd.pt. Accessed: 2019-12-26.

[29] Sara Esperto and Pedro Sousa. Generation of business process views in finance context. 2017.

[30] Nicolas Prat, Isabelle Comyn-Wattiau, and Jacky Akoka. Artifact evaluation in information systems design-science research - A holistic view. In *Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014*, 2014.

[31] Hubert Österle, Jörg Becker, Ulrich Frank, Thomas Hess, Dimitris Karagiannis, Helmut Krcmar, Peter Loos, Peter Mertens, Andreas Oberweis, and Elmar J. Sinz. Memorandum on design-oriented information systems research, 2011.